World Psychiatry

OFFICIAL JOURNAL OF THE WORLD PSYCHIATRIC ASSOCIATION (WPA)

Volume 22, Number 1 *February 2023*

WILEY

IMPACT FACTOR: 79.683

ISSN 1723-8617

The World Psychiatric Association (WPA)

The WPA is an association of national psychiatric societies aimed to increase knowledge and skills necessary for work in the field of mental health and the care for the mentally ill. Its member societies are presently 145, spanning 121 different countries and representing more than 250,000 psychiatrists.

The WPA organizes the World Congress of Psychiatry every year. It also organizes international and regional congresses and meetings, and thematic conferences. It has 66 scientific sections, aimed to disseminate information and promote collaborative work in specific domains of psychiatry. It has produced several educational programmes and series of books. It has developed ethical guidelines for psychiatric practice, including the Madrid Declaration (1996).

Further information on the WPA can be found on the website www.wpanet.org.

WPA Secretariat
Geneva University Psychiatric Hospital, 2 Chemin du Petit Bel-Air, 1226 Thônex, Geneva, Switzerland. Phone: +41223055737; Fax: +41223055735; E-mail: wpasecretariat@wpanet.org.

World Psychiatry

World Psychiatry is the official journal of the World Psychiatric Association. It is published in three issues per year and is sent free of charge to psychiatrists whose names and addresses are provided by WPA member societies and sections.

Research Reports containing unpublished data are welcome for submission to the journal. They should be subdivided into four sections (Introduction, Methods, Results, Discussion). References should be numbered consecutively in the text and listed at the end according to the following style:

1. Cuijpers P, Sijbrandij M, Koole SL et al. Adding psychotherapy to antidepressant medication in depression and anxiety disorders: a meta-analysis. World Psychiatry 2014;13: 56-67.
2. McRae TW. The impact of computers on accounting. London: Wiley, 1964.
3. Fraeijs de Veubeke B. Displacement and equilibrium models in the finite element method. In: Zienkiewicz OC, Hollister GS (eds). Stress analysis. London: Wiley, 1965:145-97.

All submissions should be sent to the office of the Editor.

Office of the Editor – Department of Psychiatry, University of Campania "L. Vanvitelli", Largo Madonna delle Grazie, 80138 Naples, Italy. Phone: +390815666502; Fax: +390815666523; E-mail: mario.maj@unicampania.it.

World Psychiatry **is indexed in PubMed, Current Contents/Clinical Medicine, Current Contents/Social and Behavioral Sciences, Science Citation Index, and EMBASE.**
All back issues of *World Psychiatry* can be downloaded free of charge from the PubMed system (http://www.ncbi.nlm.nih.gov/pmc/journals/297).

The need for a new generation of digital mental health tools to support more accessible, effective and equitable care

The potential of digital mental health to increase access to and quality of care has gained traction with the rise of smartphones and accelerated with the spread of telehealth during the COVID-19 pandemic. With at least 80% of the global population now owning a device able to capture digital phenotyping signals, analyze data, and run mental health apps, excitement about the imminent arrival of personalized, preventive and precision psychiatry is understandable.

Yet, by nearly all outcome metrics, digital mental health is not transforming care[1]. Whether measured in global trends of deaths from suicide or rising rates of depression, especially among younger people who are often the first to use digital tools, it is clear that the proclaimed paradigm shift is paused[2]. The very people who require mental health care the most, underserved populations, have not experienced a rise in access or boon in outcomes, and the burden of mental illness in low- and middle-income countries remains as high as ever.

Billions of dollars of resources have been poured into health apps, algorithms and devices with the assumption that later, with a simple step, all people would "cross over" or "trickle across" the digital divide and catch up. However, a variety of digital disparities are now emerging, which are troubling but perhaps also addressable. A focus on supporting digital literacy, improving privacy/evidence for these tools, and creating clinical connections each provides tangible steps for more equitable and impactful digital mental health.

As smartphone penetration has accelerated in all countries around the world, blaming the digital divide on a lack of access to devices has become untenable. This narrative now covers lack of Internet access, especially in rural areas. While this is indeed a barrier still requiring work today[3], it is one that can and will probably be quickly addressed. But, behind access to the Internet, lies a more challenging first inequity – that concerning digital self-determination.

Just as self-determination theory highlights the need for autonomy, competence and connection for psychological thriving, the same is necessary for any digital mental health tools, be they anything from smartphone apps to virtual reality headsets. While the data remain aloof as the topic has not yet been well explored, digital self-determination and the related sub-component of digital literacy remain underdeveloped in populations with the greatest mental health needs[4].

People may have a smartphone today, but there has not been a concomitant investment in people themselves to ensure that they can equitably engage and benefit from digital mental health tools. Evidence that older adults may find digital health tools more challenging, or that people from underserved backgrounds may engage less certainly, reflects issues with flawed designs of technology and a lack of community engagement, but may also reflect deeper inequities around educational opportunities that today's digital mental health approaches have not yet addressed[5].

Digital self-determination also means that people may say "no" to using technology for their mental health, and we should honor their choices and voices. A leading reason why people often say "no" is that today digital mental health tools have privacy practices compounded by limited evidence of efficacy. One of the clearest examples of inequity is the lack of privacy offered by most mental health apps. A report by the Mozilla Foundation in March 2022 highlighted ongoing privacy risks among well-known mental health apps. Around the same time in 2022, the suicide hotline service Crisis Textline agreed to stop sharing users' text messages with an outside company after public outcry.

The finding that less than 15% of people in the US and UK are willing to share anonymized personal health information with a company for the purposes of improving health care provides a tangible target for improvement[6]. The lack of trust engendered in health care technology must be reversed, and this can occur with better practices by app developers, demands for privacy by patients and clinicians, and regulation from governments. Without trust, there is no health or mental health, and it is understandable that people do not want their most private and vulnerable information shared in today's digital mental health ecosystem.

Furthermore, despite bold claims of efficacy on their websites, most studies in the mental health field do not recruit or sample from the patients with the highest unmet health care needs[7]. This clear lack of representativeness may explain why many digital technologies fail to offer impressive results in the real world when deployed outside clinical trial conditions. Digital mental health tools need not be perceived as second-class treatments to be utilized when a clinician is not available, but should strive for excellence that exceeds current standards of care. A more subtle but equally insidious bias rests in magnifying current inequalities when machine learning or artificial intelligence algorithms are trained on non-representative populations. As we think of the next generation of studies that can help reverse inequities, it is critical not to justify lower-quality research with the assumption that a digital intervention is better than nothing. If people have a phone, there are many free and effective interventions that can serve as an active control condition (or a digital placebo) to enable actual assessment of efficacy.

Coming to the third above-mentioned inequity, connections matter. As isolation and loneliness are recognized as public health threats, digital health tools will be most impactful when they help people form strong social connections instead of motivating them to continue focusing inward. The full potential of remote monitoring innovations, such as digital phenotyping and wearable sensors, as well as digital behavioral interventions, can only be realized when these are well integrated into care and treatment plans. That means that apps, devices and programs must transfer data to and from electronic medical records and that health workers and their workflow must be part of the design process.

Yet, less than 25% of apps today even allow such interoperabil-

ity[8], and, when supported at one major academic hospital, only 1% of people chose to link their app to their electronic health record[9]. Related, clinicians need training and support to incorporate such new digital health tools. A new workforce will be necessary, with a focus on peer support workers who may mirror the populations that are most impacted by a lack of access to and/or comfortability using technology, and who are ready to provide digital skill training and support.

Achieving optimal health, including mental health, means that we must address social/political determinants of health. Technology literacy now is considered a social determinant of health. It also impacts important aspects of people's lives, such as access to competitive employment, education, and even supportive services such as housing or access to other people, as clearly emerging during the COVID-19 pandemic. All of these aspects directly impact mental health and are as critically important as any clinical-focused use. Acknowledgment and integration of these social determinants can make digital tools more relevant and useful to a broader swath of the population with the highest need.

Thus, supporting digital self-determination should be the first priority, as it will create demand for new privacy protections, inform how the next generation of evidence will generate the highest quality of representative research, and ensure that new health care services are created to serve people with the highest needs. Developing a new generation of digital mental health tools/services to support more accessible, effective and equitable care is the true innovation ready to be stoked today by each person who becomes empowered to connect, set up, engage, start/stop, and demand more from mental health technology.

John Torous[1], Keris Myrick[2], Adrian Aguilera[3]
[1]Beth Israel Deaconess Medical Center, Boston, MA, USA; [2]Policy Liaison, US National Association of Peer Supporters; [3]Director, Digital Health Equity and Access Lab, University of California, Berkeley, CA, USA

1. Torous J, Bucci S, Bell IH et al. World Psychiatry 2021;20:318-35.
2. Stein DJ, Shoptaw SJ, Vigo DV et al. World Psychiatry 2022;21:393-414.
3. Perrin A. Mobile technology and home broadband 2021. Pew Research Center, June 3, 2021. www.pewresearch.org.
4. Hoffman L, Wisniewski H, Hays R et al. J Psychiatr Pract 2020;26:80.
5. Schueller SM, Hunter JF, Figueroa C et al. Curr Treat Options Psychiatry 2019;6:243-55.
6. Ghafur S, Van Dael J, Leis M et al. Lancet Digit Health 2020;2:e444-6.
7. Safavi K, Mathews SC, Bates DW et al. Health Aff 2019;38:115-23.
8. Barker W, Johnson C. J Am Med Inform Assoc 2021;28:2379-84.
9. Gordon WJ, Patel V, Thornhill W et al JAMA Netw Open 2020;3:e2022408.

DOI:10.1002/wps.21058

The drug treatment deadlock in psychiatry and the route forward

The US Food and Drug Administration (FDA) approved 12 new drugs in psychiatry during the decade 2011-2021 (www.fda.gov/drugs/drug-approvals-and-databases). In comparison, it approved 50 new drugs in neurology and 135 in oncology over the same period. The FDA designated two new drugs as first-in-class in psychiatry (lofexidine and brexanolone) in the most recent reviewed period (2015-2021), compared to 13 in neurology and 31 in oncology (www.fda.gov/drugs/development-approval-process-drugs). These data highlight a dearth of new drug treatments and novel mechanistic approaches across psychiatry, both in absolute and comparative terms. They indicate that psychiatry faces a deadlock in drug development.

One reason for this deadlock is represented by the challenges of conducting clinical trials in psychiatry, due to factors such as high placebo response rates in some disorders, as reviewed by Correll et al[1] in this issue of the journal. These challenges mean that trials have to be large and, consequently, expensive. Large trials generally require many sites, but having more sites has been associated with higher placebo response[1], meaning that this solution may make the problem worse. Another factor is that a number of drug companies – including Pfizer, Eli Lilly, GlaxoSmithKline and Astra-Zeneca – have largely stopped psychiatric drug development. It should be no surprise then that there are fewer new compounds coming through to approval in psychiatry. Finally, it is striking that many of the psychiatric drugs currently in development target the same mechanisms as already approved treatments, with few new classes of medications in the pipeline.

In this situation, the first necessary step is to address some of the challenges in conducting clinical trials in psychiatry. Instead of adding more sites, a potential solution is to use fewer, higher quality sites to minimize noise and reduce the placebo response rate. Another is the use of digital technologies to provide both better standardization of measures and more data. Smart designs also offer the potential to make trials more efficient and informative.

However, addressing these challenges will be of little use if there are no new drugs to test. Companies need to be attracted into psychiatry if we are to see the development of new treatments. There is some light on the horizon: new companies are entering psychiatry in some areas, notably in the development of serotonin 2A receptor agonists, such as psilocybin for major depression and related disorders. Investment in this area exceeded US$500 million in 2021[2]. This is encouraging, but needs to be replicated in other areas of psychiatry if we are to see wholesale progress.

The investment in serotonin 2A receptor agonists is also striking in that it came after well over a decade of research into the use of these compounds by academic groups[3]. This highlights the synergism between academic research and drug development: drug developers grow their ideas from mechanistic and clinical understanding of disorder. It also illustrates the need for sustained investment in translational research in psychiatry to sow the seeds for future drug development. This requires the engagement of governments and charitable funders. It is noteworthy, in this respect, that both neurology and oncology have seen large-scale, long-term research investment from charities such as Can-

cer Research UK and the Michael J. Fox Foundation, which psychiatry has not seen.

Another potential strategy is to form pre-competitive partnerships between companies and academia to generate the clinical evidence in an area to guide future drug development. Governments and regulators could also incentivize companies to invest in psychiatric drug development through, for example, tax breaks or longer patent recognition, in consideration of the challenges and unmet need in psychiatry.

Much psychiatric drug development has been based on astute clinical observation and empirical studies, followed by extensive efforts to then develop related compounds. This has given us a wide choice of medications for some conditions but few mechanistically distinct treatments. We have harvested serendipity's bounty over many decades now and, it seems, there are few low-hanging fruits left.

It is striking how much remains to be established about the link between pathophysiology and psychiatric symptoms[e.g.,4]. To develop mechanistically new treatment approaches, we will need to advance understanding of the neurobiology underlying psychiatric disorders; in particular, of the link between molecular processes and symptoms, to be able to identify new molecular targets for drugs. We also need to recognize that psychiatric disorders usually involve multiple brain systems and show clinical heterogeneity. Accordingly, successful treatment approaches of the future may need to be promiscuous in their targets and/or we will need to address clinical heterogeneity, for example by subtyping disorders to particular systems that can be targeted by more selective drugs[5,6]. This will require investment in research into neurobiology, for example in post-mortem or molecular imaging studies, and the link to psychological processes and social factors.

We also need to understand the neurobiology underlying poor response to existing treatments, not least because this is where some of the greatest unmet needs lie[7]. This has not been a focus for research traditionally, but evidence is beginning to accrue that there are neurobiological differences linked to poor treatment response, for example in schizophrenia[6,8], that identify new treatment targets[7].

Greater understanding of the neurobiology underlying psychiatric symptom domains will support the development of biomarkers that can be used to identify the right patients in whom to test a given drug, and to evaluate the effects of that drug. Furthermore, we need preclinical models that reproduce the neurobiology seen in patients. Back translation from patient findings, as has been done for the elevated striatal dopamine synthesis capacity seen in schizophrenia[9], is one approach. Another is the use of stem cell technologies that allow drugs to be tested in neurons derived from patients.

Overall, whilst in the short term strategies can be implemented to improve the design of clinical trials, ultimately much more research into the neurobiology of psychiatric disorders will be needed if we are to see the step-change in treatment approaches that has been observed in neurology and oncology.

Oliver D. Howes[1-3], Luke Baxter[1,3]
[1]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; [2]Institute of Clinical Sciences, Faculty of Medicine, Imperial College London, London, UK; [3]South London and Maudsley NHS Foundation Trust, London, UK

The authors are funded by the UK Medical Research Council, Maudsley Charity, Wellcome Trust, and the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed here are those of the authors and not necessarily those of the funding bodies.

1. Correll CU, Solmi M, Cortese S et al. World Psychiatry 2023;22:48-74.
2. Tracy HM. NeuroPerspective 2022;Winter:313-5.
3. Chi T, Gold JA. J Neurol Sci 2020;411:116715.
4. Ashok AH, Marques TR, Jauhar S et al. Mol Psychiatry 2017;22:666-79.
5. Scannell J, Blanckley A, Boldon H et al. Nat Rev Drug Discov 2012;11:191-200.
6. Veronese M, Santangelo B, Jauhar S et al. Neuropsychopharmacology 2021;46:1122-32.
7. Potkin SG, Kane JM, Correll CU et al. NPJ Schizophr 2020;6:1.
8. Mouchlianitis E, Bloomfield MA, Law V et al. Schizophr Bull 2016;42:744-52.
9. Kokkinou M, Irvine EE, Bonsall DR et al. Mol Psychiatry 2021;26:2562-76.

DOI:10.1002/wps.21059

New insights from the last decade of research in psychiatric genetics: discoveries, challenges and clinical implications

Ole A. Andreassen[1,2], Guy F.L. Hindley[1,3], Oleksandr Frei[1,4], Olav B. Smeland[1,2]

[1]NORMENT Centre, Institute of Clinical Medicine, University of Oslo, Oslo, Norway; [2]Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway; [3]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; [4]Centre for Bioinformatics, Department of Informatics, University of Oslo, Oslo, Norway

Psychiatric genetics has made substantial progress in the last decade, providing new insights into the genetic etiology of psychiatric disorders, and paving the way for precision psychiatry, in which individual genetic profiles may be used to personalize risk assessment and inform clinical decision-making. Long recognized to be heritable, recent evidence shows that psychiatric disorders are influenced by thousands of genetic variants acting together. Most of these variants are commonly occurring, meaning that every individual has a genetic risk to each psychiatric disorder, from low to high. A series of large-scale genetic studies have discovered an increasing number of common and rare genetic variants robustly associated with major psychiatric disorders. The most convincing biological interpretation of the genetic findings implicates altered synaptic function in autism spectrum disorder and schizophrenia. However, the mechanistic understanding is still incomplete. In line with their extensive clinical and epidemiological overlap, psychiatric disorders appear to exist on genetic continua and share a large degree of genetic risk with one another. This provides further support to the notion that current psychiatric diagnoses do not represent distinct pathogenic entities, which may inform ongoing attempts to reconceptualize psychiatric nosology. Psychiatric disorders also share genetic influences with a range of behavioral and somatic traits and diseases, including brain structures, cognitive function, immunological phenotypes and cardiovascular disease, suggesting shared genetic etiology of potential clinical importance. Current polygenic risk score tools, which predict individual genetic susceptibility to illness, do not yet provide clinically actionable information. However, their precision is likely to improve in the coming years, and they may eventually become part of clinical practice, stressing the need to educate clinicians and patients about their potential use and misuse. This review discusses key recent insights from psychiatric genetics and their possible clinical applications, and suggests future directions.

Key words: Genetics, genomics, psychiatry, precision medicine, common variants, rare variants, pleiotropy, polygenic risk score, nosology

(World Psychiatry 2023;22:4–24)

Psychiatric disorders are among the main causes of morbidity[1] and mortality[2] worldwide, posing a substantial burden on individuals and society. They typically begin in adolescence or young adulthood and often have a chronic course, leading to many years lived with debilitating illness. In addition, individuals with severe mental illness often have poorer socioeconomic status[3,4], frequently experience stigma[5], and have a higher occurrence of both substance use[6] and somatic disease[7], all of which negatively affect well-being and quality of life. The average life expectancy of people with severe mental illness is estimated to be approximately 10 years shorter compared to the general population[2,8], with the excess mortality due to both physical health causes, particularly cardiovascular disease[9,10], and mental health-related causes, such as suicide[11].

As emphasized by the World Health Organization[12], there is an urgent need to improve mental health care. Existing treatment modalities may provide clinically meaningful effects in many psychiatric disorders[13,14]. However, treatment is rarely curative – many patients experience relapses and unpleasant adverse effects, and lack of therapeutic response is common[15,16]. Inadequate therapeutic options can largely be attributed to the limited understanding of the causes of mental illness, despite decades of intensive research. By the same token, psychiatric nosology still relies on traditional diagnostic distinctions based on clinical observations[17,18]. In the two current leading diagnostic classification systems, the International Classification of Diseases[17] and the Diagnostic and Statistical Manual of Mental Disorders[18], psychiatric disorders are still primarily diagnosed according to their signs and symptoms. There is a lack of objective biomarkers, in contrast to most other medical fields, making clinical psychiatry more susceptible to unwanted variability in both diagnostic and therapeutic decision-making[19]. Although the present diagnostic categories have clinical utility, there is little evidence to suggest that they represent truly discrete entities with natural boundaries[20,21], as indicated by the high comorbidity and shared symptomatology across different mental disorders[22,23], and the high heterogeneity within diagnostic categories[24].

To improve the care and prevention of mental illness, a better understanding of the underlying biological mechanisms is needed. The intrinsic challenges in studying the living human brain and the uncertain validity of animal models of mental illness[25] have limited progress of biological research in psychiatry. As a consequence, there have been no major therapeutic advances in psychiatry in the past decades[26], and the potential new treatment options that currently receive most attention represent repurposing of existing drugs such as ketamine[27] or psychedelics[28]. However, the substantial heritability of psychiatric disorders[29] indicates that genetic research could uncover otherwise inaccessible pathobiological insights, and could also aid in disentangling environmental effects and gene-environment interplay.

Despite great expectations as DNA sequencing technologies became more widely available over the course of the second half of the 20th century, psychiatric genetics got off to something of a false start in the 1990s and early 2000s. A series of findings using a candidate gene approach were subsequently shown to lack reproducibility, reducing confidence that genetic research could lead to the discovery of genes for mental illness[30,31]. The major turning points came with the sequencing of the human genome

in 2003[32], and the creation of reference datasets cataloguing human genetic variation across different populations[33,34], which allowed for a systematic exploration of DNA sequence variants linked to human traits and diseases.

Since then, there has been a steady and accelerating progress in human genetics[35], driven by a combination of technological innovation, more advanced statistical analytical tools, reduced costs for genotyping and sequencing DNA, more precise knowledge about the genome, and international collaboration. Psychiatric genetics has been at the forefront of these efforts, recognizing the need to assemble large-scale case-control cohorts of psychiatric disorders to reliably identify genetic variants, most of which have very weak effects, which have gradually led to the discovery of multiple genetic risk variants for mental illness[36,37]. However, while the last decade has brought major advances in our understanding of the genetic architecture of mental illness, these discoveries have not yet been translated into improved care for people with mental illness, which remains the key challenge for the field.

Here, we aim to provide a comprehensive review of the genetic risk underlying psychiatric disorders. We summarize what we have learnt from genetic research in psychiatry during the past decade, focusing on attention-deficit/hyperactivity disorder (ADHD), anorexia nervosa, anxiety disorders, autism spectrum disorder, bipolar disorder, major depressive disorder, obsessive-compulsive disorder (OCD), post-traumatic stress disorder (PTSD), schizophrenia, and Tourette's syndrome. We also discuss how the advances in genetics may enable precision medicine approaches, and we discuss future directions, challenges and opportunities.

DISSECTING THE GENETIC RISK OF MENTAL ILLNESS

The nature vs. nurture debate on the causes of mental illness is now understood to be a false dichotomy[38,39]. Variation in risk of mental illness is neither solely due to variation in DNA or environmental factors, but both nature *and* nurture unequivocally contribute in closely intertwined processes.

For millennia, it has been observed that mental illness tends to run in families[40,41]. This familial aggregation has since been confirmed by large-scale family and population-based studies. For example, first-degree relatives of a proband with bipolar disorder or schizophrenia have approximately 6-8 and 10 times higher risk of developing these disorders, respectively, compared to relatives without an affected family member[42]. Relatives of probands with a psychiatric disorder also have increased risk of developing other psychiatric disorders[43], which indicates that familial risk of mental illness transcends diagnostic categories, suggesting shared etiology.

In the past 50 years, twin, adoption, family and population-based studies of increasing quality have demonstrated that all major psychiatric disorders have a substantial heritability, meaning that a considerable proportion of the variation in risk of developing mental illness is attributable to differences in genetic factors between individuals[29,44]. Environmental exposures, including social determinants, also influence risk of illness along with genetic factors[45], with the relative contributions varying across disorders[36]. The etiology of psychiatric disorders may also be influenced by non-inherited somatic DNA variants accumulating in brain tissue throughout development and ageing[46], as well as by stochastic variation in biological processes[47].

The estimated heritabilities are generally higher in psychotic and neurodevelopmental disorders (74-85%)[42,48-51] than in mood and anxiety disorders (37-58%)[52,53] (see Figure 1), indicating that a larger fraction of the variation in risk of developing mood and anxiety disorders is explained by environmental factors. Note that the estimated heritability of a specific disorder can vary between populations, due to population-specific variation in genetic and environmental factors, and differences in phenotypic definitions such as diagnostic criteria.

Regardless of the heritability of a trait, identifying genetic risk variants could potentially yield valuable insights into its etiology by pointing to core biological mechanisms. In human DNA, there are millions of genetic variants that differ between individuals and may confer risk or protect against illness[54]. A genetic variant may represent a difference in a single genomic position, such as a single nucleotide polymorphism (SNP), in which a single nucleotide in DNA differs between people, or larger structural changes such as copy number variants (CNVs), which are deletions or duplications of genomic regions.

According to the frequency in a population of the less frequent allele (termed the minor allele frequency, MAF), genetic variants are typically defined as common (MAF >1%), uncommon (MAF 0.1-1%), rare (MAF <0.1%), and ultra-rare (MAF <0.001%), although the exact definitions vary to some extent across studies. In addition to inherited variants, newly occurred *de novo* mutations in an individual may also influence risk of illness and potentially exert large phenotypic effects.

Importantly, genomic findings in a given population cannot be readily generalized to populations of other ancestries, since the frequency of variants, their specific effect sizes, as well as the non-random correlation pattern among variants (referred to as linkage disequilibrium, LD) vary across ancestries, in addition to the different environmental contexts[55,56].

COMMON VARIANTS

In the past decade, genome-wide association studies (GWAS) have become the most successful approach to link genetic variants to human phenotypes[57]. A GWAS systematically screens millions of common genetic variants for association with a given phenotype in a hypothesis-free manner, by comparing the frequency of variants in cases vs. controls or across a continuous measure. In order to conduct a GWAS, hundreds of thousands of common genetic variants are genotyped in each individual participant, using relatively inexpensive SNP arrays, and additional genetic variants are imputed to generate complete genotypes for

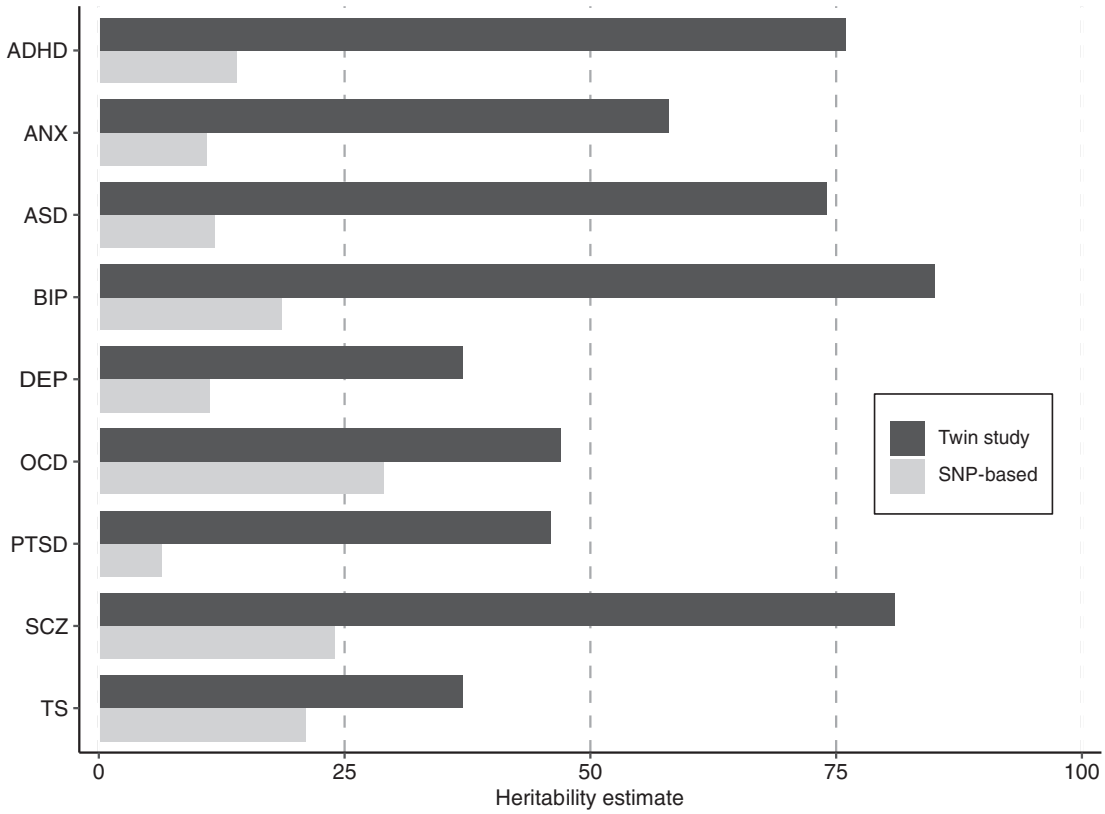

Figure 1 Estimates of twin-heritability (black) and single nucleotide polymorphism (SNP)-based heritability (grey) for major psychiatric disorders. ADHD – attention-deficit/hyperactivity disorder, ANX – anxiety, ASD – autism spectrum disorder, BIP – bipolar disorder, DEP – depression, OCD – obsessive-compulsive disorder, PTSD – post-traumatic stress disorder, SCZ – schizophrenia, TS – Tourette's syndrome.

each individual.

After the first GWAS reporting significant variant associations with a complex human phenotype was published in 2005[58], the number and sample sizes of GWAS have grown exponentially[59]. At the time of writing, GWAS have identified associations between more than 400,000 common genetic variants and hundreds of human traits and disorders according to the GWAS catalog[60], and the number is rapidly increasing. Note that GWAS typically report trait-associated genomic loci, which are DNA regions that involve multiple genetic variants highly correlated with each other due to LD, wherein one or several variants may independently influence the phenotype.

The ability of a GWAS to identify a trait-affecting genetic variant depends on the population prevalence of the variant, its strength of association with the trait, and the statistical power of the study, which corresponds to its sample size. Hence, as GWAS samples increase in size, more genetic variants are discovered. Since a GWAS scans a large number of SNPs, it is necessary to control for multiple comparisons to avoid false positive findings, which results in a stringent genome-wide significance threshold, typically $p<5x10^{-8}$. Moreover, since common genetic variants have tiny effects (e.g., small differences in the frequency of risk alleles between cases and controls), very large GWAS sample sizes are needed to achieve sufficient statistical power to discover SNPs passing the genome-wide significance threshold.

The ability of GWAS to discover SNPs also depends on the unique characteristics of the common variant architecture underlying a phenotype[61]. This includes the polygenicity of the phenotype, which refers to the number of common genetic variants influencing the phenotype, and the SNP-heritability, which refers to the proportion of phenotypic variance explained by common genetic variants. Estimates of SNP-heritability[62,63] have confirmed that part of the risk of developing psychiatric disorders is captured by common genetic variation, with SNP-heritabilities ranging between 5 and 25% for ten major psychiatric disorders[64-73] (see Figure 1 and Table 1).

The estimated SNP-heritabilities for psychiatric disorders are thus much lower than the estimated twin-heritabilities[42,48-53]. This issue is often referred to as the "missing heritability" problem[74], and also applies to other behavioral and somatic phenotypes. This problem is still not fully resolved, but may be explained by rare variants which are not included in the standard GWAS, gene-gene or gene-environment interplay, and inflated twin-heritability estimates, possibilities which are not mutually exclusive[74-76]. However, a recent study indicated pervasive downward bias of standard SNP-heritability estimates, suggesting that the SNP-heritabilities of psychiatric disorders may in reality be higher than current estimates[77].

Table 1 Summary of largest genome-wide association studies (GWAS) on major psychiatric disorders

Disorder	Largest GWAS	Cases	Controls	Ancestry	GWAS loci	SNP-heritability
ADHD	Demontis et al[68]	38,691	186,843	European	27	14%
AN	Watson et al[67]	16,992	55,525	European	8	11%
ANX	Levey et al[69]	175,163 (continuous measure)	-	European	5	5.6%
ASD	Grove et al[66]	18,381	27,969	European	5	11.8%
BIP	Mullins et al[65]	41,917	371,549	European	64	18.6%
DEP	Levey et al[70]	340,591	813,676	European	178	11.3%
OCD	Strom et al[73]	14,140	562,117	European	1	16.4%
PTSD	Stein et al[71]	59,513	329,554	European	4	6.4%
SCZ	Trubetskoy et al[64]	76,755	243,649	European (86%), East Asian (10%), African American (3%) and Latino (1%)	287	24%
TS	Yu et al[72]	4,819	9,488	European	1	21%

Risk loci identified at the genome-wide significance threshold. SNP-heritability estimated using LD score regression. ADHD – attention-deficit/hyperactivity disorder, AN – anorexia nervosa, ANX – anxiety, ASD – autism spectrum disorder, BIP – bipolar disorder, DEP – depression, OCD – obsessive-compulsive disorder, PTSD – post-traumatic stress disorder, SCZ – schizophrenia, TS – Tourette's syndrome.

One of the key insights emerging from GWAS is that most complex human phenotypes are highly polygenic, influenced by thousands of common variants with miniscule effects[59]. Hence, there is no single "disease-gene" for psychiatric disorders, but thousands of genetic variants that act together and collectively influence risk of illness. Given that most of these genetic variants are commonly occurring, every human being has a genetic risk to each psychiatric disorder, from low to high.

Compared to somatic phenotypes, both psychiatric disorders and behavioral phenotypes generally have larger polygenicities despite similar SNP-heritabilities[61,78,79]. This means that each common variant tends to have smaller effects in behavioral than somatic phenotypes. As a consequence, larger GWAS sample sizes are needed to identify a comparable fraction of the common variant architectures underlying psychiatric disorders than somatic disorders (see Figure 2). As an example, approximately one third of the heritability of Crohn's disease due to common genetic variants has been identified by GWAS with 12k cases and 34k controls[80]. In comparison, more than 10 times the number of GWAS participants are estimated to be needed to identify a similar proportion of the common genetic variance underlying schizophrenia (see Figure 2). Thus, given the high polygenicities of psychiatric disorders, which likely reflect more complex and/or heterogenous genetic etiologies, their GWAS discovery trajectories are still trailing those of somatic traits and disorders by many years.

Large-scale international collaboration, with the Psychiatric Genomics Consortium[81] as the main driving force, has led to the assembly of increasingly productive GWAS involving tens of thousands of participants, discovering reproducible common variant associations for most major psychiatric disorders[64-73] (see Table 1). In addition, several GWAS of other clinically relevant phenotypes in psychiatry have been published in recent years, such as treatment resistance in schizophrenia[82], response to lithium[83], antidepressant response[84], suicide attempt[85], cognitive function[86], insomnia[87], risky behavior[88], mood instability[89], and antisocial behavior[90]. Well-powered GWAS on substance use disorders have also been conducted in recent years[91,92]. However, there is still a lack of sufficiently powered GWAS on personality disorders[93] and eating disorders, apart from anorexia nervosa[94]. Overall, the common variant data on psychiatric disorders are consistent with a liability threshold model, in which a large number of risk alleles additively contribute to overall risk. Individually, the trait-associated common variants have minuscule effects on risk of illness, with odds ratios generally below 1.2.

The most well-powered GWAS in psychiatry to date is on schizophrenia, comprising 76,755 cases and 243,649 controls, in which 287 distinct genomic loci harboring genome-wide significant common variant associations were discovered[64]. Despite this success, the independent significant genetic variants still explain less than 10% of the SNP-heritability of schizophrenia, indicating that most of its common variant architecture remains to be identified (see Figure 2). In other psychiatric disorders, GWAS have even lower power, and the proportion of SNP-heritability explained by genome-wide significant variants is correspondingly lower (see Figure 2).

Estimates of polygenicity indicate that tens of thousands common genetic variants may influence each psychiatric disorder, although there is a considerable margin of uncertainty in these estimates[61]. In a recent cross-disorder investigation of GWAS data, depression appeared to be more than twice as polygenic as ADHD, possibly reflecting less biological heterogeneity in ADHD than depression[95]. Note that the genetic investigation of depressive disorders has focused on different phenotypic definitions, owing to different case ascertainment. While major depressive disorder refers to cases found to meet standard diagnostic criteria after structured interviews by trained interviewers, the de-

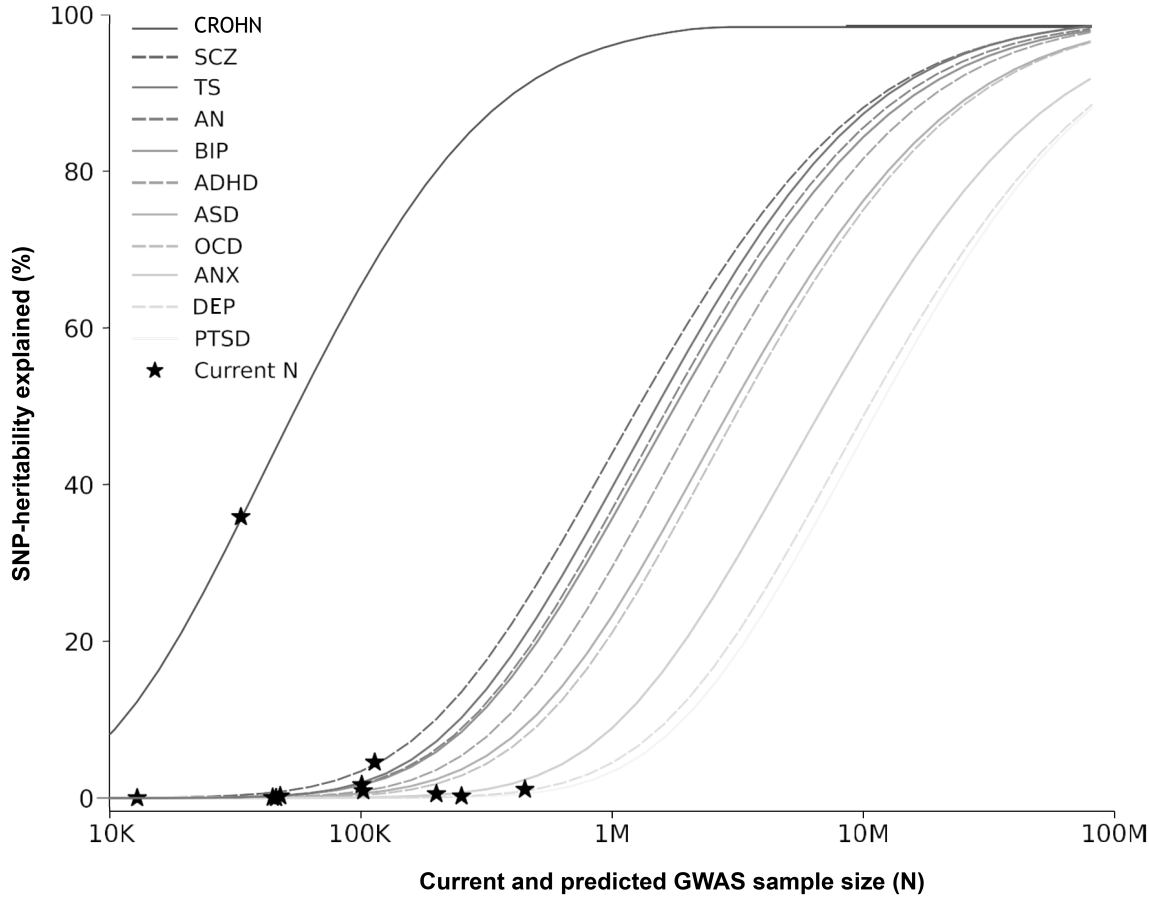

Figure 2 Statistical power calculations for current and future genome-wide association studies (GWAS) on major psychiatric disorders. The figure shows the proportion of single nucleotide polymorphism (SNP)-based heritability explained by variants detected at the genome-wide significance threshold (vertical axis) as a function of GWAS sample size across psychiatric disorders. Crohn's disease (CROHN) is included as an example of somatic disorder. For each disorder, the current effective sample size (indicated by asterisk) is shown. ADHD – attention-deficit/hyperactivity disorder, AN – anorexia nervosa, ANX – anxiety, ASD – autism spectrum disorder, BIP – bipolar disorder, DEP – depression, OCD – obsessive-compulsive disorder, PTSD – post-traumatic stress disorder, SCZ – schizophrenia, TS – Tourette's syndrome.

pression phenotype also includes self-reported treatment or diagnosis of clinical depression, and is therefore less specific[96].

RARE VARIANTS

In the past decade, rare and *de novo* sequence variants and pathogenic CNVs have been implicated in most psychiatric disorders, except for eating disorders and personality disorders. Due to their low frequency, rare variants explain less heritability in the population than common genetic variants. However, rare variants may confer substantially higher risk of illness in the individual, due to more deleterious impact on protein function or expression or, in the case of CNVs, by impacting several genes.

There is robust evidence that the burden of rare large-effect variants is highest in neurodevelopmental disorders and psychotic disorders, in particular in cases with intellectual disability or developmental delay[97-100]. This is in line with the decreased fecundity associated with neurodevelopmental and psychotic disorders[101], which prevents genetic variants with large effects on risk of illness from becoming common in the population. Correspondingly, *de novo* variation, which on average has been exposed to less selective pressure, shows more severe predicted functional consequences than inherited variation[102].

Whole exome sequencing (WES) and whole genome sequencing (WGS) studies are generally underpowered to detect specific rare single nucleotide variants (SNVs), given the rarity of these variants. To circumvent this issue, a common approach is to evaluate the burden of rare sequence variants in individual genes by comparing cases and controls or using family-based designs. To reduce the number of variants assessed, it is also common to focus on exonic SNVs using WES data, thereby ignoring the vast number of noncoding variants in WGS data.

WES studies in autism spectrum disorder[98,102,103], ADHD[98], Tourette's syndrome[104], OCD[105], schizophrenia[98,99,106,107], bipolar disorder[98,108] and major depressive disorder[109] have revealed an excess burden of ultra-rare protein-truncating and damaging missense variants in genes under strong evolutionary constraint,

and discovered many specific risk genes, in particular in autism spectrum disorder[98,102,103] and schizophrenia[98,99,106,107].

The identified protein-truncating variants typically result in partial or complete loss of protein function, while the missense variants have a less deleterious impact. The evolutionary constrained genes have a high probability of being intolerant to loss-of-function mutations, and are relatively depleted of equivalent protein-disrupting variants in the general population[110]. Recently, such genes have also been implicated by common variant findings for schizophrenia[111].

Rare CNVs at several loci have been robustly associated with autism spectrum disorder[100], schizophrenia[112,113] and ADHD[114], while only a few specific CNVs have been implicated in OCD[115], Tourette's syndrome[116], major depressive disorder[117], and bipolar disorder[118]. A CNV study in bipolar disorder found that only cases with schizoaffective disorder, bipolar type were enriched for CNVs[119], further indicating that rare CNVs may play a larger role in psychotic than mood disorders.

Despite potentially having very large effects, the penetrance of rare pathogenic variants is incomplete, meaning that only a fraction of carriers display a certain clinical outcome. Moreover, carriers may present a wide range of health outcomes, depending on the individual's DNA constitution, environmental stimuli, and chance.

By integrating datasets on both rare and common genetic variants in autism spectrum disorder[120] and schizophrenia[121,122], it has been demonstrated that genetic variation at both ends of the allele frequency spectrum jointly influences risk of these disorders in the same individuals. For instance, the clinical outcomes of 22q11.2 deletions are highly heterogenous, including schizophrenia, autism spectrum disorder, ADHD, cognitive dysfunction, neurological disorders and somatic abnormalities[123]. Among carriers of 22q11.2 deletions, a higher burden of common risk alleles for schizophrenia was linked to higher risk of that disease[124]. This indicates that common genetic risk may modulate the penetrance of rare variants such as the 22q11.2 deletion, and may eventually help prediction of clinical outcomes such as psychosis in this patient group.

In autism spectrum disorder, a recent study demonstrated an inverse correlation of the burden of rare and common genetic variants among cases, indicating a spectrum of genetic risk among cases, ranging between more monogenic to polygenic risk architectures[125]. Moreover, different aspects of the common and rare genetic risk were differently associated with clinical measures in the disorder[125]. This indicates that different genetic loadings may map to different aspects of the phenotypic spectrum, pointing to potential utility of genetic profiling in the clinic to facilitate more personalized treatment.

EMERGING BIOLOGICAL INSIGHTS

One of the key aims of human genetics is to gain insights into the underlying etiology of illness, which might inform the development of new therapeutic interventions and help identify biomarkers. However, translating genetic findings into biological mechanisms is not straightforward. To obtain a complete mechanistic understanding of a disorder's genetic risk architecture, it is necessary to: a) identify the specific causal variant underlying a genetic signal; b) determine the functional impact of the genetic variant; and c) determine how all of the genetic risk variants act together to collectively influence biological pathways in specific cell types, tissues and organs, across developmental stages, and in concert with environmental factors[126,127]. This is a tremendous challenge, warranting comprehensive animal studies, cell-biology experiments, and advanced computational approaches. The current mechanistic interpretation is also limited by the incomplete understanding of the physiological role of most genes and proteins, including how they interact in signaling networks and pathways[128].

Fine-mapping procedures, for example leveraging trans-ancestry tools[129], may help prioritize the most likely causal variants in GWAS loci[130]. However, the causal variant does not necessarily affect the closest gene. A genetic variant may exert its phenotypic effect by disrupting a single protein structure and function, or by regulating the expression of one or more genes locally or over long genomic distances. Indeed, most GWAS associations are detected in noncoding regions[131,132], suggesting that most common variants may exert their phenotypic effect through regulatory mechanisms, complicating mechanistic interpretation. To help prioritize the most likely causal genes from GWAS loci, algorithms integrating diverse functional resources have been developed in recent years[133,134].

The biological interpretation of rare variants largely depends on the type of variant in question. Since most rare pathogenic CNVs disrupt large genomic segments, often including many genes, inferring their biological consequences is challenging. By contrast, the identification of specific genes harboring rare coding variants in whole sequencing studies may provide more direct mechanistic hypotheses about disease etiology.

To evaluate the biological implications of genetic findings, it is common to evaluate whether the implicated risk genes are enriched for expression in particular cell-types or tissues, and to conduct gene-set analyses testing whether a group of genes are enriched in predefined gene-sets based on their biological functions[127]. Note that differences in methodology and power of the genetic studies limit comparisons of gene-set enrichment results across psychiatric disorders.

Expression analyses of GWAS data on schizophrenia[64,135], autism spectrum disorder[125], bipolar disorder[65,136], major depressive disorder[70,137], ADHD[68], and anorexia nervosa[67] have all revealed enrichment of expression in human brain tissue, confirming the importance of brain-expressed genes in the etiology of major psychiatric disorders. In general, the risk genes are globally expressed in the brain, with no major differential association across brain regions, although the dorsolateral prefrontal cortex (Brodmann area 9) consistently shows the strongest enrichment of expression across psychiatric disorders[64,65,68,70,72].

Furthermore, GWAS associations for schizophrenia[64], bipolar disorder[65], depression[70] and ADHD[68] are enriched in genes high-

ly expressed in neurons, with no apparent enrichment in other brain cells such as oligodendrocytes, astrocytes, endothelial cells, microglia or neural stem cells. Using neuronal subtype specific expression data, GWAS analyses on schizophrenia, bipolar disorder and ADHD implicated both excitatory and inhibitory neurons[64,65,68]. For ADHD, GWAS associations were additionally enriched for expression in dopaminergic midbrain neurons. This is consistent with the link between ADHD and deficits in the reward system, motor control and executive functioning, all of which are under dopaminergic control[68].

The recent GWAS associations for schizophrenia were strongly enriched for genes with high expression in excitatory glutamatergic neurons in the cerebral cortex and the hippocampus (pyramidal CA1 and CA3 cells, and granule cells of dentate gyrus), and in cortical inhibitory interneurons[64]. While GWAS associations for autism spectrum disorder were not significantly enriched in any specific cell type[125], which likely reflects the low power of relevant GWAS[66], risk genes for autism spectrum disorder implicated by rare variants are enriched in genes highly expressed in both excitatory and inhibitory neurons in the human cortex[125].

In schizophrenia, well-powered datasets on both common and rare variants have allowed for a more comprehensive mechanistic interrogation, with emerging biological convergence across both ends of the allelic frequency spectrum[64,106,111,138]. Both rare and common variant associations with schizophrenia have strongly implicated genes influencing synaptic organization, differentiation and signaling, at both presynaptic and postsynaptic locations[64,106,139]. One of the gene sets most strongly associated with schizophrenia is the targets of the fragile X mental retardation protein (FMRP)[140-142], a protein that is highly expressed in neurons, which binds mRNAs from multiple genes implicated in synapse development and plasticity[143].

The strongest common variant association with schizophrenia is localized to the major histocompatibility complex (MHC)[135,144,145], a genomic region that contains many genes linked to infection and autoimmunity. A comprehensive analysis demonstrated that part of the MHC association with schizophrenia is driven by structural variation in the gene *C4*, which encodes complement component 4 (C4)[146]. The complement system is part of the innate immune system and also contributes to normal brain development by eliminating immature synapses[147,148]. Schizophrenia risk at *C4* was associated with greater expression of the *C4* isotype *C4A*, which is present at human synapses and neuronal components. In mice, C4 was shown to promote synapse elimination during development. These findings indicate that at least part of the MHC association with schizophrenia may implicate inappropriate synaptic maturation[146]. However, note that the MHC risk locus only represents a minor part of the genetic risk architecture underlying schizophrenia.

Risk genes for schizophrenia implicated by both common and rare variant studies are also linked to biological processes related to excitability, in particular voltage-gated calcium channels, and multiple neurotransmitters[64,106,138]. In a recent WES study[106], two of the ten implicated genes, *GRIA3* and *GRIN2A*, encode receptor subunits involved in glutamatergic neurotransmission. These findings corroborate previous GWAS discoveries[138], providing support for the glutamatergic hypothesis of schizophrenia[149]. An analysis of the effects of schizophrenia-risk variants in neurons derived from human induced pluripotent stem cells revealed a synergistic effect on gene expression and synaptic function[150], emphasizing the importance of studying the combinatorial effects of risk variants to fully understand their biological consequences.

Genes linked to ion channels, neurotransmitter receptors and synaptic proteins have also been implicated in GWAS on bipolar disorder[65,136] and depression[151]. However, since the GWAS discoveries for these and other psychiatric disorders still trail those for schizophrenia, the biological interpretation of these data is less robust.

Risk genes for autism spectrum disorder, most of which are implicated from rare variant studies, are strongly linked to synaptic function as well as chromatin remodeling, which affect the regulation of the expression of multiple other genes, thereby complicating mechanistic interpretation[66,100,102,125,152-154]. An analysis of expression patterns of risk genes in autism spectrum disorder found that risk genes implicated by rare variants were more strongly expressed during fetal development than those implicated by common variants, which displayed relatively higher expression at later developmental stages[125].

Among risk genes shared between schizophrenia, autism spectrum disorder and developmental disorders harboring *de novo* coding variants, a recent study demonstrated that the same classes of mutations were generally involved[155]. This finding suggests that these overlapping genetic signals reflect shared biological mechanisms, further supporting a continuum in the etiology of these disorders, and impairment of neurodevelopment as part of the etiology in schizophrenia[156].

Integrating GWAS and WES data on autism spectrum disorder has revealed insights into the gender differences in risk of this disorder, which is diagnosed three to four times more often in males than in females. Female individuals with the disorder tend to have a higher burden of common and rare genetic variants than their male counterparts, indicating that a higher genetic loading is necessary to result in development of the condition in females, in line with a female protective effect[125,153]. Moreover, among parents of cases with autism spectrum disorder, who did not have the disorder themselves, the mothers had significantly higher polygenic risk for the disorder than the fathers. This supports the notion that females can accumulate more risk before being diagnosed with autism spectrum disorder[157]. Despite known gender differences in the risk for other psychiatric disorders[158], current genetic data have not yet revealed convincing insights that could explain these differences.

Gene-set analyses can also be applied to targets of existing drugs, which may inform pharmacological research and reveal opportunities for repurposing. Drugs supported by genetic evidence appear to have a higher success rate in clinical development[159]. Among 50 novel drugs approved by the US Food and Drug Administration (FDA) in 2021, two-thirds were subsequently shown to have some genetic support, although this approach is

vulnerable to confirmation bias[160].

In the latest GWAS on bipolar disorder, common variant associations were enriched in targets of several classes of pharmacological agents, including mood stabilizers, antipsychotics, antiepileptics, and calcium channel blockers[65]. These findings suggest that existing drugs in bipolar disorder have some biological support based on genetic data, and have motivated efforts to investigate the potential efficacy of calcium channel antagonists in this disorder[161], with lamotrigine being an N-type calcium channel blocker widely used in treatment of bipolar type II disorder. A recent WES study also found enrichment of rare damaging coding variants in calcium channel genes among individuals with bipolar disorder[108].

An analysis of GWAS data on major depressive disorder revealed enrichment of common variant associations in genes encoding proteins targeted by antidepressant medication[137]. Another pharmacological enrichment analysis implicated ten existing drugs, three of which have been linked to depression (riluzole, cyclothiazide and felbamate), and four modulate estrogen (tamoxifen, raloxifene, diethylstilbestrol, and Implanon – an etonogestril implant)[70]. A recent systematic umbrella review of the relationship between serotonin and depression did not find any genetic support for a role of serotonin in depression[162]. However, this conclusion is premature, given that less than 10% of the genetic risk architecture of depression is uncovered (see Figure 2), and even less is known about its biological consequences, and the biological heterogeneity between patients.

The biological interpretation of genetic data is complicated by the fact that genetic associations likely capture different types of causal relationships, at least for highly polygenic complex phenotypes such as psychiatric disorders. The genotype-phenotype associations detected in a GWAS can be decomposed into three main sources: direct genetic effects, indirect genetic effects, and confounding effects[163]. The direct genetic effects represent the causal effects of a genetic variant on a phenotype via biological pathways. The indirect effects represent situations where a genetic variant in an individual affects the phenotype in another individual through the influence on the environment, for example via parental behavior. Parental genetic variants do not need to be transmitted to the offspring to have an indirect genetic effect[164]. Confounding effects include assortative mating or population stratification, which affect the distribution of genetic variants within populations. The presence of confounding and indirect genetic effects will impact analysis of genetic data, as they dilute the genetic signal representing direct causative mechanisms.

Compared to standard population-based GWAS, family-based GWAS are less likely to be affected by confounding and indirect genetic effects. In a recent analysis of family-based and population-based GWAS for 25 phenotypes[165], the GWAS estimates for behavioral phenotypes, including depressive symptoms, were found to be considerably smaller in family-based versus population-based GWAS, while the GWAS estimates were similar for somatic molecular traits such as C-reactive protein and lipids[165]. These findings indicate that a large part of the genetic associations for behavioral phenotypes may represent indirect or confounding effects, warranting more research using large-scale family-based GWAS on psychiatric disorders. It is not yet clear how these different sources of genotype-phenotype association may affect estimates of the polygenicity of a trait.

Another aspect complicating biological interrogation of psychiatric disorders is that multiple potential causal biological pathways may be involved[166]. The clinical heterogeneity among individuals with a given psychiatric disorder is likely mirrored by biological heterogeneity of a similar extent. A case-control GWAS, however, only represents the mean differences in genetic associations between cases and controls. This summary measure may therefore conceal biological differences among potential subgroups of patients, who may have different clinical profiles and respond differently to therapeutic interventions.

Furthermore, the extent to which genetic findings and their biological consequences are generalizable across populations remains to be clarified. This is a pressing issue in human genetics, since most GWAS have been predominantly based on individuals of European descent[167], which is also the case in psychiatric genetics (see Table 1). Genetic studies are often based on one ancestral group to avoid mistaking systematic differences between ancestries for genetic influences underlying a trait. The lack of ancestral diversity also applies to functional genomic datasets, such as tissue-specific gene expression, DNA methylation and chromatin interactions[168,169], which are necessary to reliably interpret genomic data.

The transferability of genetic risk across populations may be affected by differences in allele frequencies, correlation among genetic variants (referred to as the LD structure), variation in the functional impact of a genetic variant, and the overall differences in genetic and environmental contexts. Moreover, the causes, presentation and diagnosis of psychiatric disorders may differ across populations[170]. A recent trans-ancestry GWAS analysis of schizophrenia reported a genetic correlation of 0.98 between two cohorts of East Asian and European descent, indicating that the common variant architecture of the disease is fundamentally the same in these two populations, despite differences in known environmental risk factors such as migration, urbanicity and drug abuse[171]. By contrast, a trans-ancestry GWAS analysis of major depressive disorder reported a genetic correlation of only 0.41 between two cohorts of East Asian and European descent, indicating larger differences in the genetic architecture underlying the disorder in these two populations[172]. These findings suggest that genetic heterogeneity across ancestries may differ across psychiatric diagnoses, further emphasizing the importance of prioritizing greater diversity in psychiatric genetics.

SHARED GENETIC INFLUENCES BETWEEN MENTAL DISORDERS AND WITH OTHER TRAITS AND DISEASES

Clarifying the nature of shared genetic influences between psychiatric disorders and with other traits and diseases has become an important research area in psychiatric genetics. This

research could inform ongoing processes aiming to reconceptualize psychiatric nosology[173,174], increase the understanding of the pervasive comorbidity and shared clinical features across mental disorders[22,23], help disentangle heterogeneity within diagnostic categories and identify subgroups with similar clinical features, and possibly reveal shared etiology with other traits and disorders.

Given the high polygenicity of human traits and disorders and the finite number of genetic variants, it follows that many genetic variants are expected to influence more than one phenotype, a phenomenon termed genetic pleiotropy[175]. Yet, the extent of genetic pleiotropy revealed across human traits and disorders in recent years has probably surpassed the expectations of many[59,79], and it is becoming increasingly clear that the genetic relationship between psychiatric disorders, and between psychiatric disorders and other phenotypes, is more extensive and complex than has been widely recognized[95,138,176].

Genetic influences of psychiatric disorders are shown to overlap with a wide range of brain-related and somatic human traits and disorders, including cognitive traits[86,177-180], neurological disorders[181-186], substance use[187-189], and cardiovascular disease and risk factors[190-193]. Among the many cross-trait genetic associations, it is important to emphasize that psychiatric disorders are also genetically linked to positive traits, which we believe is an important message to communicate to patients and the public. For example, risk for autism spectrum disorder is genetically correlated with higher educational attainment[194] and better cognitive performance[86], while risk for bipolar disorder and schizophrenia is genetically correlated with higher levels of the personality trait openness to experience[195] and creativity[196].

Both common and rare genetic variants exert genetic pleiotropy, but the phenomenon is more widely documented for common variants, due to the high number of well-powered GWAS reporting common variant associations[60]. In a comprehensive analysis of genetic pleiotropy across more than four thousand GWAS, 90% of the genomic loci were associated with more than one biological domain (e.g., a locus associated with both a psychiatric and an immunological phenotype), and an even greater proportion of loci had multi-trait associations within a biological domain (e.g., a locus influencing two or more psychiatric disorders)[79]. Since a locus may contain several genes and even more SNPs, multidomain associations at the gene level (63%) and SNP level (31%) were less abundant[79]. However, the extent of genetic overlap is higher when SNPs not yet identified at the genome-wide significance level are also included[78,192,197].

The assembly of well-powered GWAS on psychiatric disorders (see Table 1) has enabled systematic comparisons of their unique and shared genetic architectures. Even though most common genetic variants for complex human phenotypes remain to be identified[61], genetic overlap between two phenotypes can be investigated at the genome-wide level by including the effects of all or a subset of SNPs. The most commonly applied tools for this purpose are polygenic risk scores (PRS)[198,199] and the bivariate extension of LD score regression[200].

In line with previous findings of shared genetic risk between psychiatric disorders[201], an analysis of GWAS data from 25 common brain disorders demonstrated substantial pairwise positive genetic correlations across psychiatric disorders, which exceeded that which could be reasonably explained by potential diagnostic misclassification[184]. In Figure 3, we provide an updated overview of pairwise genetic correlations between major psychiatric disorders using the most recent GWAS available.

In comparison, there are markedly fewer and smaller pairwise genetic correlations among neurological disorders[184], and between neurological and psychiatric disorders, although there are a few exceptions[184,185,202]. This dissimilar pattern of pairwise genetic correlations among neurological and psychiatric disorders may indicate that the former represent more distinct genetic entities than the latter[184]. This is in line with the notion that neurological diagnostic categories have a stronger biological foundation. By contrast, genetic risk for psychiatric disorders evidently transcends diagnostic domains, and these disorders are more genetically interconnected. As observed for common genetic variants, rare CNVs and protein-truncating variants also show a high degree of pleiotropy across the whole group of psychiatric disorders[98,203] and with other brain-related traits such as epilepsy, developmental disorders and cognitive ability[204,205].

The emerging genetic data may be considered to be at odds with the current diagnostic classification systems[17,18], in which psychiatric disorders are considered categorically distinct from one another[206]. The genetic findings may thus be considered to support efforts to reconceptualize psychiatric nosology in a more dimensional framework[206,207], such as the proposed Hierarchical Taxonomy of Psychopathology (HiTOP)[173] or Research Domain Criteria (RDoC)[174].

Genetic risk for psychiatric disorders also overlaps with genetic variation in behavioral traits[95,208], such as the Big Five personality traits[195,209], general intelligence[86], educational attainment[210], subjective well-being[211], sleep patterns[87,212], and mental health profiles in healthy individuals[213], indicating that genetic risk for mental illness is not categorically distinct from normality[206].

A cross-disorder GWAS analysis of eight psychiatric disorders using factor analysis and genomic structural equation modelling[214] indicated broader genetic domains that may underlie a higher-order structure of psychopathology[215]. Using the same analytical approach, a recent GWAS analysis of 11 psychiatric disorders found evidence of four highly correlated groups of disorders[216]. The first group was characterized by compulsive behaviors (anorexia nervosa, OCD and Tourette's syndrome), the second group by internalizing symptoms (anxiety disorder and major depressive disorder), the third group by psychotic features (schizophrenia and bipolar disorder), and the fourth group by neurodevelopmental features (ADHD and autism spectrum disorder), surprisingly also including PTSD and problematic alcohol use[216]. Interestingly, the cross-disorder GWAS analysis did not find clear evidence that an underlying generalized liability to develop psychopathology (the p factor[217]) could adequately explain shared variance across psychiatric disorders[216].

Cross-disorder PRS analyses present a similar picture. In line with a dimensional model of psychopathology, patients with bi-

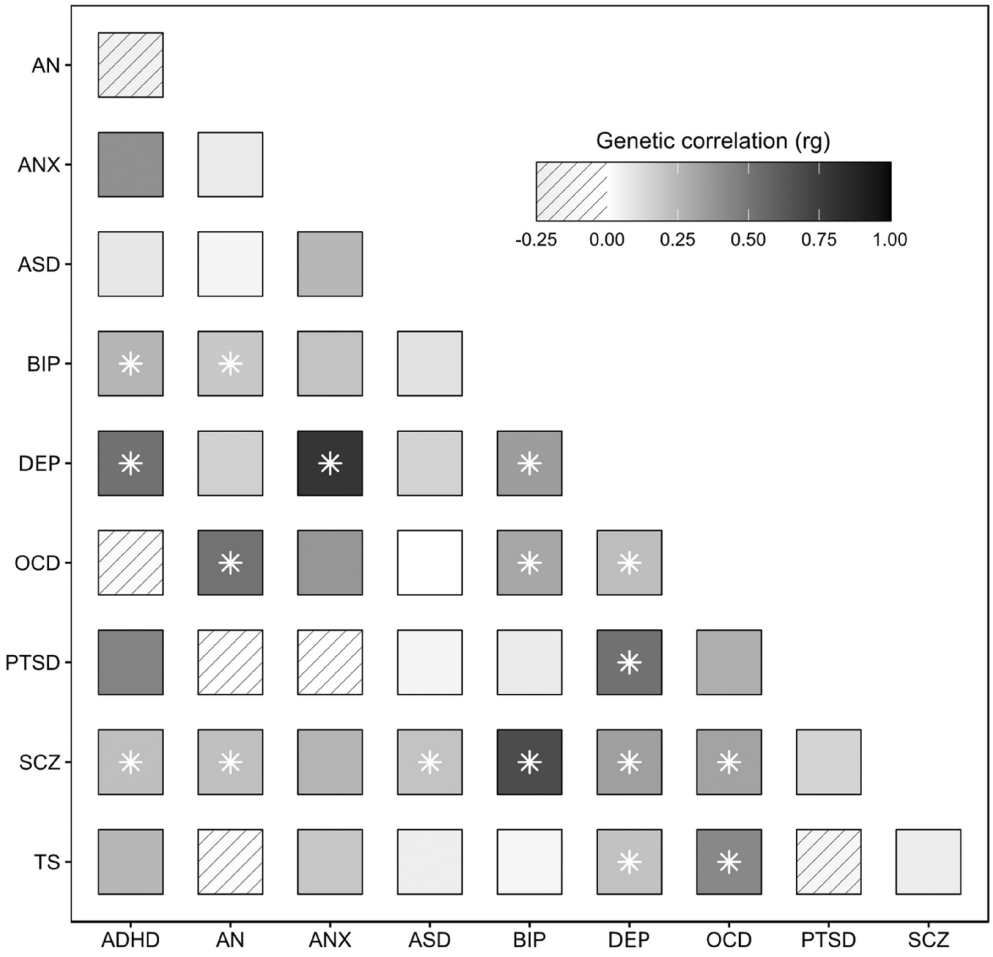

Figure 3 Pairwise genetic correlations between major psychiatric disorders estimated using LD score regression. Significant genetic correlations indicated by an asterisk. ADHD – attention-deficit/hyperactivity disorder, AN – anorexia nervosa, ANX – anxiety, ASD – autism spectrum disorder, BIP – bipolar disorder, DEP – depression, OCD – obsessive-compulsive disorder, PTSD – post-traumatic stress disorder, SCZ – schizophrenia, TS – Tourette's syndrome.

polar disorder with a history of psychotic symptoms had a higher schizophrenia PRS compared to those without such a history, which was not driven by the presence of cases with schizoaffective subtype[218]. Similarly, a history of manic symptoms in schizophrenia has been significantly associated with bipolar disorder PRS[218,219], indicating that genetic risk for mental illness influences clinical subphenotypes across diagnostic categories.

There is also increasing evidence of genetic heterogeneity among subtypes of mental disorders. For example, the genetic risk underlying childhood ADHD and ADHD persistent in adults is partially distinct, with a genetic correlation of 0.81[220]. A subsequent genetic dissection of three ADHD subgroups defined by the age at first diagnosis (childhood, adult or persistent ADHD) indicated further genetic differences, with the lowest pairwise genetic correlation (rg=0.65) between childhood and late-diagnosed ADHD[221]. The ADHD subgroups also displayed different PRS associations with related traits and disorders, with late-onset ADHD generally having the strongest associations, for example with higher risk of depression and insomnia, while childhood ADHD was most strongly associated with autism spectrum disorder[221].

Analysis of bipolar disorder has also revealed genetic heterogeneity between subtypes, with a genetic correlation of 0.89 between type I and II[136]. In line with their clinical profiles, bipolar type II disorder is more genetically correlated with major depression (rg=0.69) than with schizophrenia (rg=0.51), while bipolar type I disorder is more genetically correlated with schizophrenia (rg=0.71) than with major depression (rg=0.30)[136]. These findings clearly indicate that mood and psychotic disorders exist on a continuum, both phenotypically[222] and genetically.

Evaluating patterns of genetic overlap between psychiatric disorders and other traits has also provided significant insights. This is particularly relevant for bipolar disorder and schizophrenia, which may in some cases be difficult to differentiate diagnostically. While both disorders are associated with cognitive impairment, the cognitive deficits are generally more pronounced in individuals with schizophrenia[223]. In line with these phenotypic associations, genetic risk of both disorders extensively overlaps with cognitive function, but in a different manner, where most schizophrenia risk

variants are associated with poorer cognitive performance, while there is a balanced mix of bipolar disorder risk variants associated with worse or better cognitive performance[178]. Hence, leveraging genetic data on related traits may help distinguish the genetic architectures of highly correlated psychiatric disorders, and point to differences in their etiologies.

Additional work has indicated that the genetic overlap between psychiatric disorders is even more extensive than expressed by the pairwise genetic correlations[78,138,176214,215], as depicted in Figure 4. A comprehensive analysis of the unique and shared common variant architectures between psychiatric disorders and between psychiatric disorders and behavioral phenotypes indicated substantial genetic overlap, with only a minority of trait-specific variants, despite differences in genetic correlation[95].

Widespread genetic overlap despite divergent genetic correlations indicates that psychiatric disorders are predominantly influenced by a set of highly pleiotropic genetic variants which impact the risk of each disorder to a different degree and, in some cases, in different directions[138]. This insight is consistent with an integrated conceptualization of the neurobiology of psychiatric disorders and related traits, in which multiple, overlapping neurobiological mechanisms and systems are implicated in the development of both mental disorders and normative mental traits[95]. However, the extent to which indirect and direct genetic effects differently contribute to pleiotropy across highly

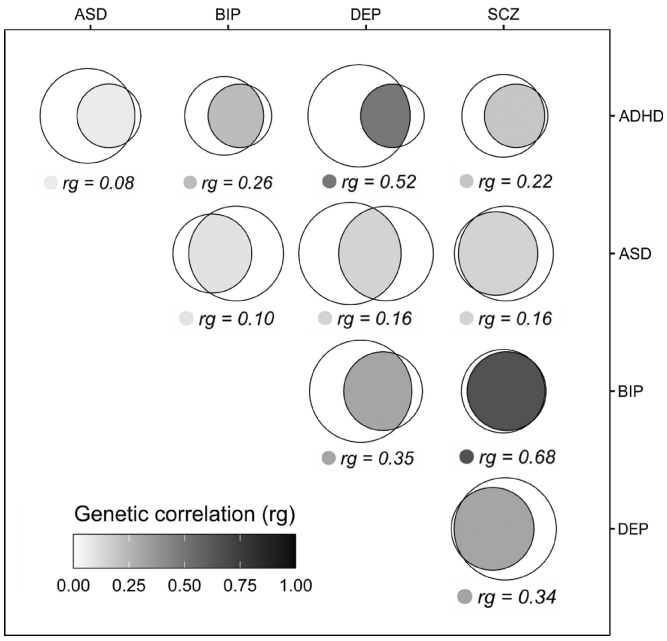

Figure 4 Extensive overlap in common genetic variants between mental disorders beyond genetic correlation. The fraction of unique and shared genetic architecture between pairs of the five psychiatric disorders is estimated using MiXeR[78]. The genetic correlations are estimated using LD score regression[200]. The disorders represented by the left circles of the Venn diagrams are listed in the horizontal axis, and right circles are represented by disorders listed in the vertical axis. ADHD – attention-deficit/hyperactivity disorder, ASD – autism spectrum disorder, BIP – bipolar disorder, DEP – depression, SCZ – schizophrenia.

polygenic phenotypes such as psychiatric disorders is currently unknown, warranting more data from family-based studies.

The recent accumulation of large publicly available genotyped neuroimaging samples through international initiatives such as ENIGMA[224] and population studies such as UK Biobank[225] has provided new opportunities to study the shared genetic foundations of human brain structure and psychiatric disorders. Global measures of brain structure, such as cortical thickness and surface area, have been shown to be highly heritable, with SNP-heritability estimates ranging from 25 to 35%[226]. However, they have been found to be 4-5 times less polygenic than mental disorders, indicating fundamental differences in their genetic architectures[227].

In particular, the genetic relationship between schizophrenia and brain structural phenotypes has been extensively studied[227-235], owing to the well-powered GWAS data on that disorder. Despite well-established findings of subtle brain structural abnormalities in schizophrenia[236-238], the genetic correlations between neuroimaging measures and schizophrenia have been absent or low[228,229]. Yet, despite a lack of genetic correlation, cortical thickness and surface area are predicted to share almost all their common genetic variants with schizophrenia, while a large majority of genetic variants associated with schizophrenia are not associated with cortical structure[227]. The difference in the proportions of overlapping genetic variants is explained by the large difference in polygenicity of the brain imaging phenotypes and schizophrenia[227]. Further, the apparent contradiction of substantial genetic overlap despite minimal genetic correlations is likely due to mixed directions of effect among the shared variants, which cancel out the overall genetic correlation[138]. Indeed, multiple specific genetic variants have been discovered in recent years which are shared between schizophrenia and various brain morphology measures[230], including cortical thickness and surface area[227], volume of subcortical regions[231-233], intracranial volume[231], cerebellar volume[234], and brainstem structures[235]. Taken together, the emerging genetic data indicate a complex genetic relationship between brain structural measures and schizophrenia, and it remains unclear to what extent imaging phenotypes can serve as endophenotypes that capture underlying mechanisms with greater biological specificity.

An important limitation of most studies of genetic overlap is the ambiguity regarding the direction of causality and whether the detected overlap implies shared biological mechanisms. A given shared genetic association may reflect so-called "horizontal" or biological pleiotropy, in which a variant influences two phenotypes through independent molecular mechanisms; "vertical" or mediated pleiotropy, in which a variant influences a trait, and this trait causally affects another trait; or "spurious" pleiotropy, in which a variant is falsely assumed to influence two traits, for example due to statistical association between two nearby variants in strong LD with each other[239].

Mendelian randomization attempts to directly address the question of causality by testing for evidence of a causal relationship between the genetic factors associated with a given "exposure" and a given "outcome" (vertical pleiotropy). For example, Mendelian randomization has provided several intriguing find-

ings regarding the link between inflammation and the etiology of psychiatric disorders. Genetically determined level of C-reactive protein was shown to have a potentially protective effect on schizophrenia risk[240]. This finding was replicated in a recent analysis[241] using the most recent schizophrenia GWAS[64], although a significant causal relationship was only present when controlling for body mass index and circulating interleukin 6 (IL-6) and its receptor[241].

In another Mendelian randomization study, IL-6 itself has been shown to exhibit a potentially causal association with grey matter volume across multiple cortical regions, and to interact with a network of co-expressed genes in the medial temporal gyrus which were found to be differentially expressed in schizophrenia, autism spectrum disorder and epilepsy[242]. IL-6 receptor levels have also been implicated in the risk for depression[243] and suicidality[244], although less is known about putative causal relationships with bipolar disorder.

Interestingly, in Mendelian randomization studies focusing on immune disorders rather than biomarkers, several psychiatric disorders where found to have a causal effect on immune disorders, rather than the other direction, including a causative effect of major depressive disorder on asthma and of schizophrenia on ulcerative colitis[245]. Nonetheless, while these findings have contributed to the growing evidence base for a possible causal association between inflammatory phenotypes and psychiatric disorders, Mendelian randomization is still based on statistical inference, and it is important to control for the extensive "horizontal" pleiotropy observed between mental traits and disorders. Thus, the validity of Mendelian randomization findings require further investigations via *in vitro*, *in vivo*, and interventional studies.

The assembly of large-scale biobanks harboring rich phenotypic data can be leveraged to discover connections between genetic markers and traits, for example using the phenome-wide association study (PheWAS) approach to systematically investigate trait-associations with a given PRS[246]. A PheWAS study investigating the link between schizophrenia PRS and electronic health record data in 106,160 patients across four large US health care systems in the PsycheMERGE Network reported that schizophrenia PRS was not only associated with psychiatric phenotypes such as diagnosis of schizophrenia and substance use, but with several non-psychiatric phenotypes, including a negative association with obesity[247]. The inverse genetic association between schizophrenia risk and obesity has been confirmed by other genetic studies[193], indicating that the increased body mass index observed in schizophrenia patients is likely due to non-genetic factors such as antipsychotic medication.

Another PheWAS study on 325,992 participants in the UK Biobank reported significant associations between schizophrenia PRS and multiple psychiatric and non-psychiatric conditions and measures, including poorer overall health ratings, more hospital inpatient diagnoses, and more specific disorders (musculoskeletal, respiratory and digestive diseases, varicose veins, pituitary hyperfunction, and peripheral nerve disorders)[248]. Although some of these PRS trait-associations may be consequences of having schizophrenia or related psychiatric disorders, the studies indicate that the genetic risk for schizophrenia also affects a wide range of somatic conditions.

Finally, a similar PheWAS study of 382,452 patients in the PsycheMERGE Network investigated the relationship between depression PRS and 315 clinical laboratory measurements[249]. A replicable yet modest association was found between higher polygenic burden of depression risk variants and increased levels of white blood cells, even after controlling for a diagnosis of depression and anxiety. In line with a neuroinflammation model[250], a potential causal link between white blood cells and depression was supported by mediation and Mendelian randomization analyses, indicating that higher genetic risk underlying depression may activate the immune system, possibly contributing to the risk of developing the disorder[249].

CLINICAL APPLICATIONS

Despite significant progress over the last decade in our understanding of the genetic foundations of psychiatric disorders, clinical translation remains conspicuous by its absence. Nevertheless, genetic-based prediction and stratification offers a promising avenue towards improved patient outcomes in the coming decades[251]. Chip-based genotyping is relatively affordable, while the price for whole-genome sequencing continues to fall[252]. What's more, genetic testing only needs to be performed once in a person's lifetime, and genotyping data can be used on multiple occasions for multiple different purposes. However, several major challenges need to be overcome before this translates into a clinically viable tool which benefits patients, including improving predictive accuracy, enabling discrimination between diagnostic categories or clinically actionable decisions, ensuring equal predictive performance across ancestral groups, and guarding against significant ethical concerns.

The main focus of research into genetic-based prediction has centered around PRS. This uses existing genetic data to construct an individualized risk score for a given trait or disorder, calculated as the sum of pre-defined risk alleles weighted according to each allele's effect on the phenotype, typically estimated by a GWAS[253]. The accumulation of massive case-control samples alongside PRS-method improvement has recently led to the development of PRS-based tools with clinically meaningful predictive accuracy in several common medical conditions[254], including cardiovascular disease[255,256], type 1 diabetes mellitus[257] and cancers[256,258]. However, even considering the improved predictive performance of the latest PRS tools, current PRSs for major psychiatric disorders are far from achieving equivalent levels of prediction[259,260].

For schizophrenia, which possesses the most well-powered GWAS to date, the best performing PRS method explained just 8.5% of the variance in liability for the disease, falling to 7.3% when non-European ancestry cohorts were included[64]. The insufficient predictive accuracy of the schizophrenia PRS is further demonstrated by an area under the receiver operating characteristic curve (AUROC) of 0.72[64], while an AUROC above 0.8 is

considered to indicate good discriminative ability[253]. Other psychiatric disorders lag even further behind, with the AUROC for major depressive disorder and bipolar disorder PRS being 0.57 and 0.65, respectively[65,137]. At the current levels of explained variance, this means that most individuals in the top PRS centiles for a given mental disorder will not develop that disorder and the majority of people who do develop mental disorders have PRS centiles closer to the median[259]. As a result, current PRSs for psychiatric disorders show poor potential for screening purposes in the general population, and do not yet have a role in genetic counselling. PRS has currently a larger potential for screening of some common medical conditions[254,256], as exemplified by the MyGeneRank application[261].

Since the predictive accuracy of PRS is also dependent on the prevalence of the disorder in the sample tested, the utility of psychiatric PRSs will vary depending on the context in which they are applied[262]. Although psychiatric disorder PRSs are far from being able to accurately predict a given disorder in the general population[259], they may provide greater clinical utility if used in clinical populations for which the pre-test probability that an individual will experience a mental disorder is higher. For example, PRS may be useful to predict risk of developing psychosis in individuals who carry large-effect rare variants, such as carriers of 22q11.2 deletion. Approximately 20-25% of 22q11.2 deletion carriers develop schizophrenia[263,264]. Among carriers of 22q11.2 deletion, schizophrenia prevalence was 9% vs. 33% in the lowest and highest deciles of the schizophrenia PRS, respectively[124], indicating potential utility for informing clinical decision-making in the near future for this patient group. Among individuals at clinical high risk of developing psychosis followed over a 2-year period, addition of schizophrenia PRS to an existing calculator slightly improved prediction of psychosis[265]. Use of disorder-specific PRS at this stage may be useful for informing decisions relating to the level of follow-up required or whether or not to initiate psychotropic medication. This may also be relevant for other patient groups, such as those presenting with depressive symptoms, for whom the clinical trajectory is highly variable and is associated with differences in genetic risk for major depressive disorder[266].

There is currently only limited evidence to support the hypothesis that disorder-specific PRSs are associated with treatment response for either depression or psychosis[267,268]. Alternatively, it may be possible to develop PRSs tailored for specific treatment decisions. High rates of non-response among patients taking both antidepressant and antipsychotic medications mean that tools which effectively predict treatment response could have a significant impact on patient outcomes[269,270]. For example, the early identification of patients with treatment-resistant schizophrenia requiring clozapine is a prime candidate for a treatment-focused PRS. Approximately 30-40% of individuals with schizophrenia do not respond to two first-line antipsychotics, but half of this group respond to clozapine[271]. A case-case GWAS of treatment responding vs. resistant patients found that treatment resistance was minimally but detectably heritable (h^2_{SNP}=1-4%) and that a PRS derived from this GWAS was weakly predictive of clozapine use in an independent sample[82].

Genetic prediction may also be helpful for identifying individuals who do not respond to pharmacological treatment whatsoever or are likely to develop specific side effects related to psychotropic medication[272]. In the coming years, large-scale, genotyped prescription registries such as FinnGen[273], in addition to deeply phenotyped clinical samples, will offer new opportunities to investigate the genetics of non-response and adverse drug reactions.

As the predictive ability of PRS largely depends on the power of the genetic study it is derived from, the performance of PRS is likely to improve in the coming years due to significant increases in sample sizes, better phenotyping procedures and further methodological refinements[96,254,260]. However, PRS performs poorly when applied to admixed individuals or individuals of other ancestries than the cohort the PRS was initially derived from[55]. Since most GWAS are based on European individuals, the poor cross-ancestry performance of PRS represents a major challenge to ensure equitable health benefits of its potential clinical implementation.

The high degree of genetic and symptomatic overlap across diagnostic categories and the lack of "gold standard" diagnostic tests also represent a unique challenge within psychiatry as opposed to other medical specialties, for which screening is already a part of routine clinical pathways. Given that the choice of psychotropic medication is often driven by diagnosis, a lack of discriminatory ability across disorder-specific PRSs may limit their clinical utility. This feeds into a wider question about the validity of the diagnostic categories themselves. Psychiatric disorders are highly heterogenous and overlapping, both clinically and neurobiologically, which may limit the predictive capability of PRSs based on the current diagnostic criteria[274,275]. This represents somewhat of a "catch-22" scenario, since PRS performance is dependent on statistical power and the largest samples to date are based on the prevailing diagnostic system, with limited phenotypic data available for large proportions of the subcohorts comprising these large-scale GWAS[206]. With increasing recognition of the need to prioritize more deeply phenotyped samples, this is likely to shift in the coming years.

It is also possible that the genetic overlap across diagnostic categories could be leveraged to improve prediction of individuals with psychiatric disorder compared to healthy controls, even if this is at the cost of discriminating between different diagnoses. A recent study combined multiple disorder-specific PRSs to improve prediction of mood disorders, anxiety, ADHD, autism spectrum disorder and substance use disorders[276]. This raises the possibility that distinct types of PRS may be applied in the future depending on the clinical question, either to maximize prediction of psychiatric disorder as opposed to its absence, or to maximize discrimination across diagnostic categories, alternative subphenotypes, or treatment options.

While psychiatric PRS is still some way from being applied clinically, advances in non-psychiatric PRS may provide more immediate benefits for individuals with psychiatric disorders. Cardiovascular disease and its metabolic risk factors are significantly more prevalent among psychiatric patients and are the single larg-

est cause of death in these patients[277]. A study in the UK Biobank showed that applying a cardiovascular disease PRS in addition to standard risk prediction for people at intermediate risk could prevent 7% more cardiovascular disease events than the standard screening approach[278]. So, while it is feasible to incorporate PRS for cardiovascular disease into routine clinical practice for the general population, this may provide particular benefit for psychiatric patients[278].

Despite the fact that PRSs are currently not deemed to be clinically useful, patients can already acquire their own PRS profile themselves at relatively low cost through direct-to-consumer genotyping companies. Although these companies do not routinely offer PRS for psychiatric disorders, individuals can download their own raw genotypes and use complementary websites to compute PRS for additional phenotypes of their choice. While this may help to democratize access to health information and increase patients' ability to take ownership for their health, these services are variably regulated across countries[279], and the information provided to help consumers accurately interpret their results varies greatly[280]. Given the common misconception that genetic testing is deterministic, this could leave consumers at risk of misinterpreting their results, which may lead to harmful outcomes.

Moreover, interpreting PRS results requires an understanding of the difference between relative risk and absolute risk, which may not be intuitive. For example, in the latest schizophrenia GWAS[64], being in the top PRS centile was only associated with an odds ratio of 5.6 relative to the rest of the sample. Hence, an individual in the top PRS centile for schizophrenia without any other risk factors is more likely to not develop the disease than get the disorder, due to the low lifetime risk of schizophrenia.

A recent news article described a particularly concerning example of consumer use of PRS, in which a couple used a company called Genomic Prediction Inc. to perform PRS-based screening of embryos derived by *in vitro* fertilization[281]. The couple then used a third-party service to compute PRS for schizophrenia and intelligence and selected their embryo based on these scores. Not only does this raise major ethical concerns given the association with eugenics and ableism, but the fact that the PRS for schizophrenia is associated with positive traits such as increased openness to new experiences[195] and creativity[196] emphasizes that selection based on tools with limited predictive ability for traits which are still poorly understood and subject to stigma and discrimination could result in unintended and unwanted consequences[282-284]. Researchers affiliated with Genomic Prediction Inc. have since constructed a polygenic health index by combining PRS for 20 impactful disease conditions, including schizophrenia[285].

Overall, the rapid methodological developments, increasing availability, and public and clinical interest in genetic prediction tools highlight the need for greater oversight and regulation in this emerging new interface between science, commerce, and the rights of the individual. Given the impact on medicine, implementation of PRS at different levels (e.g., embryo selection, risk screening in the population, informing clinical decision-making) requires a broader debate in society and the general public.

CHALLENGES AND OPPORTUNITIES FOR PROGRESS AND FUTURE IMPACT

Despite the substantial progress in the discovery of genetic variants influencing risk of mental illness in the last decade, psychiatric genetics is still in its early stages, and the genetic findings have not yet been translated into better mental health care. Most genetic risk variants affecting major psychiatric disorders remain to be uncovered (see Figure 2), and several psychiatric disorders still lack sufficiently powered genetic data. To maintain progress in the field, it is necessary to continue assembling large-scale samples of people with psychiatric disorders, including measures of the progression and severity of illness and treatment response. To this end, international cooperation is the best way forward[224,286], with support from national cohorts such as UK Biobank[225], FinnGen[273], iPSYCH[287] and deCODE[288].

It is increasingly recognized that integrated analysis of the full range of genetic variation[125,289] is necessary to provide a comprehensive understanding of how genetic variants influence risk of illness and underlie different clinical profiles, warranting greater use of sequencing technologies. Moreover, the present genetic findings have disproportionally been based on individuals of European descent, and are only partially transferrable to other ancestral groups, due to differences in genetic and environmental contexts[168,169,172], resulting in poorer performance of genomic prediction tools[55,290,291]. To ensure that the expected health benefits from the developments in human genetics are equitable, it is imperative to prioritize ancestral diversity of both genomic and functional genomic data resources in the coming years, which requires a concerted global effort[167-169]. New initiatives have been established to improve recruitment of diverse samples[167,292,293] and to develop better trans-ancestry prediction methods, with promising results in several complex human disorders[294-296].

Psychiatric disorders are multifactorial. The impact of individual genetic risk depends on the psychosocial setting of the individual, and this must be taken into account to ensure further progress in the field. To obtain a more complete understanding of the underlying causes of psychiatric disorders and account for the substantial individual variation, deeper phenotyping and incorporation of demographic and environmental data is needed. It is, therefore, necessary to go beyond unidimensional case-control studies based on diagnostic categories and adopt a multi-modal analytical framework, that incorporates clinical characteristics, genetic information, blood biomarkers, neuroimaging measures, electronic health record data, lifestyle factors, demographic data and environmental factors in a systematic manner. This will be expensive and requires extensive data harmonization, which again calls for coordinated, international collaborations[297].

Multi-modal integration is also likely to offer the best route to clinical utility for genomic precision medicine approaches[259]. Since most current PRSs are derived from common genetic variants, which explain relatively small proportions of the total

variance in liability for a given disorder, the predictive capacity of PRSs will be inherently limited without the integration of other sources of information. The large number of genetic variants affecting complex human phenotypes in a highly unspecific manner[79,138] emphasizes the need for application of frameworks for quantitative analysis of big data[61,78,298-301]. Building on the ever-increasing amount of psychiatric genetic data, it is possible to develop mathematical modeling approaches[297] that can leverage multidimensional, longitudinal and multimodal data, which may increase etiological insights and set out the roadmap towards precision medicine approaches in psychiatry[251].

In contrast to many other human disorders, psychiatric disorders typically emerge during formative years of childhood, adolescence and early adulthood[302], and they often persist throughout life. However, most of the large-scale health cohorts in the world – such as the UK Biobank[225], the Rotterdam study[303], and the Framingham Heart Study[304] – have focused on cardiovascular disease and chronic illnesses that affect older people, recruiting participants from middle age (from 45-50 years old), several decades after most psychiatric disorders have emerged. Thus, it has become increasingly apparent that birth cohorts with longitudinal follow-up assessments are required to provide insights into the etiology of psychiatric disorders and to facilitate prospective studies of the premorbid phase of these disorders. While there are some long-standing birth cohorts with approximately 15,000 participants (e.g., Avon Longitudinal Study of Parents and Children[305]), larger samples are needed. To the best of our knowledge, there are currently only four large birth cohorts, the Norwegian Mother and Child study[306], the Danish National Birth Cohort[307], the Jiaxing Birth Cohort[308], and the China Birth Cohort[309], with more than 100,000 children in each. Longitudinal samples, covering the sensitive periods of childhood and adolescence, may allow investigations of time of onset, disease trajectories, as well as the interplay between genetic variants and environmental and sociodemographic factors[310]. Here, the large Nordic and Chinese lifespan samples with genetics and real-world data from registries and hospital records will be valuable. Such samples can be used to investigate environmental stressors – e.g., the effect of COVID pandemic[311] – and to study gene-environmental interplay at sensitive periods during development.

The pace of research on human genetics will accelerate over the next decade, and eventually lead to clinical implementation of genetics in more areas of health care, beyond current applications such as neonatal screening, tumor sequencing and diagnostics of rare Mendelian diseases[35,251]. The public interest in the field will likely increase in parallel with the incremental genetic discoveries, with an increased demand for regulation of services using individual genetic data. Although it is still unclear how human genetics may be implemented in mental health care, it is important that the new knowledge about psychiatric genetics becomes an integral part of the training of health care professionals in psychiatry, which is currently not the case in many countries[312], to enable clinicians to reliably return genetic findings to patients and their relatives.

CONCLUSIONS

In the past decade, we have witnessed a series of breakthroughs in psychiatric genetics, driven by progressively larger samples and more advanced technologies and analytical methods, providing new insights into the genetic etiology of psychiatric disorders. It is now clear that thousands of common variants with small effects, as well as rare genetic variants with larger effects, collectively influence the risk of psychiatric disorders. A large proportion of these genetic risk variants influence multiple psychiatric disorders, as well as other behavioral and somatic traits and disorders, indicating a shared genetic basis. However, the biological consequences of these genetic risk variants are still poorly understood.

Psychiatric genetics is still in its early stages, but holds promise of improving mental health care, in particular through refinement of the diagnostic classification system, discovery of novel therapeutic targets and biomarkers, and paving the way for precision psychiatry.

ACKNOWLEDGEMENTS

The authors thank I.E. Sønderby for careful reading and constructive criticism of the manuscript. They are supported by grants from KG Jebsen Stiftelsen (1R01MH124839-01), European Union's Horizon 2020 (847776 and 964874), Research Council of Norway (324499, 324252, 300309, 273291, 223273), and South-East Norway Health Authority (2022-073).

REFERENCES

1. GBD 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Psychiatry 2022;9:137-50.
2. Walker ER, McGee RE, Druss BG. Mortality in mental disorders and global disease burden implications: a systematic review and meta-analysis. JAMA Psychiatry 2015;72:334-41.
3. Dohrenwend BP, Levav I, Shrout PE et al. Socioeconomic status and psychiatric disorders: the causation-selection issue. Science 1992;255:946-52.
4. Knapp M, Wong G. Economics and mental health: the current scenario. World Psychiatry 2020;19:3-14.
5. Corrigan PW, Watson AC. Understanding the impact of stigma on people with mental illness. World Psychiatry 2002;1:16-20.
6. Kessler RC. The epidemiology of dual diagnosis. Biol Psychiatry 2004;56:730-7.
7. De Hert M, Correll CU, Bobes J et al. Physical illness in patients with severe mental disorders. I. Prevalence, impact of medications and disparities in health care. World Psychiatry 2011;10:52-77.
8. Erlangsen A, Andersen PK, Toender A et al. Cause-specific life-years lost in people with mental disorders: a nationwide, register-based cohort study. Lancet Psychiatry 2017;4:937-45.
9. Correll CU, Solmi M, Veronese N et al. Prevalence, incidence and mortality from cardiovascular disease in patients with pooled and specific severe mental illness: a large-scale meta-analysis of 3,211,768 patients and 113,383,368 controls. World Psychiatry 2017;16:163-80.
10. Plana-Ripoll O, Pedersen CB, Agerbo E et al. A comprehensive analysis of mortality-related health metrics associated with mental disorders: a nationwide, register-based cohort study. Lancet 2019;394:1827-35.
11. Chesney E, Goodwin GM, Fazel S. Risks of all-cause and suicide mortality in mental disorders: a meta-review. World Psychiatry 2014;13:153-60.
12. World Health Organization. World mental health report: transforming mental health for all. Geneva: World Health Organization, 2022.
13. Leichsenring F, Steinert C, Rabung S et al. The efficacy of psychotherapies and pharmacotherapies for mental disorders in adults: an umbrella review and meta-analytic evaluation of recent meta-analyses. World Psychiatry 2022;21:133-45.

14. Espinoza RT, Kellner CH. Electroconvulsive therapy. N Engl J Med 2022;386: 667-72.
15. Linden M, Schermuly-Haupt ML. Definition, assessment and rate of psychotherapy side effects. World Psychiatry 2014;13:306-9.
16. Correll CU, Detraux J, De Lepeleire J et al. Effects of antipsychotics, antidepressants and mood stabilizers on risk for physical diseases in people with schizophrenia, depression and bipolar disorder. World Psychiatry 2015; 14:119-36.
17. World Health Organization. International statistical classification of diseases and related health problems, 11th revision. www.who.int.
18. American Psychiatric Association. Diagnostic and statistical manual of mental disorders, 5th ed. Arlington: American Psychiatric Publishing, 2013.
19. Kahneman D, Sibony O, Sunstein CR. Noise: a flaw in human judgment. Glasgow: Collins, 2021.
20. Clark LA, Cuthbert B, Lewis-Fernandez R et al. Three approaches to understanding and classifying mental disorder: ICD-11, DSM-5, and the National Institute of Mental Health's Research Domain Criteria (RDoC). Psychol Sci Public Interest 2017;18:72-145.
21. Kendell R, Jablensky A. Distinguishing between the validity and utility of psychiatric diagnoses. Am J Psychiatry 2003;160:4-12.
22. Plana-Ripoll O, Pedersen CB, Holtz Y et al. Exploring comorbidity within mental disorders among a Danish national population. JAMA Psychiatry 2019; 76:259-70.
23. Barr PB, Bigdeli TB, Meyers JL. Prevalence, comorbidity, and sociodemographic correlates of psychiatric disorders reported in the All of Us research program. JAMA Psychiatry 2022;79:622-8.
24. Galatzer-Levy IR, Bryant RA. 636,120 ways to have posttraumatic stress disorder. Perspect Psychol Sci 2013;8:651-662.
25. Anonymous. Of mice and mental health. Lancet Psychiatry 2019;6:877.
26. Tricklebank MD, Robbins TW, Simmons C et al. Time to re-engage psychiatric drug discovery by strengthening confidence in preclinical psychopharmacology. Psychopharmacology 2021;238:1417-36.
27. Krystal JH, Abdallah CG, Sanacora G et al. Ketamine: a paradigm shift for depression research and treatment. Neuron 2019;101:774-8.
28. Daws RE, Timmermann C, Giribaldi B et al. Increased global integration in the brain after psilocybin therapy for depression. Nat Med 2022;28:844-51.
29. Polderman TJ, Benyamin B, de Leeuw CA et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nat Genet 2015;47:702-9.
30. Duncan LE, Ostacher M, Ballon J. How genome-wide association studies (GWAS) made traditional candidate gene studies obsolete. Neuropsychopharmacology 2019;44:1518-23.
31. Border R, Johnson EC, Evans LM et al. No support for historical candidate gene or candidate gene-by-interaction hypotheses for major depression across multiple large samples. Am J Psychiatry 2019;176:376-87.
32. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature 2004;431:931-45.
33. International HapMap Consortium. The International HapMap Project. Nature 2003;426:789-96.
34. 1000 Genomes Project Consortium, Abecasis GR, Altshuler D et al. A map of human genome variation from population-scale sequencing. Nature 2010; 467:1061-73.
35. Claussnitzer M, Cho JH, Collins R et al. A brief history of human disease genetics. Nature 2020;577:179-89.
36. Sullivan PF, Geschwind DH. Defining the genetic, genomic, cellular, and diagnostic architectures of psychiatric disorders. Cell 2019;177:162-83.
37. Visscher PM, Yengo L, Cox NJ et al. Discovery and implications of polygenicity of common diseases. Science 2021;373:1468-73.
38. Tsuang MT, Bar JL, Stone WS et al. Gene-environment interactions in mental disorders. World Psychiatry 2004;3:73-83.
39. Weissman MM. Is depression nature or nurture? Yes. Am J Psychiatry 2020; 177:376-7.
40. Arribas-Ayllon M, Bartlett A, Lewis J. Psychiatric genetics: from hereditary madness to big biology. London: Routledge, 2019.
41. Evans K, McGrath J, Milns R. Searching for schizophrenia in ancient Greek and Roman literature: a systematic review. Acta Psychiatr Scand 2003;107: 323-30.
42. Lichtenstein P, Yip BH, Bjork C et al. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. Lancet 2009;373:234-9.
43. Smoller JW. Disorders and borders: psychiatric genetics and nosology. Am J Med Genet B Neuropsychiatr Genet 2013;162B:559-78.
44. Sullivan PF, Daly MJ, O'Donovan M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. Nat Rev Genet 2012; 13:537-51.
45. Arango C, Dragioti E, Solmi M et al. Risk and protective factors for mental disorders beyond genetics: an evidence-based atlas. World Psychiatry 2021; 20:417-36.
46. Bizzotto S, Walsh CA. Genetic mosaicism in the human brain: from lineage tracing to neuropsychiatric disorders. Nat Rev Neurosci 2022;23:275-86.
47. McAdams HH, Arkin A. Stochastic mechanisms in gene expression. Proc Natl Acad Sci USA 1997;94:814-9.
48. Faraone SV, Perlis RH, Doyle AE et al. Molecular genetics of attention-deficit/hyperactivity disorder. Biol Psychiatry 2005;57:1313-23.
49. Tick B, Bolton P, Happe F et al. Heritability of autism spectrum disorders: a meta-analysis of twin studies. J Child Psychol Psychiatry 2016;57:585-95.
50. Kieseppa T, Partonen T, Haukka J et al. High concordance of bipolar I disorder in a nationwide sample of twins. Am J Psychiatry 2004;161:1814-21.
51. Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Arch Gen Psychiatry 2003;60:1187-92.
52. Yilmaz Z, Hardaway JA, Bulik CM. Genetics and epigenetics of eating disorders. Adv Genomics Genet 2015;5:131-50.
53. Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. Am J Psychiatry 2000;157:1552-62.
54. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature 2015;526:68-74.
55. Martin AR, Kanai M, Kamatani Y et al. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet 2019;51:584-91.
56. Gurdasani D, Barroso I, Zeggini E et al. Genomics of disease risk in globally diverse populations. Nat Rev Genet 2019;20:520-35.
57. Uffelmann E, Huang QQ, Munung NS et al. Genome-wide association studies. Nat Rev Methods Primers 2021;1:59.
58. Klein RJ, Zeiss C, Chew EY et al. Complement factor H polymorphism in age-related macular degeneration. Science 2005;308:385-9.
59. Visscher PM, Wray NR, Zhang Q et al. 10 years of GWAS discovery: biology, function, and translation. Am J Hum Genet 2017;101:5-22.
60. MacArthur J, Bowler E, Cerezo M et al. The new NHGRI-EBI catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res 2017;45:D896-901.
61. Holland D, Frei O, Desikan R et al. Beyond SNP heritability: polygenicity and discoverability of phenotypes estimated with a univariate Gaussian mixture model. PLoS Genet 2020;16:e1008612.
62. Yang J, Lee SH, Goddard ME et al. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 2011;88:76-82.
63. Bulik-Sullivan BK, Loh PR, Finucane HK et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 2015;47:291-5.
64. Trubetskoy V, Pardinas AF, Qi T et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature 2022;604:502-8.
65. Mullins N, Forstner AJ, O'Connell KS et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. Nat Genet 2021;53:817-29.
66. Grove J, Ripke S, Als TD et al. Identification of common genetic risk variants for autism spectrum disorder. Nat Genet 2019;51:431-44.
67. Watson HJ, Yilmaz Z, Thornton LM et al. Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. Nat Genet 2019;51:1207-14.
68. Demontis D, Walters GB, Athanasiadis G et al. Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. medRxiv 2022. 02.14.22270780.
69. Levey DF, Gelernter J, Polimanti R et al. Reproducible genetic risk loci for anxiety: results from approximately 200,000 participants in the Million Veteran Program. Am J Psychiatry 2020;177:223-32.
70. Levey DF, Stein MB, Wendt FR et al. Bi-ancestral depression GWAS in the Million Veteran Program and meta-analysis in >1.2 million individuals highlight new therapeutic directions. Nat Neurosci 2021;24:954-63.
71. Stein MB, Levey DF, Cheng Z et al. Genome-wide association analyses of post-traumatic stress disorder and its symptom subdomains in the Million Veteran Program. Nat Genet 2021;53:174-84.
72. Yu D, Sul JH, Tsetsos F et al. Interrogating the genetic determinants of Tourette's syndrome and other tic disorders through genome-wide association studies. Am J Psychiatry 2019;176:217-27.
73. Strom NI, Yu D, Gerring ZF et al. Genome-wide association study identifies new locus associated with OCD. medRxiv 2021.10.13.21261078.
74. Manolio TA, Collins FS, Cox NJ et al. Finding the missing heritability of com-

plex diseases. Nature 2009;461:747-53.
75. Young AI. Solving the missing heritability problem. PLoS Genet 2019;15: e1008222.
76. Matthews LJ, Turkheimer E. Three legs of the missing heritability problem. Stud Hist Philos Sci 2022;93:183-91.
77. Grotzinger AD, Fuente J, Prive F et al. Pervasive downward bias in estimates of liability-scale heritability in genome-wide association study meta-analysis: a simple solution. Biol Psychiatry 2022; doi: 10.1016/j.biopsych.2022.05.029.
78. Frei O, Holland D, Smeland OB et al. Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. Nat Commun 2019;10:2417.
79. Watanabe K, Stringer S, Frei O et al. A global overview of pleiotropy and genetic architecture in complex traits. Nat Genet 2019;51:1339-48.
80. de Lange KM, Moutsianas L, Lee JC et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. Nat Genet 2017;49:256-61.
81. Psychiatric GWAS Consortium Steering Committee. A framework for interpreting genome-wide association studies of psychiatric disorders. Mol Psychiatry 2009;14:10-7.
82. Pardinas AF, Smart SE, Willcocks IR et al. Interaction testing and polygenic risk scoring to estimate the association of common genetic variants with treatment resistance in schizophrenia. JAMA Psychiatry 2022;79:260-9.
83. Hou L, Heilbronner U, Degenhardt F et al. Genetic variants associated with response to lithium treatment in bipolar disorder: a genome-wide association study. Lancet 2016;387:1085-93.
84. Pain O, Hodgson K, Trubetskoy V et al. Identifying the common genetic basis of antidepressant response. Biol Psychiatry Glob Open Sci 2022;2:115-26.
85. Mullins N, Kang J, Campos AI et al. Dissecting the shared genetic architecture of suicide attempt, psychiatric disorders, and known risk factors. Biol Psychiatry 2022;91:313-27.
86. Savage JE, Jansen PR, Stringer S et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nat Genet 2018;50:912-9.
87. Watanabe K, Jansen PR, Savage JE et al. Genome-wide meta-analysis of insomnia prioritizes genes associated with metabolic and psychiatric pathways. Nat Genet 2022;54:1125-32.
88. Karlsson Linner R, Biroli P, Kong E et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. Nat Genet 2019;51:245-57.
89. Ward J, Tunbridge EM, Sandor C et al. The genomic basis of mood instability: identification of 46 loci in 363,705 UK Biobank participants, genetic correlation with psychiatric disorders, and association with gene expression and function. Mol Psychiatry 2020;25:3091-9.
90. Tielbeek JJ, Johansson A, Polderman TJC et al. Genome-wide association studies of a broad spectrum of antisocial behavior. JAMA Psychiatry 2017;74: 1242-50.
91. Gelernter J, Polimanti R. Genetics of substance use disorders in the era of big data. Nat Rev Genet 2021;22:712-29.
92. Sanchez-Roige S, Palmer AA, Clarke TK. Recent efforts to dissect the genetic basis of alcohol use and abuse. Biol Psychiatry 2020;87:609-18.
93. Witt SH, Streit F, Jungkunz M et al. Genome-wide association study of borderline personality disorder reveals genetic overlap with bipolar disorder, major depression and schizophrenia. Transl Psychiatry 2017;7:e1155.
94. Bulik CM, Coleman JRI, Hardaway JA et al. Genetics and neurobiology of eating disorders. Nat Neurosci 2022;25:543-54.
95. Hindley G, Frei O, Shadrin AA et al. Charting the landscape of genetic overlap between mental disorders and related traits beyond genetic correlation. Am J Psychiatry 2022; doi: 10.1176/appi.ajp.21101051.
96. Cai N, Revez JA, Adams MJ et al. Minimal phenotyping yields genome-wide association signals of low specificity for major depression. Nat Genet 2020; 52:437-47.
97. Satterstrom FK, Walters RK, Singh T et al. Autism spectrum disorder and attention deficit hyperactivity disorder have a similar burden of rare protein-truncating variants. Nat Neurosci 2019;22:1961-5.
98. Ganna A, Satterstrom FK, Zekavat SM et al. Quantifying the impact of rare and ultra-rare coding variation across the phenotypic spectrum. Am J Hum Genet 2018;102:1204-11.
99. Singh T, Walters JTR, Johnstone M et al. The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. Nat Genet 2017;49:1167-73.
100. Sanders SJ, He X, Willsey AJ et al. Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. Neuron 2015;87:1215-33.
101. Power RA, Kyaga S, Uher R et al. Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings. JAMA Psychiatry 2013;70:22-30.
102. Satterstrom FK, Kosmicki JA, Wang J et al. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. Cell 2020;180:568-84.
103. De Rubeis S, He X, Goldberg AP et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature 2014;515:209-15.
104. Wang S, Mandell JD, Kumar Y et al. De novo sequence and copy number variants are strongly associated with Tourette disorder and implicate cell polarity in pathogenesis. Cell Rep 2018;25:3544.
105. Halvorsen M, Samuels J, Wang Y et al. Exome sequencing in obsessive-compulsive disorder reveals a burden of rare damaging coding variants. Nat Neurosci 2021;24:1071-6.
106. Singh T, Poterba T, Curtis D et al. Rare coding variants in ten genes confer substantial risk for schizophrenia. Nature 2022;604:509-16.
107. Genovese G, Fromer M, Stahl EA et al. Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. Nat Neurosci 2016;19:1433-41.
108. Palmer DS, Howrigan DP, Chapman SB et al. Exome sequencing in bipolar disorder identifies AKAP11 as a risk gene shared with schizophrenia. Nat Genet 2022;54:541-7.
109. Tian R, Ge T, Liu JZ et al. Whole exome sequencing in the UK Biobank reveals risk gene SLC2A1 and biological insights for major depressive disorder. medRxiv 2021.05.04. 21256398.
110. Karczewski KJ, Francioli LC, Tiao G et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020;581:434-43.
111. Akingbuwa WA, Hammerschlag AR, Bartels M et al. Ultra-rare and common genetic variant analysis converge to implicate negative selection and neuronal processes in the aetiology of schizophrenia. Mol Psychiatry 2022; doi: 10.1038/s41380-022-01621-8.
112. Marshall CR, Howrigan DP, Merico D et al. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. Nat Genet 2017;49:27-35.
113. Rees E, Kendall K, Pardinas AF et al. Analysis of intellectual disability copy number variants for association with schizophrenia. JAMA Psychiatry 2016; 73:963-9.
114. Gudmundsson OO, Walters GB, Ingason A et al. Attention-deficit hyperactivity disorder shares copy number variant risk with schizophrenia and autism spectrum disorder. Transl Psychiatry 2019;9:258.
115. Gazzellone MJ, Zarrei M, Burton CL et al. Uncovering obsessive-compulsive disorder risk genes in a pediatric cohort by high-resolution analysis of copy number variation. J Neurodev Disord 2016;8:36.
116. Huang AY, Yu D, Davis LK et al. Rare copy number variants in NRXN1 and CNTN6 increase risk for Tourette syndrome. Neuron 2017;94:1101-11.e7.
117. Kendall KM, Rees E, Bracher-Smith M et al. Association of rare copy number variants with risk of depression. JAMA Psychiatry 2019;76:818-25.
118. Green EK, Rees E, Walters JT et al. Copy number variation in bipolar disorder. Mol Psychiatry 2016;21:89-93.
119. Charney AW, Stahl EA, Green EK et al. Contribution of rare copy number variants to bipolar disorder risk is limited to schizoaffective cases. Biol Psychiatry 2019;86:110-9.
120. Weiner DJ, Wigdor EM, Ripke S et al. Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. Nat Genet 2017;49:978-85.
121. Bergen SE, Ploner A, Howrigan D et al. Joint contributions of rare copy number variants and common SNPs to risk for schizophrenia. Am J Psychiatry 2019; 176:29-35.
122. Tansey KE, Rees E, Linden DE et al. Common alleles contribute to schizophrenia in CNV carriers. Mol Psychiatry 2016;21:1085-9.
123. McDonald-McGinn DM, Sullivan KE, Marino B et al. 22q11.2 deletion syndrome. Nat Rev Dis Primers 2015;1:15071.
124. Davies RW, Fiksinski AM, Breetvelt EJ et al. Using common genetic variation to examine phenotypic expression and risk prediction in 22q11.2 deletion syndrome. Nat Med 2020;26:1912-8.
125. Antaki D, Guevara J, Maihofer AX et al. A phenotypic spectrum of autism is attributable to the combined effects of rare variants, polygenic risk and sex. Nat Genet 2022;54:1284-92.
126. Wong AK, Sealfon RSG, Theesfeld CL et al. Decoding disease: from genomes to networks to phenotypes. Nat Rev Genet 2021;22:774-90.
127. Uffelmann E, Posthuma D. Emerging methods and resources for biological interrogation of neuropsychiatric polygenic signal. Biol Psychiatry 2021; 89:41-53.
128. Flint J, Ideker T. The great hairball gambit. PLoS Genet 2019;15:e1008519.

129. Wojcik GL, Graff M, Nishimura KK et al. Genetic analyses of diverse populations improves discovery for complex traits. Nature 2019;570:514-8.
130. Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. Nat Rev Genet 2018;19:491-504.
131. Schork AJ, Thompson WK, Pham P et al. All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. PLoS Genet 2013;9:e1003449.
132. Roussos P, Mitchell AC, Voloudakis G et al. A role for noncoding variation in schizophrenia. Cell Rep 2014;9:1417-29.
133. Mountjoy E, Schmidt EM, Carmona M et al. An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. Nat Genet 2021;53:1527-33.
134. Gazal S, Weissbrod O, Hormozdiari F et al. Combining SNP-to-gene linking strategies to identify disease genes and assess disease omnigenicity. Nat Genet 2022;54:827-36.
135. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature 2014; 511:421-7.
136. Stahl EA, Breen G, Forstner AJ et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. Nat Genet 2019;51:793-803.
137. Wray NR, Ripke S, Mattheisen M et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat Genet 2018;50:668-81.
138. Smeland OB, Frei O, Dale AM et al. The polygenic architecture of schizophrenia – rethinking pathogenesis and nosology. Nat Rev Neurol 2020;16:366-79.
139. Kirov G, Pocklington AJ, Holmans P et al. De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. Mol Psychiatry 2012;17:142-53.
140. Clifton NE, Rees E, Holmans PA et al. Genetic association of FMRP targets with psychiatric disorders. Mol Psychiatry 2021;26:2977-90.
141. Szatkiewicz JP, Fromer M, Nonneman RJ et al. Characterization of single gene copy number variants in schizophrenia. Biol Psychiatry 2020;87:736-44.
142. Purcell SM, Moran JL, Fromer M et al. A polygenic burden of rare disruptive mutations in schizophrenia. Nature 2014;506:185-90.
143. Darnell JC, Van Driesche SJ, Zhang C et al. FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell 2011;146:247-61.
144. Stefansson H, Ophoff RA, Steinberg S et al. Common variants conferring risk of schizophrenia. Nature 2009;460:744-7.
145. International Schizophrenia Consortium, Purcell SM, Wray NR et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 2009;460:748-52.
146. Sekar A, Bialas AR, de Rivera H et al. Schizophrenia risk from complex variation of complement component 4. Nature 2016;530:177-83.
147. Stevens B, Allen NJ, Vazquez LE et al. The classical complement cascade mediates CNS synapse elimination. Cell 2007;131:1164-78.
148. Salter MW, Stevens B. Microglia emerge as central players in brain disease. Nat Med 2017;23:1018-27.
149. Moghaddam B, Javitt D. From revolution to evolution: the glutamate hypothesis of schizophrenia and its implication for treatment. Neuropsychopharmacology 2012;37:4-15.
150. Schrode N, Ho SM, Yamamuro K et al. Synergistic effects of common schizophrenia risk variants. Nat Genet 2019;51:1475-85.
151. Howard DM, Adams MJ, Clarke TK et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat Neurosci 2019;22:343-52.
152. Iakoucheva LM, Muotri AR, Sebat J. Getting to the cores of autism. Cell 2019; 178:1287-98.
153. Willsey HR, Willsey AJ, Wang B et al. Genomics, convergent neuroscience and progress in understanding autism spectrum disorder. Nat Rev Neurosci 2022;23:323-41.
154. Matoba N, Liang D, Sun H et al. Common genetic risk variants identified in the SPARK cohort support DDHD2 as a candidate risk gene for autism. Transl Psychiatry 2020;10:265.
155. Rees E, Creeth HDJ, Hwu HG et al. Schizophrenia, autism spectrum disorders and developmental disorders share specific disruptive coding mutations. Nat Commun 2021;12:5353.
156. Owen MJ, O'Donovan MC. Schizophrenia and the neurodevelopmental continuum: evidence from genomics. World Psychiatry 2017;16:227-35.
157. Wigdor EM, Weiner DJ, Grove J et al. The female protective effect against autism spectrum disorder. Cell Genomics 2022;2:1001342.
158. Seedat S, Scott KM, Angermeyer MC et al. Cross-national associations between gender and mental disorders in the World Health Organization World Mental Health Surveys. Arch Gen Psychiatry 2009;66:785-95.
159. Nelson MR, Tipney H, Painter JL et al. The support of human genetic evidence for approved drug indications. Nat Genet 2015;47:856-60.
160. Ochoa D, Karim M, Ghoussaini M et al. Human genetics evidence supports two-thirds of the 2021 FDA-approved drugs. Nat Rev Drug Discov 2022;21:551.
161. Harrison PJ, Tunbridge EM, Dolphin AC et al. Voltage-gated calcium channel blockers for psychiatric disorders: genomic reappraisal. Br J Psychiatry 2020;216:250-3.
162. Moncrieff J, Cooper RE, Stockmann T et al. The serotonin theory of depression: a systematic umbrella review of the evidence. Mol Psychiatry 2022; doi: 10.1038/s41380-022-01661-0.
163. Young AI, Benonisdottir S, Przeworski M et al. Deconstructing the sources of genotype-phenotype associations in humans. Science 2019;365:1396-400.
164. Kong A, Thorleifsson G, Frigge ML et al. The nature of nurture: effects of parental genotypes. Science 2018;359:424-8.
165. Howe LJ, Nivard MG, Morris TT et al. Within-sibship genome-wide association analyses decrease bias in estimates of direct genetic effects. Nat Genet 2022;54:581-92.
166. Paulus MP, Thompson WK. The challenges and opportunities of small effects: the new normal in academic psychiatry. JAMA Psychiatry 2019;76:353-4.
167. Fatumo S, Chikowore T, Choudhury A et al. A roadmap to increase diversity in genomic studies. Nat Med 2022;28:243-50.
168. Long E, Garcia-Closas M, Chanock SJ et al. The case for increasing diversity in tissue-based functional genomics datasets to understand human disease susceptibility. Nat Commun 2022;13:2907.
169. Breeze CE, Beck S, Berndt SI et al. The missing diversity in human epigenomic studies. Nat Genet 2022;54:737-9.
170. Alarcon RD. Culture, cultural factors and psychiatric diagnosis: review and projections. World Psychiatry 2009;8:131-9.
171. Lam M, Chen CY, Li Z et al. Comparative genetic architectures of schizophrenia in East Asian and European populations. Nat Genet 2019;51:1670-8.
172. Giannakopoulou O, Lin K, Meng X et al. The genetic architecture of depression in individuals of East Asian ancestry: a genome-wide association study. JAMA Psychiatry 2021;78:1258-69.
173. Kotov R, Krueger RF, Watson D et al. The Hierarchical Taxonomy of Psychopathology (HiTOP): a quantitative nosology based on consensus of evidence. Annu Rev Clin Psychol 2021;17:83-108.
174. Cuthbert BN. The RDoC framework: facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. World Psychiatry 2014;13:28-35.
175. Solovieff N, Cotsapas C, Lee PH et al. Pleiotropy in complex traits: challenges and strategies. Nat Rev Genet 2013;14:483-95.
176. Smeland OB, Frei O, Fan CC et al. The emerging pattern of shared polygenic architecture of psychiatric disorders, conceptual and methodological challenges. Psychiatr Genet 2019;29:152-9.
177. Smeland OB, Frei O, Kauppi K et al. Identification of genetic loci jointly influencing schizophrenia risk and the cognitive traits of verbal-numerical reasoning, reaction time, and general cognitive function. JAMA Psychiatry 2017;74:1065-75.
178. Smeland OB, Bahrami S, Frei O et al. Genome-wide analysis reveals extensive genetic overlap between schizophrenia, bipolar disorder, and intelligence. Mol Psychiatry 2020;25:844-53.
179. Le Hellard S, Wang Y, Witoelar A et al. Identification of gene loci that overlap between schizophrenia and educational attainment. Schizophr Bull 2017; 43:654-64.
180. Richards AL, Pardinas AF, Frizzati A et al. The relationship between polygenic risk scores and cognition in schizophrenia. Schizophr Bull 2020;46:336-44.
181. Smeland OB, Shadrin A, Bahrami S et al. Genome-wide association analysis of Parkinson's disease and schizophrenia reveals shared genetic architecture and identifies novel risk loci. Biol Psychiatry 2021;89:227-35.
182. Bahrami S, Hindley G, Winsvold BS et al. Dissecting the shared genetic basis of migraine and mental disorders using novel statistical tools. Brain 2022;145: 142-53.
183. Monereo-Sanchez J, Schram MT, Frei O et al. Genetic overlap between Alzheimer's disease and depression mapped onto the brain. Front Neurosci 2021;15:653130.
184. Brainstorm Consortium, Anttila V, Bulik-Sullivan B et al. Analysis of shared heritability in common disorders of the brain. Science 2018;360:eaap8757.
185. McLaughlin RL, Schijven D, van Rheenen W et al. Genetic correlation between amyotrophic lateral sclerosis and schizophrenia. Nat Commun 2017; 8:14774.

186. Bahrami S, Shadrin A, Frei O et al. Genetic loci shared between major depression and intelligence with mixed directions of effect. Nat Hum Behav 2021; 5:795-801.
187. Wistrom ED, O'Connell KS, Karadag N et al. Genome-wide analysis reveals genetic overlap between alcohol use behaviours, schizophrenia and bipolar disorder and identifies novel shared risk loci. Addiction 2022;117:600-10.
188. Kranzler HR, Zhou H, Kember RL et al. Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. Nat Commun 2019;10:1499.
189. Zhou H, Rentsch CT, Cheng Z et al. Association of OPRM1 functional coding variant with opioid use disorder: a genome-wide association study. JAMA Psychiatry 2020;77:1072-80.
190. Rodevand L, Bahrami S, Frei O et al. Extensive bidirectional genetic overlap between bipolar disorder and cardiovascular disease phenotypes. Transl Psychiatry 2021;11:407.
191. Torgersen K, Rahman Z, Bahrami S et al. Shared genetic loci between depression and cardiometabolic traits. PLoS Genet 2022;18:e1010161.
192. Andreassen OA, Djurovic S, Thompson WK et al. Improved detection of common variants associated with schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors. Am J Hum Genet 2013;92:197-209.
193. Bahrami S, Steen NE, Shadrin A et al. Shared genetic loci between body mass index and major psychiatric disorders: a genome-wide association study. JAMA Psychiatry 2020;77:503-12.
194. Hagenaars SP, Harris SE, Davies G et al. Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N=112 151) and 24 GWAS consortia. Mol Psychiatry 2016;21:1624-32.
195. Lo MT, Hinds DA, Tung JY et al. Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. Nat Genet 2017;49:152-6.
196. Power RA, Steinberg S, Bjornsdottir G et al. Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. Nat Neurosci 2015;18:953-5.
197. Smeland OB, Frei O, Shadrin A et al. Discovery of shared genomic loci using the conditional false discovery rate approach. Hum Genet 2020;139:85-94.
198. Purcell SM, Wray NR, Stone JL et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 2009;460:748-52.
199. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. Genome Res 2007;17:1520-8.
200. Bulik-Sullivan B, Finucane HK, Anttila V et al. An atlas of genetic correlations across human diseases and traits. Nat Genet 2015;47:1236-41.
201. Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee SH, Ripke S et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet 2013;45:984-94.
202. Wingo TS, Gerasimov ES, Canon SM et al. Alzheimer's disease genetic burden is associated with mid-life depression among persons with normal cognition. Alzheimers Dement 2022; doi: 10.1002/alz.12716.
203. Rees E, Kirov G. Copy number variation and neuropsychiatric illness. Curr Opin Genet Dev 2021;68:57-63.
204. Stefansson H, Meyer-Lindenberg A, Steinberg S et al. CNVs conferring risk of autism or schizophrenia affect cognition in controls. Nature 2014;505:361-6.
205. Huguet G, Schramm C, Douard E et al. Measuring and estimating the effect sizes of copy number variants on general intelligence in community-based samples. JAMA Psychiatry 2018;75:447-57.
206. Smoller JW, Andreassen OA, Edenberg HJ et al. Psychiatric genetics and the structure of psychopathology. Mol Psychiatry 2019;24:409-20.
207. Hyman SE. New evidence for shared risk architecture of mental disorders. JAMA Psychiatry 2019;76:235-6.
208. Abdellaoui A, Verweij KJH. Dissecting polygenic signals from genome-wide association studies on human behaviour. Nat Hum Behav 2021;5:686-94.
209. Smeland OB, Wang Y, Lo MT et al. Identification of genetic loci shared between schizophrenia and the Big Five personality traits. Sci Rep 2017;7:2222.
210. Okbay A, Beauchamp JP, Fontana MA et al. Genome-wide association study identifies 74 loci associated with educational attainment. Nature 2016;533: 539-42.
211. Okbay A, Baselmans BM, De Neve JE et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. Nat Genet 2016;48:624-33.
212. O'Connell KS, Frei O, Bahrami S et al. Characterizing the genetic overlap between psychiatric disorders and sleep-related phenotypes. Biol Psychiatry 2021;90:621-31.
213. Roelfs D, Alnaes D, Frei O et al. Phenotypically independent profiles relevant to mental health are genetically correlated. Transl Psychiatry 2021;11:202.
214. Grotzinger AD, Rhemtulla M, de Vlaming R et al. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. Nat Hum Behav 2019;3:513-25.
215. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. Cell 2019;179:1469-82. e1411.
216. Grotzinger AD, Mallard TT, Akingbuwa WA et al. Genetic architecture of 11 major psychiatric disorders at biobehavioral, functional genomic and molecular genetic levels of analysis. Nat Genet 2022;54:548-59.
217. Caspi A, Houts RM, Belsky DW et al. The p factor: one general psychopathology factor in the structure of psychiatric disorders? Clin Psychol Sci 2014;2:119-37.
218. Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. Cell 2018;173:1705-15.e1716.
219. Ruderfer DM, Fanous AH, Ripke S et al. Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. Mol Psychiatry 2014;19:1017-24.
220. Rovira P, Demontis D, Sanchez-Mora C et al. Shared genetic background between children and adults with attention deficit/hyperactivity disorder. Neuropsychopharmacology 2020;45:1617-26.
221. Rajagopal VM, Duan J, Vilar-Ribo L et al. Differences in the genetic architecture of common and rare variants in childhood, persistent and late-diagnosed attention-deficit hyperactivity disorder. Nat Genet 2022;54:1117-24.
222. Craddock N, Owen MJ. The Kraepelinian dichotomy – going, going… but still not gone. Br J Psychiatry 2010;196:92-5.
223. Kahn RS, Keefe RS. Schizophrenia is a cognitive illness: time for a change in focus. JAMA Psychiatry 2013;70:1107-12.
224. Thompson PM, Jahanshad N, Ching CRK et al. ENIGMA and global neuroscience: a decade of large-scale studies of the brain in health and disease across more than 40 countries. Transl Psychiatry 2020;10:100.
225. Bycroft C, Freeman C, Petkova D et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 2018;562:203-9.
226. Grasby KL, Jahanshad N, Painter JN et al. The genetic architecture of the human cerebral cortex. Science 2020;367:eaay6690.
227. Cheng W, Frei O, van der Meer D et al. Genetic association between schizophrenia and cortical brain surface area and thickness. JAMA Psychiatry 2021; 78:1020-30.
228. Franke B, Stein JL, Ripke S et al. Genetic influences on schizophrenia and subcortical brain volumes: large-scale proof of concept. Nat Neurosci 2016;19: 420-31.
229. Ohi K, Shimada T, Kataoka Y et al. Genetic correlations between subcortical brain volumes and psychiatric disorders. Br J Psychiatry 2020;216:280-3.
230. van der Meer D, Shadrin AA, O'Connell K et al. Boosting schizophrenia genetics by utilizing genetic overlap with brain morphology. Biol Psychiatry 2022;92:291-8.
231. Smeland OB, Wang Y, Frei O et al. Genetic overlap between schizophrenia and volumes of hippocampus, putamen, and intracranial volume indicates shared molecular genetic mechanisms. Schizophr Bull 2018;44:854-64.
232. Elvsashagen T, Shadrin A, Frei O et al. The genetic architecture of the human thalamus and its overlap with ten common brain disorders. Nat Commun 2021;12:2909.
233. Bahrami S, Nordengen K, Shadrin AA et al. Distributed genetic architecture across the hippocampal formation implies common neuropathology across brain disorders. Nat Commun 2022;13:3436.
234. Chambers T, Escott-Price V, Legge S et al. Genetic common variants associated with cerebellar volume and their overlap with mental disorders: a study on 33,265 individuals from the UK-Biobank. Mol Psychiatry 2022;27:2282-90.
235. Elvsashagen T, Bahrami S, van der Meer D et al. The genetic architecture of human brainstem structures and their involvement in common brain disorders. Nat Commun 2020;11:4016.
236. van Erp TG, Hibar DP, Rasmussen JM et al. Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. Mol Psychiatry 2016;21:547-53.
237. van Erp TGM, Walton E, Hibar DP et al. Cortical brain abnormalities in 4474 individuals with schizophrenia and 5098 control subjects via the Enhancing Neuro Imaging Genetics Through Meta Analysis (ENIGMA) Consortium. Biol Psychiatry 2018;84:644-54.
238. Moberget T, Doan NT, Alnaes D et al. Cerebellar volume and cerebellocerebral structural covariance in schizophrenia: a multisite mega-analysis of 983 patients and 1349 healthy controls. Mol Psychiatry 2018;23:1512-20.
239. Lee PH, Feng YA, Smoller JW. Pleiotropy and cross-disorder genetics among psychiatric disorders. Biol Psychiatry 2021;89:20-31.
240. Hartwig FP, Borges MC, Horta BL et al. Inflammatory biomarkers and risk of schizophrenia: a 2-sample Mendelian randomization study. JAMA Psychia-

try 2017;74:1226-33.
241. Reay WR, Kiltschewskij DJ, Geaghan MP et al. Genetic estimates of correlation and causality between blood-based biomarkers and psychiatric disorders. Sci Adv 2022;8:eabj8969.
242. Williams JA, Burgess S, Suckling J et al. Inflammation and brain structure in schizophrenia and other neuropsychiatric disorders: a Mendelian randomization study. JAMA Psychiatry 2022;79:498-507.
243. Kelly KM, Smith JA, Mezuk B. Depression and interleukin-6 signaling: a Mendelian randomization study. Brain Behav Immun 2021;95:106-14.
244. Kappelmann N, Arloth J, Georgakis MK et al. Dissecting the association between inflammation, metabolic dysregulation, and specific depressive symptoms: a genetic correlation and 2-sample Mendelian randomization study. JAMA Psychiatry 2021;78:161-70.
245. Tylee DS, Lee YK, Wendt FR et al. An atlas of genetic correlations and genetically informed associations linking psychiatric and immune-related phenotypes. JAMA Psychiatry 2022;79:667-76.
246. Denny JC, Bastarache L, Roden DM. Phenome-wide association studies as a tool to advance precision medicine. Annu Rev Genomics Hum Genet 2016; 17:353-73.
247. Zheutlin AB, Dennis J, Karlsson Linner R et al. Penetrance and pleiotropy of polygenic risk scores for schizophrenia in 106,160 patients across four health care systems. Am J Psychiatry 2019;176:846-55.
248. Zhang R, Sjolander A, Ploner A et al. Novel disease associations with schizophrenia genetic risk revealed in ~400,000 UK Biobank participants. Mol Psychiatry 2022;27:1448-54.
249. Sealock JM, Lee YH, Moscati A et al. Use of the PsycheMERGE Network to investigate the association between depression polygenic scores and white blood cell count. JAMA Psychiatry 2021;78:1365-74.
250. Beurel E, Toups M, Nemeroff CB. The bidirectional relationship of depression and inflammation: double trouble. Neuron 2020;107:234-56.
251. Denny JC, Collins FS. Precision medicine in 2030 – seven ways to transform healthcare. Cell 2021;184:1415-9.
252. He KY, Ge D, He MM. Big data analytics for genomic medicine. Int J Mol Sci 2017;18:412.
253. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. Genome Med 2020;12:44.
254. Wand H, Lambert SA, Tamburro C et al. Improving reporting standards for polygenic scores in risk prediction studies. Nature 2021;591:211-9.
255. Khera AV, Chaffin M, Aragam KG et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet 2018;50:1219-24.
256. Mars N, Koskela JT, Ripatti P et al. Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. Nat Med 2020;26:549-57.
257. Ferrat LA, Vehik K, Sharp SA et al. A combined risk score enhances prediction of type 1 diabetes among susceptible children. Nat Med 2020;26:1247-55.
258. Mavaddat N, Michailidou K, Dennis J et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. Am J Hum Genet 2019; 104:21-34.
259. Murray GK, Lin T, Austin J et al. Could polygenic risk scores be useful in psychiatry?: a review. JAMA Psychiatry 2021;78:210-9.
260. Lewis ACF, Green RC, Vassy JL. Polygenic risk scores in the clinic: translating risk into action. HGG Adv 2021;2:100047.
261. Muse ED, Chen SF, Liu S et al. Impact of polygenic risk communication: an observational mobile application-based coronary artery disease study. NPJ Digit Med 2022;5:30.
262. Janssens A, Martens FK. Reflection on modern methods: revisiting the area under the ROC curve. Int J Epidemiol 2020;49:1397-403.
263. Murphy KC, Jones LA, Owen MJ. High rates of schizophrenia in adults with velo-cardio-facial syndrome. Arch Gen Psychiatry 1999;56:940-5.
264. Bassett AS, Chow EW. 22q11 deletion syndrome: a genetic subtype of schizophrenia. Biol Psychiatry 1999;46:882-91.
265. Perkins DO, Olde Loohuis L, Barbee J et al. Polygenic risk score contribution to psychosis prediction in a target population of persons at clinical high risk. Am J Psychiatry 2020;177:155-63.
266. Rice F, Riglin L, Thapar AK et al. Characterizing developmental trajectories and the role of neuropsychiatric genetic risk variants in early-onset depression. JAMA Psychiatry 2019;76:306-13.
267. Ward J, Graham N, Strawbridge RJ et al. Polygenic risk scores for major depressive disorder and neuroticism as predictors of antidepressant response: meta-analysis of three treatment cohorts. PLoS One 2018;13:e0203896.
268. Zhang JP, Robinson D, Yu J et al. Schizophrenia polygenic risk score as a predictor of antipsychotic efficacy in first-episode psychosis. Am J Psychiatry 2019;176:21-8.
269. El-Hage W, Leman S, Camus V et al. Mechanisms of antidepressant resistance. Front Pharmacol 2013;4:146.
270. Howes OD, Thase ME, Pillinger T. Treatment resistance in psychiatry: state of the art and new directions. Mol Psychiatry 2022;27:58-72.
271. Lally J, MacCabe JH. Antipsychotic medication in schizophrenia: a review. Br Med Bull 2015;114:169-79.
272. Campos AI, Mulcahy A, Thorp JG et al. Understanding genetic risk factors for common side effects of antidepressant medications. Commun Med 2021;1: 45.
273. Kurki MI, Karjalainen J, Palta P et al. FinnGen: unique genetic insights from combining isolated population and national health register data. medRxiv 2022.03.03.22271360.
274. Jermy BS, Glanville KP, Coleman JRI et al. Exploring the genetic heterogeneity in major depression across diagnostic criteria. Mol Psychiatry 2021;26:7337-45.
275. Torrey EF, Yolken RH. Schizophrenia as a pseudogenetic disease: a call for more gene-environmental studies. Psychiatry Res 2019;278:146-50.
276. Shi Y, Sprooten E, Mulders P et al. Multi-polygenic scores in psychiatry: from disorder-specific to transdiagnostic perspectives. medRxiv 2022.05.30. 22275563.
277. Hjorthøj C, Stürup AE, McGrath JJ et al. Years of potential life lost and life expectancy in schizophrenia: a systematic review and meta-analysis. Lancet Psychiatry 2017;4:295-301.
278. Sun L, Pennells L, Kaptoge S et al. Polygenic risk scores in cardiovascular risk prediction: a cohort study and modelling analyses. PLoS Med 2021;18: e1003498.
279. Kalokairinou L, Howard HC, Slokenberga S et al. Legislation of direct-to-consumer genetic testing in Europe: a fragmented regulatory landscape. J Community Genet 2018;9:117-32.
280. Nelson SC, Bowen DJ, Fullerton SM. Third-party genetic interpretation tools: a mixed-methods study of consumer motivation and behavior. Am J Hum Genet 2019;105:122-31.
281. Goldberg C. The Pandora's box of embryo testing is officially open. Bloomberg, May 26, 2022.
282. Kumar A, Im K, Banjevic M et al. Whole-genome risk prediction of common diseases in human preimplantation embryos. Nat Med 2022;28:513-6.
283. Turley P, Meyer MN, Wang N et al. Problems with using polygenic scores to select embryos. N Engl J Med 2021;385:78-86.
284. Lencz T, Sabatello M, Docherty A et al. Concerns about the use of polygenic embryo screening for psychiatric and cognitive traits. Lancet Psychiatry 2022;9: 838-44.
285. Widen E, Lello L, Raben TG et al. Polygenic health index, general health, and disease risk. medRxiv 2022.06.15.22276102.
286. Sullivan PF. The psychiatric GWAS consortium: big science comes to psychiatry. Neuron 2010;68:182-6.
287. Pedersen CB, Bybjerg-Grauholm J, Pedersen MG et al. The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. Mol Psychiatry 2018;23:6-14.
288. Hakonarson H, Gulcher JR, Stefansson K. deCODE genetics, Inc. Pharmacogenomics 2003;4:209-15.
289. Wainschtein P, Jain D, Zheng Z et al. Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. Nat Genet 2022;54:263-73.
290. Walters RK, Polimanti R, Johnson EC et al. Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. Nat Neurosci 2018;21:1656-69.
291. Martin AR, Gignoux CR, Walters RK et al. Human demographic history impacts genetic risk prediction across diverse populations. Am J Hum Genet 2017;100:635-49.
292. Peterson RE, Kuchenbaecker K, Walters RK et al. Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. Cell 2019;179:589-603.
293. Ramirez AH, Sulieman L, Schlueter DJ et al. The All of Us research program: data quality, utility, and diversity. Patterns 2022;3:100570.
294. Huynh-Le MP, Fan CC, Karunamuni R et al. Polygenic hazard score is associated with prostate cancer in multi-ethnic populations. Nat Commun 2021;12:1236.
295. Kurniansyah N, Goodman MO, Kelly TN et al. A multi-ethnic polygenic risk score is associated with hypertension prevalence and progression throughout adulthood. Nat Commun 2022;13:3549.
296. Amariuta T, Ishigaki K, Sugishita H et al. Improving the trans-ancestry port-

ability of polygenic risk scores by prioritizing variants in predicted cell-type-specific regulatory elements. Nat Genet 2020;52:1346-54.
297. Manchia M, Vieta E, Smeland OB et al. Translating big data to better treatment in bipolar disorder – a manifesto for coordinated action. Eur Neuropsychopharmacol 2020;36:121-36.
298. van der Meer D, Frei O, Kaufmann T et al. Understanding the genetic determinants of the brain with MOSTest. Nat Commun 2020;11:3512.
299. O'Connor LJ. The distribution of common-variant effect sizes. Nat Genet 2021; 53:1243-9.
300. Zeng J, de Vlaming R, Wu Y et al. Signatures of negative selection in the genetic architecture of human complex traits. Nat Genet 2018;50:746-53.
301. Zhang Y, Qi G, Park JH et al. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. Nat Genet 2018;50:1318-26.
302. Solmi M, Radua J, Olivola M et al. Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. Mol Psychiatry 2022;27:281-95.
303. Hofman A, Grobbee DE, de Jong PT et al. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. Eur J Epidemiol 1991;7:403-22.
304. Mahmood SS, Levy D, Vasan RS et al. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. Lancet 2014; 383:999-1008.
305. Fraser A, Macdonald-Wallis C, Tilling K et al. Cohort profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. Int J Epidemiol 2013;42:97-110.
306. Magnus P, Birke C, Vejrup K et al. Cohort profile update: the Norwegian Mother and Child Cohort Study (MoBa). Int J Epidemiol 2016;45:382-8.
307. Olsen J, Melbye M, Olsen SF et al. The Danish National Birth Cohort – its background, structure and aim. Scand J Public Health 2001;29:300-7.
308. Zheng JS, Liu H, Jiang J et al. Cohort profile: the Jiaxing Birth Cohort in China. Int J Epidemiol 2017;46:1382-2g.
309. Yue W, Zhang E, Liu R et al. The China Birth Cohort Study (CBCS). Eur J Epidemiol 2022;37:295-304.
310. Choi KW, Wilson M, Ge T et al. Integrative analysis of genomic and exposomic influences on youth mental health. J Child Psychol Psychiatry 2022; doi: 10.1111/jcpp.13664.
311. Magnusdottir I, Lovik A, Unnarsdottir AB et al. Acute COVID-19 severity and mental health morbidity trajectories in patient populations of six nations: an observational study. Lancet Public Health 2022;7:e406-16.
312. Besterman AD, Moreno-De-Luca D, Nurnberger JI Jr. 21st-century genetics in psychiatric residency training: how do we get there? JAMA Psychiatry 2019; 76:231-2.

DOI:10.1002/wps.21034

The alliance in mental health care: conceptualization, evidence and clinical applications

Bruce E. Wampold[1,2], Christoph Flückiger[3,4]

[1]Modum Bad Psychiatric Center, Vikersund, Norway; [2]University of Wisconsin-Madison, Madison, WI, USA; [3]Department of Psychology, University of Zürich, Zürich, Switzerland; [4]Department of Psychology, University of Kassel, Kassel, Germany

The concept of alliance reflects the collaborative relationship between a clinician and a patient, defined as consisting of three elements: a) the agreement on the goals of treatment; b) the agreement on a task or series of tasks; c) the development of a bond. Although much of the theory and research on the alliance comes from the domain of psychotherapy, the concept is applicable to any practice involving a person seeking help and a socially sanctioned healer. An extensive research evidence suggests that the alliance (typically measured at the third or fourth session) is a robust predictor of the outcomes of various forms of psychotherapy, even when prior symptom improvement and other factors are considered. Both the clinician and the patient bring to the therapy situation different capacities to form an alliance. Factors concerning the patient include, among others, the diagnosis, attachment history and style, motivation, and needs for affiliation. However, the benefits of the alliance have been found to be mostly due to the therapist's contribution, in particular his/her facilitative interpersonal skills, including verbal fluency, communication of hope and positive expectations, persuasiveness, emotional expression; warmth, acceptance and understanding; empathy, and alliance rupture-repair responsiveness. Placebo studies have allowed to experimentally manipulate aspects of the relationship between a therapist and a patient in non-psychotherapy contexts. In these settings, two components of the relationship have emerged: an emotional one (involving being cared for and understood by the clinician) and a cognitive one (including the belief in the competence of the therapist to select and administer an effective treatment). Here we propose a model that describes three pathways through which the alliance creates benefits, named CARE (caring, attentive, real and empathic), EXPECTANCY, and SPECIFIC. Although research and clinical attention have mostly focused on the alliance between a clinician and a patient in face-to-face interactions, there is preliminary evidence concerning the alliance between patients and other clinic staff, systems of care, or the program in Internet-mediated services. These new research areas clearly require further development.

Key words: Alliance, relationship, bond, expectations, treatment goals, competence, warmth, empathy, placebos, trust

(World Psychiatry 2023;22:25–41)

In many instances, there is a propensity for humans to disregard phenomena that permeate everyday life. For example, we converse using language much of the day without paying it the least regard. Of course, we become acutely aware of language when confronted with an unusual situation, such as an interaction with a person with aphasia, when interacting with others who are using an unfamiliar language, or when having to pick our words carefully in a challenging situation. Yet language, when examined, is exceedingly complex and is studied and understood from a variety of perspectives, including linguistics, psychology, neuroscience, anthropology, sociology, and literature. Language is vital to human life – without it, humans could not exist.

The alliance is the "language" of mental health care. To varying degrees, it is present in all interactions between a clinician and a patient but, like language, it is typically ignored until it is disrupted or vanishes. Examining the alliance from multiple perspectives unveils its nature and highlights aspects of it that could lead to improved quality of care.

In this paper, we trace the historical roots of the alliance concept, and provide a definition of it. We then review the evidence related to the alliance, which demonstrates its importance for the outcomes of mental health treatments. These discussions lead to a presentation of the psychological mechanisms that explain how the alliance produces benefits, and of clinical applications, including some recent developments which involve systems of care.

HISTORICAL ROOTS AND DEFINITION OF THE ALLIANCE CONCEPT

The concept of the alliance is usually traced to E. Bordin's seminal 1979 paper entitled *The Generalizability of the Psychoanalytic Concept of the Working Alliance*[1]. Bordin intertwined two psychoanalytic threads. The first involved the relationship between the analyst and the patient's rational ego as well as the notion of a therapeutic contract[2,3]. The second borrowed the psychoanalytic concept of the "real relationship", which is the transference-free relationship between the patient and the analyst[4,5].

Bordin's contribution was to weave the two threads together to define a concept that he labeled the *working alliance*, which applied to all forms of psychotherapy as well as to other relationships that involved a person seeking help and a person designated as a helper. He defined the alliance as containing three elements: a) the agreement on the goals of treatment; b) the assignment of a task or series of tasks; and c) the development of a bond. Several of the issues discussed by Bordin over a half century ago remain central to current discussions of the alliance.

The title of Bordin's paper mentioned *generalizability* to emphasize that the importance of the alliance was not limited to psychoanalysis. Indeed, he stated: "I propose that the working alliance between the person who seeks change and the one who offers to be a change agent is one of the keys, if not the key, to the change process… A working alliance between a person seeking change and a change agent can occur in many places besides the locale of psychotherapy. The working alliance can be defined and elaborated in terms which make it universally applicable"[1, p.252]. Accordingly, his model is often referred to as *trans-theoretical*, although he did not use that label. However, he did emphasize that aspects of the alliance will depend on the nature of the treatment used to create change. That is, the nature of the alliance and how it leads to improved outcomes depends on the particular treatment.

The expression *agreement on the goals of treatment* suggests to many that the therapist and the patient explicitly discuss the goals of treatment, coming to an agreement, after which the treatment can begin. However, it rarely happens this way. It seems that experienced therapists in high-alliance and successful cases rarely explicitly discuss the very specific goals of treatment, although they do induce a future orientation through various techniques[6,7]. This raises the question of what is meant by *goals of treatment*, particularly the level of specificity of such goals.

As Bordin discussed, the choice of therapist and therapy determines much about the goals of treatment. Treatment by a psychoanalyst or a psychodynamically oriented therapist "rests on the mutual agreement that the patient's stresses, frustrations and dissatisfactions are to a significant extent a function of his own ways of thinking, feeling and acting"[1, p.253], but this understanding may not be realized until therapy has progressed for some time. On the other hand, cognitive and behavior therapists direct attention toward more concrete and circumscribed goals related to behavior, cognitions, emotions and values. Some therapies emphasize character or personality change, while others are focused on symptoms or well-being. The goals for a patient receiving psychopharmacological treatment will be typically focused on symptoms of the disorder.

Clearly, agreement on goals is not a simple matter. The use of the terms *goals* and *agreement on goals* suggest to many a degree of specificity; alternative language could refer to *general aims of treatment* and *clarification of aims of treatment*. Moreover, as any clinician knows, what the patient identifies as problematic in his/her life may change as therapy provides insight or understanding. Further complicating the situation, patients may report that they have come to an agreement on the goals of therapy in the absence of any discussion of goals[7], suggesting that an implicit understanding might be sufficient. Anyway, the degree to which psychotherapy is focused on the patient's perceived problems is related to the efficacy of the treatment[8].

The second element of the alliance, as formulated by Bordin, is therapist's *assignment of tasks*. Bordin was clear that the choice of therapeutic tasks is not unilaterally made by the therapist and presented to the patient, and noted that "collaboration between patient and therapist involves an agreed-upon contract"[1, p.254]. However, he recognized that the choice of therapist determined the range of tasks that would be utilized in therapy.

The particular tasks assigned by therapists will be different across orientations. For example, a patient presenting to a biologically oriented psychiatrist will not be surprised to receive a prescription for psychotropic medication, and the patient's task will involve taking the medication as prescribed. Thus, the patient has expectations about the nature of the tasks that will be assigned, which predisposes to collaboration and creates expectations for the outcomes of the therapy, thereby increasing its effectiveness, as will be discussed later[9-11].

Despite the frequent citation of Bordin when discussing the alliance, the *assignment of tasks* element of the alliance is commonly referred to as *agreement on the tasks of treatment*, although it is important to remember the asymmetric relationship in mental health care, where the clinician has a particular expertise and various therapeutic skills that influence the tasks of treatment. As will be discussed, the clinician's persuasiveness and verbal fluency increase collaboration between the clinician and the patient. That is, the manner in which the clinician explains the treatment influences the degree to which the patient believes that the treatment will be effective.

The *bond* between the clinician and the patient is the least well defined and understood of the three elements of the alliance, and is the most controversial. According to Bordin, goal setting and collaboration on the tasks of treatment "appear intimately linked to the nature of the human relationship between therapist and patient"[1, p.254]. Calling the third therapeutic element the *bond* conveys the idea that it is linked to the relationship, but there are two central ways that the bond has been discussed in the literature.

First, the bond has been conceptualized as the "real relationship", which refers to the collaborative quality of a genuine, caring, unconditional and understanding stance of the clinician, something akin to C. Rogers's "core conditions"[12]. Such a collaborative relationship quality can be healing in and of itself, as discussed later. A second interpretation of the bond is one of trust: for example, does the patient sufficiently trust that the clinician has the expertise to be helpful, so that the patient is willing to engage in the difficult and sometimes distressing work involved in the treatment? The former is oriented toward the person of the clinician, and the latter toward the competence of the clinician. Both aspects are valuable, but the distinction is important.

Moreover, the nature of the bond might well depend on the nature of the treatment, the treatment stage, and the patient's characteristics, as noted by Bordin: "Some basic level of trust surely marks all varieties of therapeutic relationships, but when attention is directed toward the more protected recesses of inner experience, deeper bonds of trust and attachment are required and developed… One bond may not necessarily be stronger than the other, but they do differ in kind"[1, p.254].

There is a characteristic of the alliance that separates it from all, or almost all, other healing concepts. The alliance is, by definition, a dyadic concept. The alliance is created by the work that the clinician and the patient do together. Other therapeutic concepts involve conditions created or actions taken by the clinicians, although patients will be affected by or react to such conditions and actions differently. Consider empathy: a therapist can offer an empathic response to a patient after the patient describes a difficult event in his/her life, and such a response can be seen as empathic regardless of how the patient receives, understands and is affected by the response. By definition there is no "alliant-ic" therapist response, as alliance is created in the dyadic interaction and is a phenomenon that occurs as a consequence of the therapist and patient interaction.

As such, both participants contribute to the alliance. The therapist creates the conditions under which the alliance will develop, but importantly patients perceive this as having a collaborative quality. The ontological distinction between the alliance and other therapeutic factors has been highlighted most convincingly by R. Hatcher[13], who emphasized that the alliance is a collaborative

construct. The dyadic nature of the alliance is central to understanding its role in leading to effective treatment.

Although Bordin's discussion of the alliance was ground-breaking and his ideas have persisted, there have been theoretical variations on his conceptualization, one of which offers particularly important insights. L. Luborsky and A.O. Horvath[14-16] discussed the alliance from a variety of perspectives, including its psychodynamic origins, its Rogerian client-centered relational aspects, the social influence concept, and the pan-theoretical perspective.

From these multiple perspectives, two types of alliance were identified as well as a sequencing of these types over the course of treatment. Luborsky suggested that the alliance is a dynamic rather than a static entity, responsive to the changing demands of different phases of therapy. Type 1 alliance is "based on the patient's experiencing the therapist as supportive and helpful with himself as a recipient"; Type 2 alliance is "a sense of working together in a joint struggle against what is impeding the patient… on shared responsibility for working out treatment goals… a sense of 'we-ness'"[14, p.563]. According to Luborsky, Type 1 alliance is more evident in the beginning of therapy, and Type 2 more typical of later phases of treatment.

Although much of the theory and research on the alliance comes from the psychotherapy domain, the concept is applicable, as Bordin emphasized, to any practice involving a person seeking help and a socially sanctioned healer. Accordingly, we will discuss alliance with a psychotherapist and then expand the concept by discussing other domains, including psychiatry, medicine and placebos, among others.

As the alliance became to be seen as central to mental health treatments, researchers needed to have a reliable and valid way to measure it. We now discuss several of the measures of the alliance. Because the alliance is a dyadic phenomenon, respondents using these instruments are giving their own sense of the alliance. Consequently, clinicians and patients typically assess identical items, but rate the alliance as they perceive it. The clinician and the patient may not perceive the alliance similarly, as each rates the alliance filtered through his/her own lens and interpretation of the interaction. There are some instruments in which an observer rates the alliance, providing an outsider's perspective, although observers are still rating on the basis of their perspective of a dyadic construct.

MEASUREMENT OF THE ALLIANCE

Measurement of interpersonal perceptions of individuals in a social context has been a lasting challenge in psychological sciences[e.g.,17,18]. For example, a person may love his/her partner but, at the same time, his/her evaluation will also consider how much it feels that this kind of love is reciprocated[e.g.,19,20]. Evaluating the alliance needs to consider the relationship of two persons as well as the two persons, as individuals, with individual characteristics. According to Kenny's social relations model[21], the evaluation focuses on three components: perceiver, target and relationship.

Alliance scores are thus based on the two actors and their general rating tendencies as well as their perceptions of the other and the relationship[16]. More specifically, alliance is assessed by particular measures completed by raters (patient, therapist, or sometimes an observer) evaluating a relational phenomenon at a particular time in therapy. The majority of studies assessing the alliance refer to overall reports at the end of a session (item examples: "I feel that my therapist appreciates me"; "As a result of these sessions I am clearer as to how I might be able to change"; "I believe the way we are working with my problem is correct"). These items do provide a more general alliance evaluation across sessions, and they are not focused on a particular intervention or time during a session. There is some empirical indication that the alliance assessed at post-session is rated somewhat higher than the alliance immediately before the next therapy session, even though no additional interaction occurred[22].

Four post-session alliance measures – the California Psychotherapy Alliance Scale (CALPAS)[23], the Helping Alliance Questionnaire (HAQ)[24], the Vanderbilt Psychotherapy Process Scale (VPPS)[25], and the Working Alliance Inventory (WAI)[26] – are used in approximately two-thirds of the alliance-outcome studies. Over time, there has been a trend toward developing and using shorter versions of these measurement instruments. About 70% of the published papers in the past decade have used an inventory based on WAI items[18]. Separate versions for patient, therapist and observer ratings have been developed. Each of the above-mentioned four core instruments has demonstrated acceptable internal consistency, in the range of .81 to .87 (Cronbach's alpha).

Various studies of the factor structure of the measures range from multiple factors to more coordinated perceptions across the alliance elements (e.g., coordinated view of tasks, goals and bond[e.g.,27]). The shared variance of alliance across measures and evaluators is low, indicating that there is much uniqueness in the alliance ratings of particular evaluators[28].

Despite these issues of measurement, the evidence for the importance of the alliance converges across raters, measures and assessment times, and how the alliance is involved in producing therapeutic benefits is in many ways unambiguous.

EVIDENCE FOR THE BENEFITS OF THE ALLIANCE

A search for the term "alliance" in the titles of articles indexed in the PsycINFO database yields approximately 5,000 publications that deal with the alliance in the sense used here. Consistent with Bordin's observation that the alliance spans an array of healing settings, the concept is also referenced in medicine (>900 hits in PsycINFO), social work (>800 hits), nursing (>200 hits), school counseling (>600 hits), and pharmacotherapy (>100 hits). The emphasis on the alliance is also central in the emerging patient-centered care movement[29].

In this section, we review the evidence for the benefits of the alliance. It will be clear that making valid conclusions from the available research is challenging, because the alliance is complex and designing research to investigate it is difficult. There are threats to validity to each alliance study as well as to all studies using a particular design. To rule out various threats, the design

of the studies has evolved. The evidence produced by the studies also reveals important aspects of the alliance, showing that research and theory development go hand-in-hand.

Due to the volume of the alliance research, various meta-analyses have been conducted, the results of which will be cited to summarize the evidence. For various critical issues, particular studies will be discussed.

The association between the alliance and outcomes of treatment

At the most basic level, if the alliance is an important aspect of mental health care, then the alliance measured during the course of therapy should predict the final outcomes of treatment. Said another way, the stronger the alliance, the better the outcomes of treatment.

The first study that investigated the association of the alliance with outcomes was a doctoral dissertation by A.O. Horvath in 1981[30], who studied 29 patients receiving various types of treatment. The alliance was measured by the WAI (rated by both patient and therapist) early in therapy, and outcomes were measured by the Psychotherapy Questionnaire (also rated by both patient and therapist). Across the various measures, the alliance-outcome correlation was .49, suggesting a rather strong association.

By 1991, there was a sufficient number of studies (i.e., 24) to conduct a meta-analysis of the alliance-outcome association. The typical study measured the alliance early in treatment (at the third or fourth session) and then the correlation of the alliance score with outcomes as a criterion variable was calculated. The results of this meta-analysis[31] are presented in Table 1. The 24 studies involved 1,148 patients and yielded an aggregate correlation of .26, which is generally considered of moderate size. When converted to standardized mean difference (SMD), the effect was .54, which would be regarded as sizable and clinically important. This effect size indicates that seven percent of the variability in outcomes (i.e., R^2) is due to the alliance. Although this may not appear impressively large, there is no variable measured early in therapy, except for initial severity of the patient's condition, that predicts the outcomes better than the alliance.

The number of studies examining the correlation between the alliance and outcomes has remarkably increased over the years. Four additional meta-analyses have been conducted since 1991[32-35], whose results are summarized in Table 1. Clearly, the range of the aggregate correlation of alliance with outcomes exceeds .20, and in the most recent and comprehensive meta-analysis approaches .30. Due to the number of studies (almost 300) and number of patients (over 30,000) within the studies in the most recent meta-analysis[35], it is safe to conclude that there is a robust association between alliance and outcomes of psychotherapy. Indeed, the standard error of estimate for the aggregate correlation of .28 was approximately .011.

Importantly, the association of the alliance between the therapist and youth is also predictive of outcomes[36]. Furthermore, the alliance is associated with outcomes also in marital, family and group therapy, although in these cases there are multiple alliances to consider[37,38].

The adage that "correlation does not mean causation" provides a cautionary note to making claims about the alliance from these meta-analyses, even if they are comprehensive and precise. However, research has burgeoned to address many of the threats to the validity of the conclusion that the alliance is a central therapeutic factor, and also provides clinical insight into how the alliance is therapeutic. We now briefly review this additional evidence.

Is the alliance an epiphenomenon of early symptom change?

The correlation between the alliance and outcomes discussed earlier involves a measurement of the alliance early in therapy, typically at the third or fourth session. The alliance, it is thought, cannot be validly assessed earlier, because it is a dyadic construct that needs sufficient clinician-patient interaction to develop. However, by the time the alliance is measured, many patients will have experienced a significant decrease in distress[39,40], which has generated two conjectures about early treatment gains.

The first conjecture, put forth by DeRubeis et al[41] among others, is that the specific treatment actions create early change, and it may well be that the patients who have experienced significant benefits early in treatment will tend to rate all aspects of the treatment favorably, including the alliance, and will have better

Table 1 Summary of meta-analyses of the correlation of alliance and outcome

	Population	N. studies	N. patients	Aggregate correlation (r)	Equivalent SMD	R^2
Horvath & Symonds[31]	Adults	24	1,148	.26	.54	.07
Martin et al[32]	Adults	79	4,770	.22	.45	.05
Horvath & Bedi[33]	Adults	100	5,741	.21	.43	.04
Horvath et al[34]	Adults	190	17,422	.28	.58	.08
Flückiger et al[35]	Adults	295	>30,000	.28	.58	.08
Karver et al[36]	Children and adolescents	43	3,447	.20	.40	.04
Friedlander et al[37]	Couples and families	40	4,113	.30	.62	.08

SMD – standardized mean difference

final outcomes. In this case, it could be said that the alliance is a consequence of the benefits of treatment. This epiphenomenon argument has been proposed as an explanation for the alliance-outcome correlation and to suggest that the alliance may not be an important therapeutic factor[41].

The second conjecture is that early treatment progress is due to remoralization, a tenet of the psychotherapy model proposed by J. Frank[10]. Remoralization is related to the patient taking action to solve his/her problems (i.e., partake in psychotherapy) as well as to the expectation that the treatment will be effective (which is intimately tied to the agreement about the goals of treatment and the acceptance of the therapeutic tasks, and to the unconditional acceptance by a clinician who shows understanding and caring). In the former epiphenomenon case, it is the specific treatment action itself that results in symptom change[42] as well as a strong alliance, whereas in the latter it is the engagement in the therapeutic process and feeling accepted by the clinician that is important[43].

The evidence for these two conjectures partially clarifies their relative validity[44]. The first issue, which has been examined quite extensively, is whether the alliance is predictive of the outcome of therapy beyond the early progress of treatment observed before the alliance was measured. Indeed, there are other processes occurring in therapy prior to alliance measurement that might generate higher alliance ratings and better treatment outcomes, such as adherence to the treatment protocol and therapist competence at delivering the treatment. Moreover, there are several characteristics of patients that might present confounds, such as patient personality, demographics, and context (racial, ethnic or cultural variables), as well as the initial severity of the patient's condition.

Over the years, there have been several attempts to statistically control for patient characteristics and early processes. Recently, a meta-analysis examined studies that partialled out factors occurring before measurement of the alliance and found that the alliance-outcome correlation was not attenuated by these factors[45]. Thus, there is evidence that the alliance is not simply an epiphenomenon of factors occurring before it is measured. However, early symptom change also predicts the final outcomes of therapy[46] and mediates change[47,48], a result which beseeches further investigation of how symptom change and alliance are related over the course of treatment.

An advance in statistical methods has clarified to some extent the alliance-symptom association. The evidence discussed up to now is known as a *between-patient* effect. The alliance-outcome correlation is a bivariate statistic indicating that, with patients for whom the rated alliance is larger than *for other patients*, the outcome is better than *for other patients*. Such statistics say nothing about the temporal aspects of the alliance. An important question is whether the level of the alliance for a particular patient at a particular session is followed subsequently by a reduction in symptoms for that patient. Conversely, is a reduction of symptoms followed by an increase in the rated alliance? Such questions are answered by a *within-patient* analysis[49]. This analysis requires that the two variables are assessed at regular intervals over the course of therapy (i.e., a longitudinal design)[49].

Increasingly, researchers have examined alliance and symptoms over the course of psychotherapy, providing a sufficient number of longitudinal studies to be meta-analytically synthesized[50]. The meta-analysis examined 17 primary studies of the alliance and symptoms over the course of the first phase of treatment, which was designated as the first seven sessions. A between- and within-patient analysis was conducted with the data from each primary study, and the results from the 17 studies were then aggregated, yielding several informative findings.

First, early alliance was related to the level of symptoms at post-treatment, consistent with the meta-analyses reviewed earlier. Second, at the within-patient level, the relative level of the alliance for a patient predicted the subsequent level of symptoms, but as well the relative level of symptoms for a patient predicted subsequent level of the alliance. That is, there is a reciprocal relationship between alliance and symptoms as treatment unfolds during the initial phase. The reciprocal relationship between alliance and symptoms was stronger for patients with stronger alliance relative to other patients, whereas it was stronger for patients with lower symptom level than for other patients. The results of this meta-analysis demonstrate that the alliance is not simply a consequence of symptom improvement, but suggest that symptom improvement and alliance work synergistically.

Whose contribution (therapist or patient?) to the alliance mostly leads to change?

The alliance is a dyadic construct that reflects the interaction between a therapist and a patient. However, each of the participants brings to the therapy situation different capacities to form an alliance[51,52]. Patients have, for example, varying attachment histories, attachment styles, motivation, and needs for affiliation – all these factors may affect the strength of the alliance. Similarly, therapists will differ in their ability to form alliances with patients[51,53]. The correlation of the alliance with outcomes is what is called a *total correlation*[51], in that it ignores that the phenomenon under investigation is due to two sources. When the total correlation is disaggregated, there are two possibilities.

First, it might be the patient contribution to the alliance that is more important for the outcomes of therapy. For example, a patient may have a secure attachment style, lack of stress in life (e.g., adequate economic resources and social support), no comorbid personality disorder, and be motivated to reduce his/her distress. This patient would likely form a good alliance with the therapist and would likely have relatively satisfactory outcomes. If this were the case over a sample of such patients, there would be a positive correlation of alliance with outcomes, and this correlation would be due primarily to the patient's capacity to form an alliance.

On the other hand, if some therapists are able to form better alliances than others, then it could well be that therapists who are able to form strong alliances across a range of patients also produce better outcomes. In this case, there would be a strong total correlation, but this would be mostly due to the therapist contribution to the alliance. Of course, the total correlation could

be due to both the therapist and the patient contribution.

Disaggregating the total correlation into therapist and patient contributions is possible with multilevel modeling, that takes into consideration that the patients (level 1) are nested within therapists (level 2). For example, Baldwin et al[51] disaggregated the total alliance-outcome correlation, which allowed identification of whose contribution to the alliance was mostly associated with outcomes. They examined the outcomes of 331 patients who were treated by 80 therapists. The outcomes of therapy were measured by the Outcome Questionnaire 45 (OQ-45) at baseline and termination, and alliance was measured by the WAI early in therapy from the patients' perspective.

The total correlation of WAI and post-treatment OQ-45 was –.24 (negative because lower OQ scores indicate better outcomes). When the baseline OQ-45 score was included in the model as a covariate, the total correlation was –.21. These total correlations were approximately equal to the values estimated in various meta-analyses[35,45]. Using multilevel models that disaggregated the patient and therapist contribution to the alliance, it was found that the therapist contribution to the alliance predicted outcomes (y_{02} = –0.33, p<0.01), but the patient contribution did not (y_{20} = –0.08, not significant).

The differential effectiveness of therapists has been labeled *therapist effects*[54]. A therapist who generally forms stronger alliances with his/her patients than other therapists also generally has better outcomes than other therapists. However, an apparently surprising result of Baldwin et al's study[51] was that patients with a stronger alliance with that particular therapist did not have better outcomes than the same therapist's other patients with a lower alliance.

To understand this result, consider a chronically depressed patient with a comorbid Cluster B personality disorder, who has a difficult attachment history, an insecure attachment style, and little social support. This patient's alliance with a therapist who generally forms strong alliances will likely be weak relative to the other patients of that therapist. However, this alliance will likely be *stronger than it would have been had this patient been treated by another therapist*. This patient is accustomed to having a chaotic relationship with everyone in his/her world and here is a therapist who is able to form with him/her a relatively stable relationship, albeit less strong than with other patients. This stronger alliance than usual for this patient will generate positive outcomes.

There have been several investigations that have disaggregated the patient and therapist contributions to the alliance, some of which have replicated Baldwin et al's findings and some others have not[54]. However, two meta-analyses have examined the corpus of alliance-outcome correlation by utilizing an innovative method. Del Re et al[55,56] examined several potential moderators of the alliance-outcome correlation, and found that a significant moderator was the patient-to-therapist ratio (i.e., the number of patients in each study divided by the number of therapists). It was found that the lower that ratio, the higher the alliance-outcome correlation. This result, which remained significant even when several potential covariates were controlled, confirms the significance of therapists' impact on the alliance-outcome relationship.

That the benefits of the alliance are mostly due to the therapist contribution raises the fundamental question of what are the characteristics and actions of therapists who form strong alliances across a range of patients. Psychotherapy research has shown that the age, ethnicity, gender, profession of therapist, therapist's theoretical orientation, therapist's experience, size of therapist's caseload, self-reported social skills on a valid inventory, and expert interviewer's rating of trainees' clinical skills, do not differentiate more effective from less effective therapists[54]. The strongest predictor of effectiveness is a set of interpersonal skills of the therapists displayed in interpersonally challenging situations[57,58].

In Anderson et al's study[57], the *facilitative interpersonal skills* of the therapist were the only factor accounting for variability of therapy outcomes. These skills included verbal fluency; therapist communication of hope and positive expectations; persuasiveness; emotional expression; warmth, acceptance and understanding; empathy; alliance bond capacity; and alliance rupture-repair responsiveness. Anderson et al[59,60] as well as others[58] assessed the interpersonal skills of psychotherapy trainees and were able to use these skills to predict therapy outcomes two to five years in the future.

Does the alliance differ among various forms of psychotherapy?

According to Bordin[1], the alliance is important for all healing practices involving a person seeking help and a clinician offering help, although he recognized that the nature of the alliance might be different among the various therapies. Plumbing the depths of the psyche in psychoanalysis might well require a different type of alliance than exposure for a socially anxious patient in cognitive-behavior therapy (CBT), although both tasks can be extremely demanding emotionally.

The most basic question is whether the alliance predicts outcomes across various types of therapy. In their meta-analysis, Flückiger et al[35] examined the size of the correlation for different treatments, including CBT, counseling, psychodynamic therapy, humanistic therapy, interpersonal therapy, and unspecified and eclectic therapies. They found no statistically significant differences in the size of the correlation among the various treatments, which indicates that the *magnitude* of the impact of alliance is high for all psychotherapies. This result is in line with Bordin's suggestion that alliance is vital for change in all psychotherapies, and indeed in all healing practices. However, it is important to examine Bordin's conjecture that the *nature* of the alliance may be different among various treatments.

There are several investigations that shed light on the nature of the alliance in different treatments. Webb et al[61] examined data from two randomized trials of cognitive therapy (CT) for depression, with WAI measured early and later in therapy. Early in therapy, only the agreement on tasks and goals of therapy predicted depression symptom change, whereas the bond factor did not. Later in therapy, the bond factor, as well as the agreement on goals

and tasks, predicted symptom change. These results suggest that in CBT the goals and tasks dimensions of the alliance are more important than the bond dimension in the critical early phase of therapy.

Hagen et al[62] disaggregated the therapist and patient contributions to the alliance in exposure and response prevention treatment for obsessive-compulsive disorder. They found that the therapist contribution to the goals and tasks dimensions predicted outcomes, but the therapist contribution to the bond dimension did not. This result suggests again that the bond dimension is not as important in CBT, but it also corroborates the notion that the therapist contribution to the alliance (here only to the goals and tasks aspects) is what is important to the outcomes of the treatment.

The impact of the bond dimension on the outcome of psychodynamic psychotherapy and of CT for patients with Cluster C personality disorders was investigated by Ulvenes et al[63]. They found that, in psychodynamic psychotherapy, therapist's avoidance of affect negatively influenced symptom reduction and suppressed the relation of bond to that reduction. In contrast, in CT, therapist's avoidance of affect was positively related to both the formation of the bond and to symptom reduction. Thus, the impact of the bond dimension is different in the various forms of psychotherapy, and this dimension interacts with therapeutic actions.

Clearly, the alliance is important across therapies, but exactly how it works in various treatments is complex and needs further investigation.

How are characteristics of the patients related to the alliance-outcome correlation?

Are there patient variables that affect the size of the alliance-outcome correlation? There is reason to expect that the patient's diagnosis might be relevant in this regard. For example, the alliance, which depends on agreement on the goals and tasks of therapy, may not be strong for a patient who is ambivalent about change[64], such as in substance use disorders and eating disorders[65,66]. Furthermore, a patient with attachment difficulties may have problems to form an alliance; therefore, treatment may not progress adequately, unless the relationship with an empathic therapist provides an attachment corrective experience[52] resulting in therapeutic benefits.

Flückiger et al[35] examined the size of the alliance-outcome correlation across various diagnoses and reported several informative findings. For eating disorders, the alliance-outcome correlation was smaller than it was generally (r=.15 vs. r=.28 in general). Some experts in the field have gone so far as to affirm that the alliance is relatively unimportant in the treatment of patients with eating disorders[67]. However, a meta-analysis[68] suggested that the alliance has a stronger relationship to outcomes in younger (vs. older) patients, over and above the variance shared with early symptom improvement, and that early alliance shows a greater association with outcomes in non-behavioral therapies than in those with a strong behavioral component. Clearly, the role of the alliance in the treatment of eating disorders is complex and not well understood.

A second diagnosis where the alliance-outcome is attenuated relative to other diagnoses is substance use disorders (r=.14). Similar to those with eating disorders, patients with substance use disorders may have difficulties to agree on the goals and tasks of therapy. However, there is evidence that adding motivational interviewing to CBT in the presence of ambivalence and resistance to treatment[69,70] can improve the alliance and the outcomes in these patients[71].

Many of the outcome-alliance correlation studies of substance use disorders have been conducted in the US, and the samples contained a high proportion of patients from racial/ethnic minority groups, particularly African Americans. There is evidence that cultural micro-aggressions perceived by the patient during therapy are negatively associated with psychological well-being, and that the alliance mediates this relationship[72]. Here, the alliance may well be the consequence of a therapy process (e.g., perceived cultural micro-aggressions), which leads to a further discussion of the mechanisms involved in the alliance as well as of the therapist actions that may lead to stronger alliances.

A third diagnosis that is theoretically and clinically interesting is personality disorder. In Flückiger et al's meta-analysis[35], the alliance-outcome correlation for borderline personality disorder (r=.32) and other personality disorders (r=.32) was larger than the average correlation across various diagnoses (r=.28), but the differences were not statistically significant. A large variability was observed: the alliance-outcome correlation for borderline personality disorder in the nine relevant studies ranged from r=.00 to r=.78. This variability suggests that the alliance in personality disorder is particularly complex.

It would be informative to examine other characteristics of patients that moderate aspects of the alliance-outcome association. As an example, Zimmermann et al[73] found that the bond feature of the alliance was not predictive of outcomes among patients with sufficient social support, whereas it was a strong predictor in patients with little social support. Further research is clearly warranted in this area.

Are there methodological aspects that affect the size of the alliance-outcome correlation?

There are a number of methodological threats to the validity of the alliance-outcome association. It may well be that the rater of the alliance makes a difference in the size of the correlation. Typically, in the alliance-outcome studies, the outcome measures are rated by the patient, so it might be that, if the patient also rates the alliance, the correlation might be larger because of method variance. However, Flückiger et al's meta-analysis[35] did not find significant differences based on who made the rating, although there was a trend, when observers rated the alliance, for the correlation to be slightly lower. Similarly, there were no differences in the alliance-outcome correlation due to who rated the outcomes. So, it seems that method variance is not a major threat to the validity of the association between the alliance and

outcomes.

We have reported that the alliance measured early in treatment predicts outcomes, which is the typical study method. However, there are studies that measure the alliance mid-treatment or near the end of treatment (e.g., the last three sessions). The correlations for early, mid and late assessment were r=.22, .21 and .30, respectively. It is not surprising that the alliance measured late in therapy is a stronger predictor of outcomes, as variables measured proximally tend to have a larger effect than variables measured distally, regardless of what psychological variables are being assessed. What is important to reiterate here is that the alliance measured early in treatment is predictive of outcomes.

Previously we discussed several alliance measures. Although all of them have demonstrated adequate reliability and validity, it is informative to determine whether the various measures produce different magnitudes of alliance-outcome correlation. Flückiger et al's meta-analysis[35] found no differences in the alliance-outcome correlation among the various alliance assessment instruments. In terms of outcomes, there was a slightly larger alliance-outcome correlation for broader outcome measures, such as quality of life, than for disorder-specific symptom measures. Furthermore, there was no difference in the size of the alliance-outcome correlation depending on whether the data were derived from randomized trials or from naturalistic settings.

It appears that the alliance is a robust predictor of treatment outcomes, regardless of many factors that might have mitigated the size of the correlation. The alliance is associated with outcomes controlling for early symptom change; the level of the alliance at each session predicts subsequent level of symptoms in longitudinal analyses; and the therapist contribution to the alliance predicts outcomes. On the basis of this evidence, it can be argued that the alliance is clearly an important therapeutic factor. Nevertheless, there is a perspicuous limitation to the evidence cited: this evidence heretofore is correlational. It is true that major threats to the causal validity of the alliance have been addressed and adequately ruled out, yet experimental evidence would be needed to bolster a causal relationship between the alliance and outcomes. In psychotherapy, it is unethical to randomly assign patients to levels of the alliance as well as pragmatically difficult to design therapies with different levels of the alliance. However, in medicine and particularly in placebo studies, experimental designs have been used to examine various aspects of the relationship between the clinician and the patient. That evidence will now be reviewed.

ALLIANCE IN MEDICINE AND PLACEBO STUDIES

Up to now our focus has been on the alliance in psychotherapy, but, as Bordin[1] discussed, the alliance is germane to all healing practices that involve a clinician and a patient. The nature of the alliance depends on the particular healing practice. Moreover, various healing practices use the term *alliance* without much thought about the classical definition of the concept.

Our review of research in medicine and placebo studies will demonstrate the importance of the alliance and its generalizability to practices other than psychotherapy. We begin with a general discussion of healing, as this discussion will clarify the role of the alliance in non-psychotherapy contexts.

Natural, specific and contextual effects

When exposed to disease or trauma, human healing is composed of three effects: *natural*, *specific* and *contextual*[74,75].

Biological mechanisms have evolved to protect humans from disease and enable the organism to heal (e.g., blood coagulation, immune functions, barriers such as the skin). Healing that occurs as a result of these defenses is called *natural healing*[75]. *Natural effects* refer to the change in the patient's status due to the natural course of disease as impacted by these defenses.

Specific effects are those due to the particular treatment administered to a patient with a given diagnosis. The medicine or procedure addresses a particular biological deficit or process, resulting in patient cure or improvement. A patient with a gastric ulcer will respond to a course of antibiotics and proton pump inhibitors. Cataract surgery will restore vision, which would have progressively failed without intervention (i.e., natural healing is insufficient in this case). *Specific effects* compose what is generally referred to as modern or Western medicine.

The final component of healing involves *contextual effects*. These effects are due to a number of psychosocial factors, including patient expectations, symbolic meaning of a healing setting (e.g., a physician's white coat, syringes, diplomas on the wall), the relationship between the healer and the patient, and conditioned responses to various medications or procedures[74,76,77]. These psychosocial factors are closely related to the factors that have been identified as generating the placebo response[75,78-80]. However, contextual effects in medicine are not placebo effects, because no placebo has been administered. They have been called *placebo-like effects*[81].

There are two critical points to make here. First, the contextual effects are, to varying degrees, present in all healing practices, including medicine, psychiatry and psychotherapy, contributing to healing experienced by the patient. Second, the alliance is the backbone of the contextual factors – the various contextual factors are, in one way or another, wired to the alliance as conceptualized by Bordin[1].

We now review the literature in medicine that establishes the importance of the relationship for healing. The term alliance is rarely used in this literature and, when it is, it is often misused. Nevertheless, this literature confirms experimentally the importance of the alliance and adds to our understanding of it. We will use the generic term *relationship* and make reference to the alliance for particular studies.

Alliance in somatic medicine

There is a limited number of experimental studies in medicine that have examined variables related to the relationship. This is due to two factors: first, there is little interest in medicine in establishing the importance of the relationship for producing

health outcomes; second, it is difficult to manipulate relationship in medical settings.

In the studies that do examine the relationship in medical settings, this is often discussed as consisting of two components: an *emotional* and a *cognitive* one[76,82]. The emotional component corresponds to the "real relationship" conceptualization of the bond, comprising warmth, empathy and genuineness. The cognitive component is usually described as "information gathering, sharing medical information, patient education, and expectation management"[82, p.1], and is conceptualized as effective communication about the disorder and the treatment.

There is an unstated assumption that an effective communication will lead to belief in the treatment and to belief that the clinician has the technical expertise to produce positive outcomes, which are similar to aspects of the alliance, particularly the emphasis on *agreement on goals* and on the component of *bond* oriented toward the competence of the clinician.

Di Blasi et al[76] found 25 randomized controlled trials (RCTs) exploring the effects of contextual factors, although most of them examined the extent to which the clinicians provided information about the treatment. Clinicians who attempted to influence patient's beliefs about the treatment achieved better outcomes. No studies examined the effects of emotional care only, but four trials evaluated the combination of providing information and emotional care. The results of these studies suggested that providing information in a warm and accepting way produced better health outcomes than a neutral situation. The authors concluded: "Practitioners who attempted to form a warm and friendly relationship with their patients, and reassured them that they would soon be better, were found to be more effective than practitioners who kept their consultations impersonal, formal, or uncertain"[76, p.760].

Kelley et al[82] meta-analyzed medical studies that manipulated the clinician-patient relationship and used validated or objective health outcomes. The results indicated that better relationship conditions produced better health outcomes than poorer relationship, although the effect was small (SMD=0.11). The authors concluded: "This systematic review and meta-analysis of RCTs suggests that the patient-clinician relationship has a small, but statistically significant effect on healthcare outcomes.... relatively few RCTs met our eligibility criteria, and... the majority of these trials were not specifically designed to test the effect of the patient-clinician relationship on healthcare outcomes"[82, p.1].

Thus, the experimental evidence for a relationship effect in medicine is sparse and the quality of evidence available is relatively poor. On the other hand, there are several well-conducted and informative experimental studies of relationship variables using placebos.

Placebos

Placebos are substances or procedures without ingredients that should, from a biological perspective, affect the health status of an individual[83]. They are designed to resemble the *verum* (i.e., the treatment under investigation) in every way except the presence of the therapeutic ingredients. They may consist of sham pills, inoculations, creams or surgery.

Placebos have demonstrated effects on subjective outcomes (e.g., pain ratings) as well as creating physiological changes for a variety of conditions, including pain (acute, chronic as well as experimentally induced), Parkinson's disease, menopausal symptoms, irritable bowel syndrome, headaches, osteoarthritis, respiratory illnesses, and mental disorders (primarily anxiety and depression)[78-80].

The effects of placebos "depend on a person's psychological and brain responses to the *treatment context*, which influence appraisals of future well-being"[78, p.73] (emphasis added). The treatment context includes the relationship between the patient and the clinician, the information about the intervention that is communicated to the patient, the physical healing space, the healing rituals, and cultural beliefs about healing and healers. These psychosocial factors create in the patient the experience of being cared for and understood by the clinician, and the expectation that the treatment delivered by that particular clinician will be effective. Placebo effects can be induced without a face-to-face interaction, say by written materials, or by prior conditioning[77,84-87]. The placebo studies we will review first are those in which aspects of the relationship were experimentally manipulated.

Kaptchuk et al[88] explored if augmenting the therapeutic relationship would increase the placebo response for the treatment of irritable bowel syndrome. The placebo was sham acupuncture (the needles did not pierce the skin although they provided the sensation of doing so). The first arm was usual treatment by the physician, but no sham acupuncture. In the second arm, the patient received sham acupuncture twice a week for three weeks, with the acupuncturist who explained the acupuncture procedure but did not exhibit warmth or caring (called a *limited interaction*). In the third condition, called the *augmented interaction*, the same procedure was implemented, but with a 45 min interaction prior to the first sham acupuncture session, including questions about the patient's symptoms, curiosity about the effects of irritable bowel syndrome on functioning, and inquiries about how the patient understood the cause and meaning of the syndrome. In this condition, the acupuncturist did not provide any advice, treatment or coping strategies.

The results of the study showed that the limited interaction procedure was superior to treatment-as-usual with regard to reduction of symptom severity, relief from distress, global improvement, and quality of life, but the augmented interaction provided additional benefit on all outcomes. According to the authors, "the magnitude of non-specific effects in the augmented arm is not only statistically significant but also clearly clinically significant in the management of irritable bowel syndrome"[88, p.6], supporting the notion that the relationship effect on healing is clinically important. In this study, the actions in the augmented interaction condition resemble those associated with the *bond*, although there were some actions that might be associated with *agreement on goals* (e.g., talking about the symptoms that were distressing).

Notably, a follow-up analysis[89] showed that there were differences between acupuncturists in patient improvements. Indeed, after controlling for treatment condition (augmented vs. limited) and patient characteristics, acupuncturists accounted for an additional 6.9% of the variance in outcomes. In contrast, after controlling for acupuncturist and patient characteristics, treatment condition accounted for 3.0% of outcome variance. So, the effect attributable to different acupuncturists was more than twice as large as the effect attributable to treatment condition (augmented vs. limited), supporting the psychotherapy evidence about the role of the interpersonal skills of the therapist in shaping the alliance-outcome correlation.

In a study of pain intensity and pain sensitivity of patients with chronic back pain, Fuentes et al[90] explored how the "alliance" augmented the effect of both placebo and verum. Patients received either active interferential current therapy (IFC, the verum) or sham IFC in conjunction with either a limited relationship or an enhanced relationship, which the authors labelled as "alliance". In the limited relationship condition, the practitioners introduced themselves and explained the purpose of the treatment, whereas in the other condition "the therapeutic interaction was enhanced through verbal behaviors, including active listening (i.e., repeating the patient's words, asking for clarifications), tone of voice, nonverbal behaviors (i.e., eye contact, physical touch), and empathy"[90, p.480]. Again, the clinician actions were oriented toward the "real relationship" conceptualization of the *bond*. The clinicians left the room during the procedure in the limited relationship condition, but they remained in the enhanced condition. For both the verum and the placebo, the augmented relationship condition produced superior outcomes relative to the limited relationship condition. The authors concluded: "The context in which physical therapy interventions are offered has the potential to dramatically improve therapeutic effects"[90, p.477].

As mentioned previously, there is a conjecture that the therapeutic relationship in medicine is composed of two components, emotional and cognitive[76,82,91]. Howe et al[92] examined physician warmth and perceived competence, two characteristics that map onto the emotional and cognitive components of the relationship. In their study, the participants were given a physical examination, which was explained to the participants as a screen for a subsequent purported medical study. The examination included measurement of vital signs, respiration, as well as a skin prick "allergy test". In actuality, the skin was pricked with histamine, which caused a reaction in all participants. The participants were informed that this outcome disqualified them from the subsequent study, and they were administered a cream, which they were told would attenuate the skin irritation. The cream was a placebo (i.e., contained no antihistamine). These procedures were executed in four conditions: warmth (high vs. low) crossed with competence (high vs. low). High warmth involved an inviting office furnishing (e.g., posters with calming images) and physician use of the participant's name and warm nonverbal behavior (eye contact, proximal seating, and smiling facial features), whereas the low warmth condition did not include these features. In the high competence condition, the physician was verbally fluent (e.g., gave a confident and cogent explanation of various procedures), the tests were administered efficiently without mistakes, and the examination room was well organized, whereas the low competence lacked these features. The diameter of the wheal (circle of irritated tissues) on the skin and the rate of change in diameter were the outcome measures. The wheal diameter decreased most rapidly and the final wheal diameter was smallest in the high warmth/high competence condition, whereas the wheal diameter decreased most slowly and the final wheal diameter was largest in the low warmth/low competence condition. The results of the mismatched conditions (low competence/high warmth and high competence/low warmth) were intermediate between the low/low and high/high conditions, indicating that warmth and competence both contributed to the response to placebo. In this study, the warmth and perceived competence of the clinician affected the physiological response to the administered histamine, experimentally establishing relationship effects.

Czerniak et al[93] manipulated the relationship between healer and recipient in relation to pain tolerance. An actor portraying a physician administered placebo cream to healthy volunteers who participated in a cold-pressor test. In one condition, the "physician" portrayed a traditional doctor-patient relationship and in the other the "physician" role emphasized "attentiveness and strong suggestion, elements… present in ritual healing"[93, p.1]. Pain tolerance was assessed before and after placebo administration. In the enhanced relationship condition, participants showed greater change in pain tolerance after administration. The authors concluded that a "structured manipulation of physician's verbal and non-verbal performance, designed to build rapport and increase faith in treatment, is feasible and may have a significant beneficial effect on the size of the response to placebo analgesia"[93, p.2].

Implications of medical and placebo research for understanding the alliance

The design of the above experimental studies establishes the importance of the relationship in healing. Whereas the previously reviewed alliance-outcome studies were correlational, the placebo studies (and some medical studies) have experimentally manipulated the relationship. Furthermore, placebos are inert and therefore an interaction of the relationship with specific effects is ruled out. Moreover, some of these studies establish that the relationship between healer and patient does not simply have an effect on the patient's subjective experience, as an effect on physiology was also demonstrated (e.g., the size of the wheal created by histamine).

A second consideration is how the relationship in these studies maps onto the alliance. As mentioned, in the medical context, two aspects of the relationship have been emphasized: a) warmth, caring, trust and understanding (emotional component), and b) competence and conveyance of information (cognitive component). These two dimensions need further clarification. Clearly,

the first aspect maps well onto the "real relationship", which to many is the essence of the *bond* feature of the alliance. This aspect has obviously an emotional dimension. The second aspect is not simply conveying information in a clear and cogent manner. The relationship enhances the persuasive salience of the information, thereby influencing the patient to believe that the treatment will effectively remediate distress and restore health[94-96]. In this way, the patient comes to believe that goals can be accomplished through adherence to the recommended actions. Thus, this second aspect of the medical interaction maps onto *agreement on goals* as well as *assignment of tasks* of treatment.

Howe et al[91] authored an article on the above two dimensions of the relationship with the memorable title *When Your Doctor "Gets it" and "Gets You": The Critical Role of Competence and Warmth in the Patient-Provider Interaction*. Actually, the two factors converge with various theoretical and empirical claims, starting from J. Frank's classic discussion of psychotherapy as an example of the universe of healing practices[10,97-99]. The belief in the healing myth and ritual, central to Frank's exposition, is essentially the belief that the clinician understands the nature of the problem, will administer a treatment that will be remedial to the problem, and has the competence to administer the treatment. On the other hand, Frank also discussed the importance of the patient's belief that the clinician understands, cares for, and will make extraordinary efforts to assist him/her (i.e., the *bond* that is created).

Over the years, there have been many relationship concepts discussed in the literature. Recently, Norcross and Lambert[100] published an anthology of meta-analyses on relationship factors in psychotherapy, including the alliance itself (as measured by the instruments discussed earlier), collaboration, goal consensus, empathy, positive regard and affirmation, congruence/genuineness, cultivating positive expectations, real relationship, and treatment credibility, all of which were associated with better outcomes. Clearly, these constructs are not independent, which raises the question about what latent factors underlie the various relationship constructs.

Finsrud et al[101] conducted a study to identify the latent factors of various relationship constructs. In this study, a large sample (N=332) of patients undergoing intensive psychotherapy for a variety of disorders completed at each session a compressive measure of the relationship, with items assessing agreement on goals, agreement on tasks, expectations, treatment credibility, therapist empathy, and perceptions of therapist expertise. The results yielded two factors, which were invariant over the course of treatment and were validated across subsamples. These two factors were described as "confidence in the therapist" and "confidence in the treatment", which mirror the two factors discussed by Howe et al[91] and are consistent with the theoretical positions of Bordin[1], Frank[10,97-99], Horvath and Luborsky[14], and Wampold[44,102].

It appears that the alliance is not distinct from other relationship concepts that have been discussed and investigated. As well, the various relationship constructs, including the alliance, might best be considered as being composed of two factors: being cared for and understood by the clinicians (corresponding to Bordin's *bond*), and belief in the competence of the therapist to select and administer an effective treatment (corresponding to Bordin's *agreement on goals* and therapist's *assignment of tasks*).

We have previously reported the evidence suggesting that in psychotherapy the benefits of the alliance are mostly due to the therapist contribution, in particular the *facilitative interpersonal skills* of the therapist[57]. This has been confirmed in healing contexts other than psychotherapy. In the context of a double-blind RCT[103,104], psychiatrists administered either an antidepressant or placebo "plus minimal supportive therapy", which involved a warm, empathic and caring atmosphere, but no advice or coping strategies. The antidepressant was found to be superior to placebo, accounting for about 3% of the variability in outcomes[104]. However, differences in outcomes due to psychiatrists themselves accounted for about 9% of that variability[105]. The more effective psychiatrists delivering placebo had better outcomes than the less effective psychiatrists delivering antidepressant medication. Because this was a double-blind RCT, the difference among the psychiatrists was likely due to what took place in the clinical management, supporting the role of clinicians' interpersonal skills.

Alliance in other contexts and beyond the therapist-patient dyad

There is evidence to support the idea that face-to-face interaction is not needed to develop a collaborative relationship. For example, various Internet-based therapies have been developed, most of which are variations of CBT (IBCT)[106]. These therapies involve the following components. First, the patient is screened to ensure that his/her problem is consistent with the goals and tasks of the treatment. Second, the therapist, through asynchronous text messages, orients the patient to the program, describing the sequence of modules to be completed. The modules mirror the components of the CBT for the particular disorder. Third, after each module is completed, the patient answers an essay question, and the therapist provides a brief personalized comment on patient progress (although there are efforts to use artificial intelligence to provide this feedback). Meta-analytic evidence indicates that IBCT is as effective as face-to-face CBT for various psychiatric and somatic conditions[106].

In these Internet-based therapies, the assessed alliance between the patient and the clinician/program, despite the distal and short interaction, is reported to be correlated with outcomes. For example, Zalaznik et al[107], examining the alliance with the program and with the therapist in ICBT for panic disorder, found that patient-rated alliance with the program predicted treatment outcomes, whereas alliance with the therapist predicted adherence to treatment. There have been two meta-analyses of the association of the alliance and outcomes in electronically mediated treatments, and both detected an effect comparable to face-to-face psychotherapy[35,108].

The findings with Internet-based therapies suggest that the concept of alliance extends beyond the individual clinician and applies to a program or treatment and the context in which it is implemented. A patient's belief that the treatment will be effec-

tive for the disease or distress he/she is experiencing (*agreement about goals and tasks of treatment*) seems to be forged by multiple factors other than the clinician.

This system perspective is supported by other mental health care findings. Wampold and Brown[109] studied the variability of outcomes due to psychotherapists in a naturalistic study in managed care. Consistent with the previously reported therapist effects studies, about 5% of the variability in the outcomes was due to the therapists: some of them consistently achieved better outcomes than others. Of these therapists, fifteen had 586 patients who began pharmacotherapy with a psychiatrist. A remarkable finding was that the patients of the most effective psychotherapists had the largest medication effects, even though the psychotherapists had no or little contact with the psychiatrist. Thus, the relationship between the patient and the psychotherapist, and the expectations for medication that were created therein, affected the outcomes of care from a different mental health professional.

Further evidence for system effects comes from a meta-analysis by Falkenström et al[110], based on 19 studies that examined the variability in the outcomes of mental health treatments due to organizational differences. They found that "all studies showed some evidence for organization effects, and there was some evidence for *organizational climate and culture* explaining differences in outcome"[110, p.76] (emphasis added).

The alliance, and in particular its component related to confidence in the treatment, is influenced by many contextual variables. The relationship between the clinician and the patient is the most proximal place for the alliance to be formed. This level of understanding has attracted the greatest attention, theoretically, clinically and empirically. However, the context where the treatment takes place also contributes to the alliance.

It has been speculated that a high prestige clinic will increase belief in the efficacy of a treatment[86]. There is also evidence that the climate and culture of the clinic matter, most likely at least in part by creating an organization where therapists can thrive[111]. Furthermore, it is a mistake to assume that the treating clinician is the only influencer in such organizations. Patients interact on the phone, through email, and in person with non-clinician staff. Do these interactions communicate warmth, caring, respect as well as competence? As well, how a patient perceives a clinician and the treatment being offered may well depend on the patient's interaction with other clinicians.

It is important to consider the context in which a treatment is delivered, with attention to the alliance of the patient with other clinicians and the clinic staff, as well as to aspects of the physical space and clinic reputation. Mental health services are increasingly being delivered electronically, and patients use various Internet-based mediated services not involving a face-to-face interaction with a clinician; nevertheless, as the research suggests, the alliance with the program and a presumed clinician is critical to the optimal effectiveness of such programs. Clearly, more research into how consideration of the alliance in such programs can improve outcomes is needed.

MECHANISMS OF THE ALLIANCE AND CLINICAL ACTIONS

We will discuss now how the alliance might be healing and what might promote clinically a strong alliance. We describe three pathways to healing, each involving the alliance, which are shown in Figure 1.

The caring, attentive, real and empathic (CARE) pathway

The CARE pathway has been described in several ways. In Bordin's[1] conceptualization of the alliance, this pathway is described as the *bond*. In the medical literature, it is often called the *emotional* component of the relationship[76,83]. In placebo studies, the terms *warmth*[90] and *interpersonal healing*[75] have been used. The therapist actions associated with this pathway have been labeled as

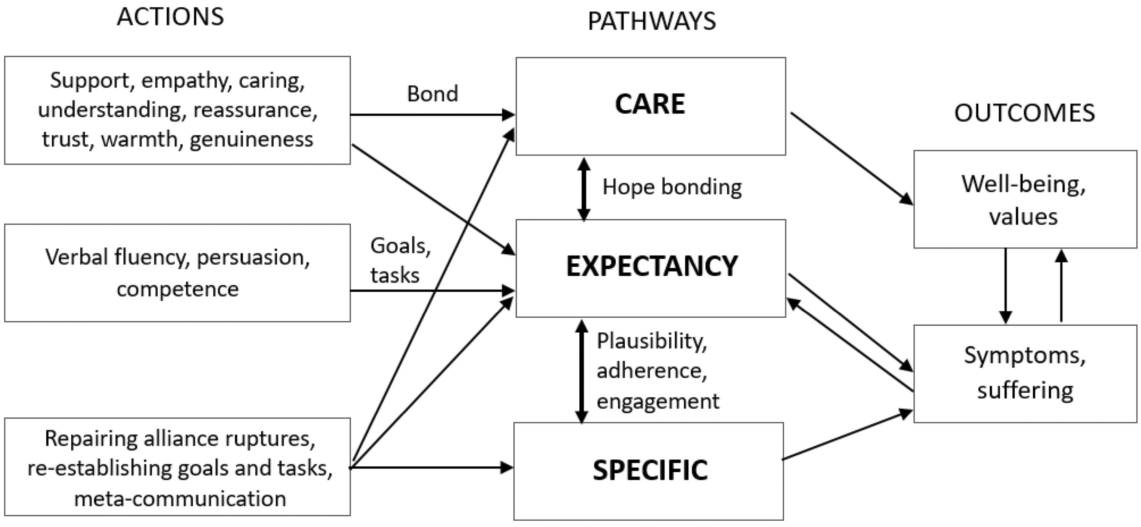

Figure 1 Three pathways to healing involving the alliance

support, empathy, reassurance, warmth, caring, and non-transference-based real relationship, among others. The question is: what about these therapist actions leads to healing? Here we tentatively suggest a few mechanisms that underlie this pathway to healing.

When patients present to a clinician for treatment, they often experience emotional distress that originates from the disorder, disease or injury. A pain in the gut may create fear of cancer; a diagnosis of Parkinson's disease may lead to depression due to an understanding of the progressive nature of the illness. The clinician, through his/her empathic and reassuring behavior, reduces the patient's emotional distress.

Humans are a social species, and rely on the assistance of others for survival[18,94,112,113]. Individuals without adequate social support and connection will not flourish, particularly when under threat. Lack of exercise, smoking, obesity, excessive drinking, and environmental pollution increase the risk of morbidity and mortality; interestingly, loneliness is a greater risk for mortality than any of these factors[114,115]. A warm and understanding clinician may well provide emotional support to patients who lack social connection, perceive themselves as lonely, or who feel that those close to them do not understand their problems. In mental health care, the clinician – with some exceptions – is available, in an understanding way, at each and every session, regardless of what the patient discloses and however shameful, fearful or difficult the material may be. With increased pressure to expand services, the time spent with each patient is becoming shorter, which increases the need to focus on the relationship.

Patients' emotional dysregulation negatively affects mental and physical health, and consequently several mental health treatments are focused on reducing this dysregulation. In these interventions, the locus is typically the patient. For example, meditation is predicated on assisting the patient regulate his/her emotions. However, there is evidence that emotion regulation is an unconscious dyadic process, in that the presence of an intimate other can attenuate arousal and distress through a process that is referred to as co-regulation, social regulation, or interpersonal emotion regulation[116-118]. Dyadic emotion regulation "refers to the process by which relationship partners form a dyadic emotional system involving an oscillating pattern of affective arousal and dampening that dynamically maintains an optimal emotional state"[116, p.202].

Co-regulation between intimates has been investigated experimentally. In a study of maritally satisfied women, it was found that holding the hand of their husbands reduced arousal in a stressful situation in comparison to holding the hand of a stranger or not holding anyone's hand; furthermore, the more maritally satisfied the women were, the greater the effect[119]. In psychotherapy, interpersonal co-regulation has been detected in moment-to-moment emotional states of the patient and therapist[120,121]. Indeed, the beneficial effects of empathy in medicine have been attributed to co-regulation[74,91,122].

The CARE pathway is not focused on particular patient problems and should have its effect primarily on the general well-being of the patient. This was evident in the study on irritable bowel syndrome we discussed earlier, as the largest effect of the enhanced therapeutic relationship was on the quality of life outcome[88].

The EXPECTANCY pathway

Expectations have a strong influence on our experience of the world, particularly our expectations of our internal sensations, both physical and mental[78,79,123]. For example, taste aversions, which have evolved to protect organisms from ingesting harmful substances and which are easily conditioned, can be influenced in humans by expectations[124,125].

The influence of expectations on well-being is established most persuasively in the placebo literature, where placebo administration influences health outcomes. Placebos "depend on a person's psychological and brain responses to the *treatment context*, which influence appraisals of future well-being"[78, p.73] (emphasis added). The effects of placebos on mental disorders are well documented[126]. The EXPECTANCY pathway will affect primarily symptoms (or, more accurately, it will affect the purported outcomes of the treatment on which the clinician and patient agree).

There are many ways to acquire expectations. As discussed earlier, placebo effects can be generated without face-to-face interactions[84-86]. However, an effective and efficient way to create expectations is through verbal persuasion[95]. The verbal transmission of information about healthy behaviors is important in everyday life, as well as in health settings. Wampold[74] describes how the expectation that inserting a metal object into an electrical socket will create a painful shock is unlikely to have been acquired through classical conditioning or vicarious learning. Most people have learned not to insert metal objects into electrical sockets by being told by someone they trust, most likely a parent, that this was a dangerous practice.

Indeed, as Lieberman[94] pointed out, "our brains are designed to be influenced by others". That is, patients are wired to believe in the explanations provided by a clinician, particularly if the clinician is perceived to be competent and expert and the patient trusts that the therapist is acting in his/her best interest. As shown in Figure 1, expectations are created by both the "warmth" and the "competence" dimensions of Howe et al's conceptualization[91]. Attention to how the clinician informs the patient about the disorder and the persuasiveness of the explanation of the treatment to be delivered are critical aspects of mental health care.

The SPECIFIC pathway

To varying degrees, the specific ingredients of mental health and in general medical treatments have an effect on the disorders. For both psychotropic medications and psychotherapies, there is a debate about the size of this effect[44,127,128]. This debate is orthogonal to the discussion of the alliance, as the alliance is necessary in most cases for the specific effects to occur. Without an agreement about the goals and tasks of therapy as well as a trusting relationship, the patient is unlikely, or at least less likely, to be engaged in and adhere to the treatment.

In medicine, there is evidence that physician's communication is associated with patient's adherence[129,130]. In the schema of Figure 1, it is conjectured that expectations partially mediate the relationship between clinician's actions and the specific ef-

fects. Agreement about the tasks of therapy implies that the patient believes that the treatment will be effective, which is essentially expectations.

There is one complication of the distinction between specific effects and the alliance, not emphasized heretofore. To this point the alliance has been treated as a static entity – measurement of the alliance at a particular point in time is associated with symptoms, say at another time. However, the alliance is not stationary, but rather oscillates over the course of a session, between sessions, and over the course of therapy. Relational psychodynamic approaches to psychotherapy consider the alliance a specific effect, in that the development of the alliance over the course of therapy is therapeutic in and of itself[52,131-133]. The primary mechanism is that disordered relationships underlie mental disorders and that the creation of a strong relationship with the therapist is reparative. Moreover, according to this school, there will be inevitable relationship disruptions in therapy, often called "ruptures", due to the difficult work, and addressing these issues is therapeutic, as it models how strong interpersonal bonds are negotiated.

Whether one agrees with this approach or not, it is clear that addressing ruptures in the alliance is critical, as unaddressed problems will lead to decrements in the bond and in agreement about the goals and tasks of therapy. There is relatively strong meta-analytic evidence that "repairing ruptures" in psychotherapy is associated with better outcomes[131]. Such repairs can be addressed by renegotiating the goals and tasks of therapy or by meta-communication about the patient-clinician relationship[131,133].

Interdependence of pathways

In the previous discussions of the alliance and how it relates to outcomes, it is clear that there are reciprocal and interdependent effects. For example, over time alliance predicts subsequent symptoms, and level of symptoms predicts the alliance[50]. As well, expectations reflected by agreement on goals and tasks predict final outcomes, but alliance mediates the effects of outcome expectations at the beginning of therapy and final outcomes[48]. Feeling cared for and understood by a trustworthy clinician will increase expectations. In Figure 1, we have shown various recursive effects. The pathways to healing are presented as a means to understand the complexity of how the alliance can be therapeutic.

CONCLUSIONS

The alliance, a concept that originated with Bordin's[1] discussion in 1979, has been generally accepted and empirically established in psychotherapy, and, as Bordin predicted, is now acknowledged as a therapeutic factor in any healing setting. A patient who has a warm, understanding, caring and empathic clinician, and who perceives that the treatment offered by the clinician will effectively remediate distress and restore health, will have better treatment outcomes. Understanding how the alliance works and using the interpersonal skills needed to produce a strong alliance will improve outcomes, in psychotherapy, in other mental health care, and most likely in all healing contexts.

Despite the rather extensive research on the alliance, there are a number of areas that need further exploration. There is evidence to suggest that a set of facilitative interpersonal skills demonstrated by the therapist in challenging interpersonal situations creates stronger alliances and better outcomes. However, there is a need for further research on how these skills should be applied in different contexts as well as with different patients. It is important to be cognizant that some patients will respond to the same therapeutic action differently. A patient with attachment difficulties, who has difficulty decoding emotions in interpersonal situations, or who is culturally different from the clinician, may respond in ways different from what the clinician routinely expects.

It was beyond the scope of this paper to discuss whether the interpersonal skills are *born* or *made*. There is evidence that the interpersonal skills of psychotherapy trainees at the beginning of training predict outcomes several years later[58,60], suggesting that these skills are formed before an individual receives training for professional practice. However, from studies of expert performance[134,135], there is also evidence that therapists can deliberately practice interpersonal skills and improve performance[111,136-138].

Research and clinical attention have mostly focused on the alliance between the clinician and the patient in face-to-face interactions. However, there is preliminary evidence concerning the alliance of patients with other clinic staff, systems of care, or the program in Internet mediated services. Those involved in the design and delivery of mental health services, whether in person or delivered electronically, should attend to how the alliance can be strengthened in ways that improve the quality of care. Education and training of mental health professionals need to incorporate deliberate efforts to utilize what is known about the alliance, in order to foster the development of the interpersonal skills necessary for these professionals to form strong alliances across a range of patients.

REFERENCES

1. Bordin ES. The generalizability of the psychoanalytic concept of the working alliance. Psychotherapy 1979;16:252-60.
2. Sterba R. The fate of the ego in analytic therapy. Int J Psychoanal 1934;15:117-26.
3. Menninger KA. Theory of psychoanalytic technique. New York: Basic Books, 1958.
4. Greenson RR. The technique and practice of psychoanalysis. New York: International Universities, 1967.
5. Gelso CJ. The real relationship in a postmodern world: theoretical and empirical explorations. Psychother Res 2009;19:253-64.
6. Oddli HW, McLeod J, Nissen-Lie HA et al. Future orientation in successful therapies: expanding the concept of goal in the working alliance. J Clin Psychol 2021;77:1307-29.
7. Oddli HW, McLeod J, Reichelt S et al. Strategies used by experienced therapists to explore client goals in early sessions of psychotherapy. Eur J Psychother Couns 2014;16:245-66.
8. Yulish NE, Goldberg SB, Frost ND et al. The importance of problem-focused treatments: a meta-analysis of anxiety treatments. Psychotherapy 2017;54:321-38.
9. Wampold BE. The great psychotherapy debate: model, methods, and findings. Mahwah: Lawrence Erlbaum, 2001.
10. Frank JD, Frank JB. Persuasion and healing: a comparative study of psycho-

therapy, 3rd ed. Baltimore: Johns Hopkins University, 1991.
11. Constantino MJ, Vîslă A, Coyne AE et al. A meta-analysis of the association between patients' early treatment outcome expectation and their posttreatment outcomes. Psychotherapy 2018;55:473-85.
12. Rogers CR. Client-centered therapy. Boston: Houghton Mifflin, 1951.
13. Hatcher RL, Barends AW. How a return to theory could help alliance research. Psychotherapy 2006;43:292-9.
14. Horvath AO, Luborsky L. The role of the therapeutic alliance in psychotherapy. J Consult Clin Psychol 1993;61:561-73.
15. Luborsky L. Helping alliance in psychotherapy. In: Cleghhorn JL (ed). Successful psychotherapy. New York: Brunner/Mazel, 1976:92-116.
16. Horvath AO. Research on the alliance: knowledge in search of a theory. Psychother Res 2017;28:499-516.
17. Kenny DA. Interpersonal perception: the foundation of social relationships. New York: Guilford, 2020.
18. Baumeister RF. The cultural animal: human nature, meaning, and social life. New York: Oxford University Press, 2005.
19. Falconier MK, Jackson JB, Hiplert P et al. Dyadic coping and relationship satisfaction: a meta-analysis. Clin Psychol Rev 2015;42:28-46.
20. Malloy TE. Social relations modeling of behavior in dyads and groups. London: Academic Press, 2018.
21. Kenny DA, Judd CM. Consequences of violating the independence assumption in analysis of variance. Psychol Bull 1986;99:422-31.
22. Huppert JD, Kivity Y, Cohen L et al. A pilot randomized clinical trial of cognitive behavioral therapy versus attentional bias modification for social anxiety disorder: an examination of outcomes and theory-based mechanisms. J Anxiety Disord 2018;59:1-9.
23. Marmar CR, Horowitz MJ, Marziali E. The development of the therapeutic alliance rating system. In: Greenberg LS, Pinsof WM (eds). Psychotherapeutic process: a research handbook. New York: Guilford, 1986:367-90.
24. Alexander LB, Luborsky L. The Penn Helping Alliance Scales. In: Greenberg LS, Pinsof WM (eds). Psychotherapeutic process: a research handbook. New York: Guilford, 1986:325-56.
25. Suh CS, Strupp HH, O'Malley SS. The Vanderbilt process measures: the Psychotherapy Process Scale (VPPS) and the Negative Indicators Scale (VNIS), In: Greenberg LS, Pinsof WM (eds). Psychotherapeutic process: a research handbook. New York: Guilford, 1986:285-323.
26. Horvath AO, Greenberg LS. Development and validation of the Working Alliance Inventory. J Couns Psychol 1989;36:223-33.
27. Flückiger C, Horvath AO, Brandt H. The evolution of patients' concept of the alliance and its relation to outcome: a dynamic latent-class structural equation modeling approach. J Couns Psychol 2022;69:51-62.
28. Hoyt WT, Melby JN. Dependability of measurement in counseling psychology: an introduction to generalizability theory. Couns Psychol 1999;27:325-52.
29. Hamovitch EK, Choy-Brown M, Stanhope V. Person-centered care and the therapeutic alliance. Community Ment Health J 2018;54:951-8.
30. Horvath AO. An exploratory study of the working alliance: its measurement and relationship to therapy outcome. EdD Thesis, University of British Columbia, Vancouver, 1981.
31. Horvath AO, Symonds BD. Relation between working alliance and outcome in psychotherapy: a meta-analysis. J Couns Psychol 1991;38:139-49.
32. Martin DJ, Garske JP, Davis MK. Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review. J Consult Clin Psychol 2000;68:438-50.
33. Horvath AO, Bedi RP. The alliance. In: Norcross JC (ed). Psychotherapy relationships that work: therapist contributions and responsiveness to patients. New York: Oxford University Press, 2002:37-69.
34. Horvath AO, Del Re AC, Flückiger C et al. Alliance in individual psychotherapy. Psychotherapy 2011;48:9-16.
35. Flückiger C, Del Re AC, Horvath AO et al. The alliance in adult psychotherapy: a meta-analytic synthesis. Psychotherapy 2018;55:316-40.
36. Karver MS, De Nadai AS, Monahan M et al. Meta-analysis of the prospective relation between alliance and outcome in child and adolescent psychotherapy. Psychotherapy 2018;55:341-55.
37. Friedlander ML, Escudero V, Welmers-van de Poll M et al. Meta-analysis of the alliance-outcome relation in couple and family therapy. Psychotherapy 2018;55:356-71.
38. Alldredge CT, Burlingame GM, Yang C et al. Alliance in group therapy: a meta-analysis. Group Dyn 2021;25:13-28.
39. Owen J, Adelson J, Budge S et al. Trajectories of change in psychotherapy. J Clin Psychol 2015;71:817-27.
40. Baldwin SA, Berkeljon A, Atkins DC et al. Rates of change in naturalistic psychotherapy: contrasting dose-effect and good-enough level models of change. J Consult Clin Psychol 2009;77:203-11.
41. DeRubeis RJ, Brotman MA, Gibbons JC. A conceptual and methodological analysis of the nonspecifics argument. Clin Psychol 2005;12:174-83.
42. Tang TZ, DeRubeis RJ. Reconsidering rapid early response in cognitive behavioral therapy for depression. Clin Psychol 1999;6:283-8.
43. Ilardi SS, Craighead WE. The role of nonspecific factors in cognitive-behavior therapy for depression. Clin Psychol 1994;1:138-56.
44. Wampold BE, Imel ZE. The great psychotherapy debate: the evidence for what makes psychotherapy work, 2nd ed. New York: Routledge, 2015.
45. Flückiger C, Del Re AC, Wlodasch D et al. Assessing the alliance-outcome association adjusted for patient characteristics and treatment processes: a meta-analytic summary of direct comparisons. J Couns Psychol 2020;67:706-11.
46. Beard JIL, Delgadillo J. Early response to psychological therapy as a predictor of depression and anxiety treatment outcomes: a systematic review and meta-analysis. Depress Anxiety 2019;36:866-78.
47. Baier AL, Kline AC, Feeny NC. Therapeutic alliance as a mediator of change: a systematic review and evaluation of research. Clin Psychol Rev 2020;82:101921.
48. Constantino MJ, Coyne A. Goodwin BJ et al. Indirect effect of patient outcome expectation on improvement through alliance quality: a meta-analysis. Psychother Res 2021;31:711-25.
49. Curran PJ, Bauer DJ. The disaggregation of within-person and between-person effects in longitudinal models of change. Annu Rev Psychol 2011;62:583-619.
50. Flückiger C, Rubel JA, Del Re AC et al. The reciprocal relationship between alliance and early treatment symptoms: a two-stage individual participant data meta-analysis. J Consult Clin Psychol 2020;88:829-43.
51. Baldwin SA, Wampold BE, Imel ZE. Untangling the alliance-outcome correlation: exploring the relative importance of therapist and patient variability in the alliance. J Consult Clin Psychol 2007;75:842-52.
52. Zilcha-Mano S. Is the alliance really therapeutic? Revisiting this question in light of recent methodological advances. Am Psychol 2017;72:311-25.
53. Baldwin SA, Imel ZE. Therapist effects: finding and methods. In: Lambert MJ (ed). Bergin and Garfield's handbook of psychotherapy and behavior change. New York: Wiley, 2013:258-97.
54. Wampold BE, Owen J. Therapist effects: history, methods, magnitude, and characteristics of effective therapists. In: Barkham M, Lutz W, Castonguay LG (eds). Bergin and Garfield's handbook of psychotherapy and behavior change. New York: Wiley, 2021:301-30.
55. Del Re AC, Flückiger C, Horvath AO et al. Therapist effects in the therapeutic alliance-outcome relationship: a restricted-maximum likelihood meta-analysis. Clin Psychol Rev 2012;32:642-9.
56. Del Re AC, Flückiger C, Horvath AO et al. Examining therapist effects in the alliance-outcome relationship: a multilevel meta-analysis. J Consult Clin Psychol 2021;89:371-8.
57. Anderson T, Ogles BM, Patterson CL et al. Therapist effects: facilitative interpersonal skills as a predictor of therapist success. J Clin Psychol 2009;65:755-68.
58. Schöttke H, Flückiger C, Goldberg SB et al. Predicting psychotherapy outcome based on therapist interpersonal skills: a five-year longitudinal study of a therapist assessment protocol. Psychother Res 2017;27:642-52.
59. Anderson T, Crowley MEG, Himawan L et al. Therapist facilitative interpersonal skills and training status: a randomized clinical trial on alliance and outcome. Psychother Res 2016;26:511-29.
60. Anderson T, McClintock AS, Himawan L et al. A prospective study of therapist facilitative interpersonal skills as a predictor of treatment outcome. J Consult Clin Psychol 2016;84:57-66.
61. Webb CA, DeRubeis RJ, Amsterdam JA et al. Two aspects of the therapeutic alliance: differential relations with depressive symptom change. J Consult Clin Psychol 2011;79:279-83.
62. Hagen K, Solem S, Opstad HB et al. Therapist variability in the task/goal dimension of the early working alliance predicts outcome in exposure and response prevention treatment for obsessive-compulsive disorder. Clin Neuropsychiatry 2016;13:94-9.
63. Ulvenes PG, Berggraf L, Hoffart A et al. Different processes for different therapies: therapist actions, therapeutic bond, and outcome. Psychotherapy 2012;49:291-302.
64. Krebs P, Norcross JC, Nicholson JM et al. Stages of change and psychotherapy outcomes: a review and meta-analysis. J Clin Psychol 2018;74:1964-79.
65. Vrabel KR, Ulvenes PG, Wampold BE. Alliance and symptom improvement in inpatient treatment for eating disorder patients: a study of within-patient

processes. Int J Eat Disord 2015;48:1113-21.
66. Flückiger C, Del Re AC, Horvath AO et al. Substance use disorders and racial/ethnic minorities matter: a meta-analytic examination of the relation between alliance and outcome. J Couns Psychol 2013;60:610-6.
67. Brown A, Mountford VA, Waller G. Is the therapeutic alliance overvalued in the treatment of eating disorders? Int J Eat Disord 2013;46:779-82.
68. Graves TA, Tabri N, Thompson-Brenner H et al. A meta-analysis of the relation between therapeutic alliance and treatment outcome in eating disorders. Int J Eat Disord 2017;50:323-40.
69. Miller WR. Sacred cows and greener pastures: reflections from 40 years in addiction research. Alcohol Treat Q 2016;34:92-115.
70. Moyers TR, Miller WR, Hendrickson SML. How does motivational interviewing work? Therapist interpersonal skill predicts client involvement within motivational interviewing sessions. J Consult Clin Psychol 2005;73:590-8.
71. Westra HA, Norouzian N. Using motivational interviewing to manage process markers of ambivalence and resistance in cognitive behavioral therapy. Cogn Ther Res 2018;42:193-203.
72. Owen J, Imel Z, Tao KW et al. Cultural ruptures in short-term therapy: working alliance as a mediator between clients' perceptions of microaggressions and therapy outcomes. Couns Psychother Res 2011;11:204-12.
73. Zimmermann D, Wampold BE, Rubel JA et al. The influence of extra-therapeutic social support on the association between therapeutic bond and treatment outcome. Psychother Res 2021;31:726-36.
74. Wampold BE. Healing in a social context: the importance of clinician and patient relationship. Front Pain Res 2021;2:684768.
75. Miller FG, Colloca L, Kaptchuk TJ. The placebo effect: illness and interpersonal healing. Perspect Biol Med 2009;52:518-39.
76. Di Blasi Z, Harkness E, Ernst E et al. Influence of context effects on health outcomes: a systematic review. Lancet 2001;357:757-62.
77. Moerman DE, Jones WB. Deconstructing the placebo effect and finding the meaning response. Ann Int Med 2002;136:471-6.
78. Ashar YK, Chang LJ, Wager TD. Brain mechanisms of the placebo effect: an affective appraisal account. Annu Rev Clin Psychol 2017;13:73-98.
79. Benedetti F. Placebo effects: understanding the other side of medical care, 3rd ed. Oxford: Oxford University Press, 2021.
80. Price DP, Finniss DG, Benedetti F. A comprehensive review of the placebo effect: recent advances and current thought. Annu Rev Psychol 2008;59:565-90.
81. Benedetti F. Mechanisms of placebo and placebo-related effects across diseases and treatments. Annu Rev Pharmacol Toxicol 2008;48:33-60.
82. Kelley JM, Kraf-Todd G, Schapira L et al. The influence of the patient-clinician relationship on healthcare outcomes: a systematic review and meta-analysis of randomized controlled trials. PLoS One 2014;9:e94207.
83. Grünbaum A. The placebo concept. Behav Res Ther 1981;19:157-67.
84. Kam-Hansen S, Jakubowski N, Kelly JM et al. Altered placebo and drug labeling changes the outcome of episodic migraine attacks. Sci Transl Med 2014;6:218ra5.
85. Gaab J, Kossowski J, Ehlert U et al. Effects and components of placebos with a psychological treatment rationale – three randomized-controlled studies. Sci Rep 2019;9:1421.
86. Liberman BL. The role of mastery in psychotherapy: maintenance of improvement and prescriptive change. In: Frank JD, Hohen-Saric R, Imber SD et al (eds). Effective ingredients of successful psychotherapy. Baltimore: Johns Hopkins University, 1978:35-72.
87. Colloca L, Miller FG. How placebo responses are formed: a learning perspective. Philos Trans R Soc Lond B Biol Sci 2011;366:1859-69.
88. Kaptchuk TJ, Kelly JM, Conboy LA et al. Components of placebo effect: randomised controlled trial in patients with irritable bowel syndrome. BMJ 2008;336:999-1003.
89. Kelley JM, Lembo AJ, Ablon JS et al. Patient and practitioner influences on the placebo effect in irritable bowel syndrome. Psychosom Med 2009;71:789-97.
90. Fuentes J, Armijo-Olivo S, Funabashi M et al. Enhanced therapeutic alliance modulates pain intensity and muscle pain sensitivity in patients with chronic low back pain: an experimental controlled study. Phys Ther 2014;94:477-89.
91. Howe LC, Leibowitz KA, Crum AJ. When your doctor "gets it" and "gets you": the critical role of competence and warmth in the patient-provider interaction. Front Psychiatry 2019;10:475.
92. Howe LC, Goyer JP, Crum AJ. Harnessing the placebo effect: exploring the influence of physician characteristics on placebo response. Health Psychol 2017;36:1074-82.
93. Czerniak E, Biegon A, Ziv A et al. Manipulating the placebo response in experimental pain by altering doctor's performance style. Front Psychol 2016;7:874.
94. Lieberman MD. Social: why our brains are wired to connect. New York: Crown Publishing Group, 2013.
95. Geers AL, Briñol P, Vogel A et al. The application of persuasion theory to placebo effects. Int Rev Neurobiol 2018;138:113-36.
96. Petty RE, Briñol P. The elaboration likelihood and metacognitive models of attitudes. Implications for prejudice, the self, and beyond. In: Sherman JW, Gawronski B, Trope Y (eds). Dual-process theories of the social mind. New York: Guilford, 2013:172-87.
97. Frank JB, Wampold BE. Persuasion and healing, 4th ed. Baltimore: Johns Hopkins University (in press).
98. Frank JD. Persuasion and healing: a comparative study of psychotherapy, 2th ed. Baltimore: Johns Hopkins University, 1973.
99. Frank JD. Persuasion and healing: a comparative study of psychotherapy. Baltimore: Johns Hopkins University, 1961.
100. Norcross JC, Lambert MJ (eds). Psychotherapy relationships that work, Vol. 1: Evidence-based therapist contributions, 3rd ed. New York: Oxford University Press, 2019.
101. Finsrud I, Nissen-Lie, HA, Vrabel K et al. It's the therapist and the treatment: the structure of common therapeutic relationship factors. Psychother Res 2022;32:139-50.
102. Wampold BE. How important are the common factors in psychotherapy? An update. World Psychiatry 2015;14:270-7.
103. Elkin I, Parloff MB, Hadley et al. NIMH Treatment of Depression Collaborative Research Program: background and research plan. Arch Gen Psychiatry 1985;42:305-16.
104. Elkin I, Shea MT, Watkins JT et al. National Institute of Mental Health Treatment of Depression Collaborative Research Program. General effectiveness of treatments. Arch Gen Psychiatry 1989;46:971-82.
105. McKay KM, Imel ZE, Wampold BE. Psychiatrist effects in the psychopharmacological treatment of depression. J Affect Disord 2006;16:236-42.
106. Carlbring P, Andersson G, Cuijpers P et al. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018;47:1-18.
107. Zalaznik D, Strauss AY, Halaj A et al. Patient alliance with the program predicts treatment outcomes whereas alliance with the therapist predicts adherence in internet-based therapy for panic disorder. Psychother Res 2021;31:1022-35.
108. Kaiser J, Hanschmidt F, Kersting A. The association between therapeutic alliance and outcome in internet-based psychological interventions: a meta-analysis. Comp Hum Behav 2021;114:106512.
109. Wampold BE, Brown GS. Estimating therapist variability: a naturalistic study of outcomes in managed care. J Consult Clin Psychol 2005;73:914-23.
110. Falkenström F, Grant J, Holmqvist R. Review of organizational effects on the outcome of mental health treatments. Psychother Res 2018;28:76-90.
111. Goldberg SB, Babins-Wagner R, Rousmaniere T et al. Creating a climate for therapist improvement: a case study of an agency focused on outcomes and deliberate practice. Psychotherapy 2016;53:367-75.
112. Wilson EO. On human nature. Cambridge: Harvard University Press, 1978.
113. Wilson EO. The social conquest of earth. New York: Liveright, 2012.
114. Cacioppo S, Cacioppo JT. Decoding the invisible forces of social connections. Front Integr Neurosci 2012;6:51.
115. Holt-Lunstad J, Robles TF, Sbarra DA. Advancing social connection as a public health priority in the United States. Am Psychol 2017;72:517-30.
116. Butler EA, Randall AK. Emotional coregulation in close relationships. Emot Rev 2013;5:202-10.
117. Reeck C, Ames DR, Ochsner KN. The social regulation of emotion: an integrative, cross-disciplinary model. Trends Cogn Sci 2016;20:47-63.
118. Williams WC, Morelli SA, Ong DC et al. Interpersonal emotion regulation: implications for affiliation, perceived support, relationships, and well-being. J Pers Soc Psychol 2018;115:224-54.
119. Coan JA, Schaefer HS, Davidson RJ. Lending a hand: social regulation of the neural response to threat. Psychol Sci 2006;17:1032-9.
120. Soma CS, Baucom BRW, Xiao B et al. Coregulation of therapist and client emotion during psychotherapy. Psychother Res 2020;30:591-603.
121. Wieder G, Wiltshire TJ. Investigating coregulation of emotional arousal during exposure-based CBT using vocal encoding and actor-partner interdependence models. J Consult Clin Psychol 2020;67:337-48.
122. Decety J, Fotopoulou A. Why empathy has a beneficial impact on others in medicine: unifying theories. Front Behav Neurosci 2015;8:457.
123. Kirsch I. How expectancies shape experience. Washington: American Psychological Association, 1999.
124. Nitschke JB, Dixon GE, Sarinopoulos I et al. Altering expectancy dampens neural response to aversive taste in primary taste cortex. Nat Neurosci

2006;9:435-42.
125. Sarinopoulos I, Dixon GE, Short SJ et al. Brain mechanisms of expectation associated with insula and amygdala response to aversive taste: implications for placebo. Brain Behav Immun 2006;20:120-32.
126. Kirsch I. Placebo effect in the treatment of depression and anxiety. Front Psychiatry 2019;10:407.
127. Kirsch I, Deacon BJ, Huedo-Medina TB et al. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. PLoS Med 2008;5:e45.
128. Kirsch I, Moore TJ, Scoboria A et al. The emperor's new drugs: an analysis of antidepressant medication data submitted to the U.S. Food and Drug Administration. Prev Treat 2002;5:23.
129. Stavropoulou C. Non-adherence to medication and doctor-patient relationship: evidence from a European survey. Patient Educ Couns 2011;83:7-13.
130. Zolnierek KBT, DiMatteo MR. Physician communication and patient adherence to treatment: a meta-analysis. Med Care 2009;47:826-34.
131. Eubanks CF, Muran JC, Safran JD. Alliance rupture repair: a meta-analysis. Psychotherapy 2018;55:508-19.
132. Zilcha-Mano S, Muran JC, Eubanks CF et al. Not just a non-specific factor: moderators of the effect of within- and between-clients alliance on outcome in CBT. Cogn Ther Res 2018;42:146-58.
133. Safran JD, Muran JC. Negotiating the therapeutic alliance. New York: Guilford, 2000.
134. Ericsson KA, Krampe RT, Tesch-Römer C. The role of deliberate practice in the acquisition of expert performance. Psychol Rev 1993;100:363-406.
135. Ericsson KA, Lehmann AC. Expert and exceptional performance: evidence of maximal adaptation to task constraints. Annu Rev Psychol 1996;47:273-305.
136. Rousmaniere T. Deliberate practice for psychoherapists: a guide to improving clinical effectiveness. New York: Routledge, 2016.
137. Rousmaniere T, Goodyear R, Miller SD et al (eds). The cycle of excellence: using deliberate practice to improve supervision and training. Hoboken: Wiley, 2017.
138. Chow DL, Miller SD, Seidel JA et al. The role of deliberate practice in the development of highly effective psychotherapists. Psychotherapy 2015;52:337-45.

DOI:10.1002/wps.21035

Accelerating Medicines Partnership® Schizophrenia (AMP® SCZ): developing tools to enable early intervention in the psychosis high risk state

Schizophrenia is a severe mental illness that presents with positive, negative and cognitive symptoms and ranks among the top 15 leading causes of disability worldwide[1]. Signs of risk for developing this illness can occur months to years before diagnosis. This early period, referred to as the clinical high risk (CHR) for psychosis state, reflects a time during which attenuated psychotic symptoms, marked declines in social and role functioning, help-seeking behavior, and non-psychotic comorbidity are noted. Intervention in the CHR state can prevent future illness-related disability[2].

Longitudinal studies of CHR individuals show that, at two-year follow-up, approximately 20% transition to psychosis[3], 41% undergo remission[4], but many of the remainder experience significant symptoms and problems in functioning[4]. Studies are underway to establish risk calculators and biomarkers that can help identify CHR individuals who are most likely to convert to psychosis, but more work is needed to develop tools that use mechanistic input to stratify CHR populations by predicted clinical outcomes beyond psychosis[5]. The CHR stage represents a unique opportunity to develop interventions guided by such tools, focused on reducing conversion to psychosis and improving long-term functional outcomes.

Aimed at capitalizing on this opportunity, the Accelerating Medicines Partnership® Schizophrenia (AMP® SCZ) is a large international collaboration to develop algorithms using a set of clinical and cognitive assessments, multi-modal biomarkers, and clinical endpoints that can be used to predict the trajectories and outcomes of CHR individuals and advance the testing of pharmacological interventions for CHR individuals in need. The goal is to accurately predict which individuals are likely to remit, experience an acute psychotic episode, or have intermediate outcomes that feature persistent attenuated psychotic and/or mood symptoms along with functional impairment. The algorithms will have the potential to serve as early indicators of treatment efficacy in CHR persons.

The AMP SCZ partnership, managed by the Foundation for the National Institutes of Health (FNIH), brings together a breadth of scientific and regulatory expertise and lived experience from the partners: the US National Institute of Mental Health (NIMH), the US Food and Drug Administration (FDA), and the European Medicines Agency (EMA); private industry (Boehringer Ingelheim; Janssen Research & Development; Otsuka Pharmaceutical Development & Commercialization); non-profit and patient advocacy organizations (American Psychiatric Association Foundation; National Alliance on Mental Illness; One Mind; Schizophrenia & Psychosis Action Alliance); and a charitable foundation (Wellcome). The partnership will contribute $117.7 million over 5 years ($99.4 million from NIMH, $7.5 million from industry, and $10.8 million from non-profit organizations) to support implementation of the program.

The AMP SCZ program is composed of two Research Networks – the Psychosis-Risk Outcomes Network (ProNET) at Yale University, and the Trajectories and Predictors in the CHR for Psychosis Population: Prediction Scientific Global Consortium (PRESCIENT) at the University of Melbourne/Orygen – and a Data Processing, Analysis and Coordination Center (DPACC) at Harvard Medical School[6]. ProNET and PRESCIENT form a harmonized research network focused on CHR individuals: identifying biological markers, clinical endpoints, and other measures that predict disease trajectory and outcomes for this group. The DPACC is responsible for managing, processing, disseminating, archiving and analyzing AMP SCZ data, which will be rapidly disseminated and made accessible to all qualified researchers and the public within the NIMH Data Archive[7].

The AMP SCZ research network will recruit a large cohort (N=1,977) of individuals between the ages of 12 and 30 years who meet CHR criteria – based on the Positive SYmptoms for CAARMS Harmonized with SIPS (PSYCHS) interview, a new psychometric instrument for defining CHR and associated outcomes – and healthy controls (N=640) across 42 sites from 14 countries (US, Canada, UK, Spain, Italy, Switzerland, The Netherlands, Germany, Denmark, Australia, Singapore, South Korea, Chile and China). CHR participants will complete screening, baseline assessments, and a battery of follow-up assessments across 24 months. Healthy controls will complete screening and baseline assessments, and a subset (approximately 5 per site) will complete month 2, 12 and 24 visits.

The CHR cohort and healthy controls will be assessed with a core set of measures at baseline and 2 months post-baseline, with additional assessments completed at other timepoints. CHR subjects will be assessed longitudinally for up to 2 years. Subjects who develop their first episode of psychosis ("converted" cases) over the course of study participation will continue to be followed and assessed as scheduled. Measures will include clinical and cognitive assessments; neurophysiology, neuroimaging, genetics and fluid biomarkers; speech and facial expression (audio/video recording); and digital assessments[8].

The digital assessments will collect active (e.g., daily survey on social interactions and feelings of connectedness) and passive (e.g., time spent sleeping, number of texts and phone calls received or made; time participants spend in green space, home, school, exercising, therapy visits, and social relationships) data, along with an automated assessment of social and community functioning from global positioning system (GPS) data. Through the digital measures, AMP SCZ will be able to assess bio-psycho-social data in CHR individuals and elucidate their role in affecting trajectories which could be targeted by psychosocial interventions.

The primary endpoint of interest is conversion to psychosis by

24-month follow-up as defined by psychosis threshold criteria on the PSYCHS. Secondary clinical endpoints of interest include remission or recovery of CHR state, and non-conversion/non-remission. Clinical outcomes of interest cover multiple domains such as attenuated positive symptoms, mood and anxiety, psychosocial functioning, and persistent negative symptoms[8].

The biomarker data collected by ProNET and PRESCIENT will be analyzed by the DPACC to develop multi-modal prediction models and risk calculators by drawing on recent theoretical and methodological advances (e.g., dynamic prediction, probabilistic multimodal modeling). These models will leverage existing prediction models in the field[9] and guide selection and stratification of CHR participants for future clinical trials based on the primary endpoint of interest. For example, stratification can identify a subset of CHR participants who are at higher risk of developing psychosis relative to the rest.

The developed tools may have clinical utility in decision making about stepping interventions up or down as risk is assessed over time (clinical trajectory, treatment response) and in response to incoming biomarker information. Some tools, such as the risk calculators, will prioritize the less invasive and more readily available biomarkers for prediction, to enable clinical tools that could be used in community-based settings and are more tolerable by subjects. The novel prediction models generated for the AMP SCZ dataset will be tested using cross-validation approaches designed to improve generalizability of the derived algorithms to other CHR cohorts.

By integrating the strengths of multiple international stakeholders, sharing discoveries openly, and priming future research, the AMP SCZ program aims to catalyze advances in knowledge about the CHR population to enable intervention at the earliest stages of schizophrenia, with the goal of maximizing functional outcomes for CHR patients.

Linda S. Brady[1], Carlos A. Larrauri[2], AMP SCZ Steering Committee
[1]National Institute of Mental Health, Bethesda, MD, USA; [2]National Alliance on Mental Illness, Arlington, VA, USA

The AMP SCZ Steering Committee includes E. Appelmans, R. Benabou, L. Bilsland, T. Brister, F. Butlen-Ducuing, M.C. Davis, K. Duckworth, G.K. Farber, B.A. Fischer, S. Frangou, S.T. Garcia, N. Gogtay, S. Gopal, R.K. Heinssen, W. Horton, B.R. Johnson, P.S. Joshi, N.I. Keren, S.H. Lisanby, G. Pandina, S.E. Roth, M. Sand, A.J. Savitz, B. Staglin, M. Tomé, E. Velthorst, D. Wholley, and J.A. Gordon. A complete listing of participating sites and study investigators can be found at ampscz.org. AMP SCZ data are held in the NIMH Data Archive and available at nda.nih.gov/ampscz. Those expressed in this paper are personal views of the authors and may not be understood or quoted as being made on behalf of or reflecting the position of the European Medicines Agency or one of its committees or working parties. The authors acknowledge the seminal contributions of W.Z. Potter to earlier versions of the AMP SCZ research plan. They also acknowledge S. Morris and L. Rowland for their programmatic leadership of the AMP SCZ Research Networks; A. Wijtenburg for her efforts in support of the harmonized PSYCHS instrument; and J. Pevsner, M. Zhan, R. Beer and S. Vaziri for their programmatic leadership of the AMP SCZ Data Processing and Analysis Coordinating Center.

1. GBD 2019 Diseases and Injuries Collaborators. Lancet 2020;396:1204-22.
2. Fusar-Poli P, Salazar de Pablo G, Correll CU et al. JAMA Psychiatry 2020;77:755-65.
3. Salazar de Pablo G, Radua J, Pereira J et al. JAMA Psychiatry 2021;78:970-8.
4. Salazar de Pablo G, Besana F, Arienti V et al. EClinicalMedicine 2021;36:100909.
5. Polari A, Yuen HP, Amminger P et al. Early Interv Psychiatry 2021;5:642-51.
6. Accelerating Medicines Partnership® Schizophrenia. https://www.ampscz.org/about/networks-coordination.
7. NIMH Data Archive. https://www.ampscz.org/scientists/data.
8. Accelerating Medicines Partnership® Schizophrenia. https://www.ampscz.org/scientists/design.
9. Worthington MA, Cannon TD. Front Psychiatry 2021;12:770774.

DOI:10.1002/wps.21038

A critical assessment of NICE guidelines for treatment of depression

The UK National Institute for Health and Care Excellence (NICE) recently updated its recommendations for the treatment of depression[1]. This effort has many strengths, including the meticulous documentation of the process; systematic reviews, meta-analyses and cost-effectiveness analyses; and inclusion of stakeholder comments that feed into the guidelines. Here we attempt a constructive critical appraisal of areas where future improvements for this but also for other similar initiatives are feasible, with a special focus on psychotherapies for depression.

We first notice that the methods and analyses of the NICE guidelines were not subjected to formal external peer review for any of the addressed questions. Asking stakeholders for comments is welcome, but it is unlikely to be equally rigorous, leaving it to the guideline committee how these comments are considered. External peer review is recommended as a default quality standard for treatment guidelines[2].

Furthermore, study protocols were pre-registered only for some conditions (e.g., for new episodes of depression and treatment-resistant depression), but not for others (including chronic depression, depression with personality disorder, and psychotic depression). Pre-registering should be established as a default standard in guidelines for all reviewed conditions.

For the primary analysis concerning new episodes of depression, network meta-analysis (NMA) was chosen[1]. NMA has the advantage of incorporating both direct and indirect evidence, but complex assumptions need to be fulfilled, and the level of evidence provided is still debated[3]. For these reasons, NMA results and the derived inferences require extra caution.

For treatment ranking, the guideline committee primarily focused on effect sizes from NMA treatment comparisons with placebo or treatment-as-usual, and compared these effect sizes between treatments. From these comparisons, the committee concluded that some treatments appeared to be "more effective" than others[1]. For most treatments, however, the differences between treatment and control effect sizes were below the minimal clinically significant difference defined by the committee (standardized mean difference, SMD >0.5 or <–0.5)[1]. This applies to comparisons between individual cognitive or cognitive-behavioral therapy (CT/CBT), individual interpersonal therapy (IPT), individual problem solving, individual short-term psychodynamic psychotherapy (STPP), and

group behavior activation. Thus, with only subtle effect size differences, treatment ranking carries large uncertainty. Furthermore, assuming differences between two treatments if one of them shows descriptively a larger effect size than the other compared to a control condition, without comparing them directly, should be avoided[4].

The guideline committee reported head-to-head comparisons of active treatments only in a supplement. These comparisons show that, in more severe depression, the differences between individual behavioral therapy, individual CBT, individual IPT and individual STPP are neither statistically nor clinically significant (SMDs <0.50)[1]. In less severe depression, only a few clinically significant differences were found: for example, in a pairwise comparison, STPP was statistically and clinically significantly superior to counselling (SMD=–0.61, 95% CI: –1.05 to –0.17), but was ranked below counselling.

Thus, the committee's conclusions about differences in efficacy between active treatments are not consistent with its own head-to-head comparisons. They are also not compatible with independent peer-reviewed evidence of no substantial differences in efficacy between psychotherapies[5]. The committee, however, erroneously interpreted this independent evidence[5] as confirming its treatment ranking[1,B, p.165]. In summary, procedures for treatment ranking need to be pre-defined, and subtle differences below the threshold of clinically meaningful values should not be overstated.

In principle, possible allegiance and conflicts of interests need to be controlled for[2], for example by including methodologists, patients, and different-field experts, and by limiting the involvement of field specialists to a consultation role[6]. Avoidance of stacking is also essential, ensuring that guideline developers do not have an over-representation of believers in one or another treatment modality[6].

The guideline committee based the hierarchy of treatment recommendations on both efficacy and cost-effectiveness, which is useful in trying to optimize the use of treatments for conditions with high prevalence[1]. For cost-effectiveness, however, peer reviews and pre-registration are missing. Moreover, the cost-effectiveness literature is notoriously replete with biases. This further complicates matters in a field such as depression where the primary studies are often also biased (e.g., sponsor bias in pharmacotherapy trials and allegiance bias in psychotherapy trials). Furthermore, the studies used by the committee for cost-effectiveness analysis did not cover all relevant treatment types. For those not covered, it is not clear whether cost-effectiveness estimates are valid. Additional cost-effectiveness analyses commissioned by the committee were based on the NMA treatment-control effect sizes shown above to be questionable, which further limits the derived treatment ranking.

Another challenge is whether extrapolations from new episodes of depression to other conditions are valid, when there is no solid evidence for these other categories of depression. For example, in depression with personality disorder, the committee recommends combining antidepressants and psychotherapy. For the choice between psychotherapies, readers are referred to the treatments for new episodes of depression. Then, for patients not sufficiently responding to pharmacotherapy alone, switching to psychotherapies listed for new episodes of more severe depression is recommended as one option. In reviewing new episodes of depression, however, the committee excluded depression with personality disorder and treatment-resistant depression. Thus, the committee's ranking of psychotherapies for new episodes of depression may not be valid for these other conditions. Finally, for the cost-effectiveness of chronic depression and depression with personality disorder, the committee also used the economic data for new episodes of depression.

As another problem, the guideline committee found the quality of studies to be quite low. The committee tried to adjust results for bias, but a pre-registered threshold analysis for assessing confidence in recommendations was not carried out. Quality of evidence was evaluated narratively using the GRADE system, but without assessing confidence. Assessing confidence in evidence is essential for guidelines[6].

The committee also draws an arbitrary distinction between the more complex forms of depression, which not only reduces generalizability to clinical practice but appears to have led to the exclusion of relevant studies. Available randomized controlled trials have not clearly distinguished between chronic depression and treatment-resistant depression. For chronic depression, the committee recommends CBT, antidepressants or their combination[1]. However, these recommendations do not take into account the evidence for STPP and long-term psychodynamic therapy in treatment-resistant depression and in depression with personality disorder[7,8], conditions highly associated with chronic depression. Guidelines need to avoid arbitrary distinctions of disorders.

Moreover, the committee did not sufficiently consider the limitations of the available evidence[2], especially the limited remission rates (about 30%) of short-term psychotherapies (4-20 sessions), with SMDs of 0.30[9]. Aggravating this problem, most effect sizes of short-term treatments are not stable at follow-up[1]. Especially for chronic depression, success rates may be improved with longer-term treatments[9]. The committee, however, considered long-term treatments only as an option for depression with personality disorder.

Finally, an explicit link between evidence and recommendations is missing[2]. We acknowledge that the evidence in this field is uncertain, and this may be the reason why the committee found it "difficult… to link the recommendations directly to the NMA results"[1,B, pp.48,66], and based its recommendations ultimately on "clinical experience"[1,B, p.66]. However, it is unclear whether clinical experience can offer any solid guidance when treatment differences are modest, uncertainty is high and bias is substantial. Guidelines should fully admit this uncertainty and avoid over-simplified, over-confident recommendations[6].

Falk Leichsenring[1,2], Christiane Steinert[3], Felicitas Rost[4], Allan Abbass[5], Nikolas Heim[3], John P.A. Ioannidis[6-10]

[1]Department of Psychosomatics and Psychotherapy, University of Giessen, Giessen, Germany; [2]Department of Psychosomatics and Psychotherapy, University of Rostock, Rostock, Germany; [3]International Psychoanalytic University, Berlin, Germany; [4]Tavistock and Portman NHS Foundation Trust, London, UK; [5]Dalhousie University, Halifax, NS, Canada; [6]Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA; [7]Department of Epidemiology and Population Health, Stanford University School

of Medicine, Stanford, CA, USA; [8]Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, CA, USA; [9]Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, CA, USA; [10]Meta-Research Innovation Center at Stanford, Stanford, CA, USA

1. National Institute for Health and Care Excellence. Depression in adults: treatment and management. www.nice.org.uk/Guidance/ng222.
2. Brouwers MC, Kho ME, Browman GP et al. CMAJ 2010;182:E839-42.
3. Faltinsen EG, Storebo OJ, Jakobsen JC et al. BMJ Evid Based Med 2018;23: 56-9.
4. Makin TR, Orban de Xivry JJ. Elife 2019;8:e48175.
5. Cuijpers P, Quero S, Noma H et al. World Psychiatry 2021;20:283-93.
6. Lenzer J, Hoffman JR, Furberg CD et al. BMJ 2013;347:f5535.
7. Fonagy P, Rost F, Carlyle JA et al. World Psychiatry 2015;14:312-21.
8. Abbass A, Town J, Driessen E. Psychiatry 2011;74:58-71.
9. Leichsenring F, Steinert C, Rabung S et al. World Psychiatry 2022;21:133-45.

DOI:10.1002/wps.21039

Cyberbullying: next-generation research

Cyberbullying, or the repetitive aggression carried out over electronic platforms with an intent to harm, is probably as old as the Internet itself. Research interest in this behavior, variably named, is also relatively old, with the first publication on "cyberstalking" appearing in the PubMed database in 1999.

Over two decades later, the broad contours of the problem are generally well understood, including its phenomenology, epidemiology, mental health dimensions, link to suicidality, and disproportionate effects on minorities and individuals with developmental disorders[1]. Much remains understudied, however. Here we call for a "next generation" of research addressing some important knowledge gaps, including those concerning self-cyberbullying, the bully-victim phenomenon, the bystander role, the closing age-based digital divide, cyberbullying subtypes and how they evolve with technology, the cultural specificities of cyberbullying, and especially the management of this behavior.

Defined as the anonymous online posting, sending or otherwise sharing of hurtful content about oneself, "self-cyberbullying" or "digital self-harm" has emerged as a new and troubling manifestation of cyberbullying. Rather than a fringe phenomenon, self-cyberbullying is thought to affect up to 6% of middle- and high-school students[2]. Is this a cry for help by someone who might attempt "real" self-harm or even suicide if not urgently treated? Is it "attention-seeking" in nature, meant to drive Internet traffic in a very congested social media landscape where it can be hard to get noticed and where "likes" are the currency of self-worth? Research is needed to better characterize self-cyberbullying, including how it relates to depression and offline self-harm and suicide.

The bully-victim phenomenon refers to the permeable boundaries between roles that can make it relatively easy for a cyberbullying victim to become a cyberbully and vice versa. Unlike traditional bullying, visible markers of strength are not a requirement in cyberbullying. Assuming the identity of the cyberbully is known, all that the victims need to attack back and become cyberbullies themselves is a digital platform and basic digital know-how. Do cyberbullying victims feel in any way "empowered" by this permeability, as some do express in clinical settings? And does knowledge that perpetrators can be attacked back have any deterrent effect on them, or is the bi-directional violence that can ensue an unmitigated race to the bottom that further impairs well-being?

What of the bystander role? Depending on the platform, the audience witnessing a cyberbullying attack can potentially be limitless – attacks that go viral are an extreme example of this. While this can magnify the humiliation inflicted on the victim, it also introduces the possibility of enlisting bystanders to protect victims and push back against perpetrators. Research examining how to leverage bystanders as part of anti-cyberbullying interventions would have significant management and public health utility.

Recent scholarship has brought attention to cyberbullying beyond the young age group. What had been called the "digital divide", which in this context refers to the notion that children and adolescents are more active online and therefore at higher risk, has narrowed to the point where a significant risk of cyberbullying now appears to exist among college students and perhaps adults overall. Cyberbullying is no longer a middle- and high-school problem, as suggested by a 30-country United Nations-sponsored survey that recruited nearly 170,000 youth up to 24 years of age and found that 33% of them had been victims of that behavior[3]. To better protect against cyberbullying and implement age-appropriate interventions, new research should better delineate the upper limits of the high-risk cyberbullying age bracket, if they exist.

There is also insufficient research into the culturally-specific dimensions of cyberbullying. Co-authoring analyses reveal that the most influential cyberbullying scholarship comes from the US, and that the top 5 universities in publication productivity are in the European Union[4]. Given the different relationship to violence across cultures and the diverging definitions of, and reactions to, trauma worldwide, a broader culturally-centered research perspective is essential for a more thorough understanding of cyberbullying's global impact.

As we "zoom out" and investigate across cultures, we should also "zoom in" on the specific cyberbullying behavior. Are all cyberbullying attacks similar in terms of prevalence, perpetrator and victim profiles, short- and long-term consequences, and management strategies? Several forms of cyberbullying have been identified[5], but their similarities and differences require elucidation, especially as technology continues to change and new forms emerge. Therefore, future research should compare diverse behaviors, such as cyberstalking, "excluding" (deliberately leaving someone out), "doxing" (revealing sensitive information about the victim), "fraping" (using the victim's social media account to post inappropriate content under the victim's name), "masquerading" (creating a fake identity with which to attack the victim), "flaming" (posting insults against the victim), and sex-based cyberbullying through the non-consensual sending of sexual text messages or imagery.

To better understand and address cyberbullying, we must explore its existing subtypes – some of which have only been described in blogs – and, as technology evolves, its emerging forms.

Most urgently, the lack of agreement upon "best practices" for the management of cyberbullying must be remedied. Expanding access to psychiatric and psychological care – given the mental health dimension of cyberbullying – is imperative, as is a better understanding of school-based interventions, which remain the most popular management approach.

Data from school-based studies suggest that programs which adopt a broad, ecological approach to the school-wide climate and which include specific actions at the student, teacher and family levels are more effective than those delivered solely through classroom curricula or social skills trainings[6]. However, the best meta-analytic evidence for school-based programs demonstrates mostly short-term effects[7], while long-term data suggest small benefits[8]. Further, success appears more likely when programs target cyberbullying specifically as opposed to general violence prevention[7], and when they are delivered by technology-savvy content experts as opposed to teachers[8]. Evidence also suggests that programs are most successful when they provide informational support through interactive modalities (e.g., peer tutoring, role playing, group discussion), and when they nurture stakeholder agency (e.g., offer quality teacher training programs, engage parents in program implementation)[9].

Future research into cyberbullying management should expand on these findings and examine how management interfaces with the legislative process and with law enforcement when it comes to illegal behavior, including privacy breeches and serious threats.

Much has been learned about cyberbullying, but much remains to be explored. The knowledge gaps are all the more challenging given that Internet-related technologies evolve at a breakneck pace and in a way that reveals new exploitable vulnerabilities. Along with the previously cited statistic that no less than 33% of young people worldwide have been victimized[3], this should give the field added urgency to "keep up" and investigate some understudied areas that are critical to a more nuanced understanding of cyberbullying and its effective management.

Elias Aboujaoude[1], Matthew W. Savage[2]
[1]Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA, USA; [2]School of Communication, San Diego State University, San Diego, CA, USA

1. Aboujaoude E, Savage MW, Starcevic VD et al. J Adolesc Health 2015;57:10-8.
2. Patchin JW, Hinduja S. J Adolesc Health 2017;61:761-6.
3. United Nations International Children's Emergency Fund (UNICEF). More than a third of young people in 30 countries report being a victim of online bullying. https://www.unicef.org.
4. Peker A, Yalçın RU. Front Commun 2022;7:768494.
5. Slonje R, Smith PK, Frisén A. Comput Hum Behav 2013;29:26-32.
6. Cantone E, Piras AP, Vellante M et al. Clin Pract Epidemiol Ment Health 2015;11(Suppl. 1 M4):58-76.
7. Polanin JR, Espelage DL, Grotpeter JK et al. Prev Sci 2022;23:439-54.
8. Ng ED, Chua J, Shorey S. Trauma, Violence & Abuse 2022;23:132-51.
9. Lan M, Law N, Pan Q. Comput Hum Behav 2022;130(Suppl. 1 M4):107200.

DOI:10.1002/wps.21040

The role of gamification in digital mental health

In the face of high unmet mental health needs and overburdened mental health systems, scalable approaches to increase use of evidence-based interventions are essential. Smartphone apps, e-therapies and other digital interventions offer promise in this regard.

Digital interventions can be effective for a range of clinical disorders. These tools, particularly those that can be used without clinical support, can have enormous reach[1]. However, early optimism that they could be placed online, optimally utilized by those who need them, and thereby improve population mental health, has not been realized. Both the uptake of tools and sustained engagement with them have often been disappointing[1,2].

More sophisticated efforts, in both *systems around* digital interventions and *features within* the digital tools themselves are required. Promising areas in systems around the tools include improved public messaging, clinician training, and embedding tools within clinical, educational or workplace settings[1]. In terms of improving digital interventions themselves, there is potential in further increasing appeal (so that people are willing to try the tools), improving usability (thus addressing the major reason for early disengagement in apps) and enhancing "stickiness". By "stickiness", we mean the degree to which users' adherence or engagement is supported by aspects of the intervention itself, rather than relying on their personal effort or external support. A key opportunity for both appeal and stickiness is increased use of gamification.

Gamification refers to the use of features from gaming in contexts that are not games as such[3,4]. Commonly used features include small achievable challenges (often building toward larger objectives), rapid feedback or rewards, and personalization. Other features include unpredictability, increasing complexity, narrative, themes or imaginary settings, opportunities to choose and explore, and social interaction or competition[3,4].

Gamification can allow users to test and rehearse skills in a safe yet responsive environment, offer extrinsic motivation, and support intrinsic motivation (e.g., by noticing progress)[3]. It often includes elements of user control, supporting a sense of autonomy[5], and may facilitate a sense of flow or immersion, important for enjoyment and sustained attention[3]. From step counters to supermarket loyalty schemes, gamification has burgeoned with the development of digital technologies.

Within the field of digital psychiatry, gamification offers three key areas of potential[3]. First, an appeal or attractiveness potential. Games are among the most popular forms of entertainment

globally, reaching a hugely diverse audience. Far from the popular stereotype of gaming as a teenage male phenomenon, the average gamer is over 30 years old and 45% are female. A gamified intervention may be more appealing to some users than traditional models due to fun elements. Gamification might also reduce barriers to therapy such as stigma and help-negation[4]. Second, gamification may offer potential for alternative mechanisms of change to those emphasized in more traditional approaches. For example, facilitating the visualization of complex ideas, such as negative thoughts, and allowing manipulation of such images. Third, gamification offers an engagement potential, keeping users engaged in the tool longer than they otherwise might be, via the use of rewards, fun and other features, meaning that users get a higher "dose" of the intervention[3].

While gamification has been used in diverse areas, there is little evidence to date in psychiatry. A meta-analysis did not identify higher adherence or impact for gamified compared to non-gamified apps for depression[6], and there is a lack of recent evaluative reviews[4]. Reviews are hampered by heterogeneity and lack of specificity about gamification processes and by time delays between rapidly changing digital approaches and publication of trials. However, studies have reported that gamified mental health options are appealing for some users. Young adults with internalizing symptoms selected a game promoted as a mental health intervention over an entertainment game[7] and, in a community sample, many adolescents considered gamified interventions appealing[8]. That said, the latter study reported polarized views: some adolescents advised that gamification might be trivializing of their distress and highlighted the need for choice in digital approaches[8].

In the face of interest but limited evaluative literature, it is useful to consider illustrative examples. Gamification techniques have been widely used in mental health tools. Here we outline two contrasting examples: Headspace, one of the most popular mental well-being apps, and SPARX, a cognitive behaviour therapy (CBT)-based treatment for adolescent depression.

Headspace is a meditation app boasting tens of millions of downloads. While it does not look like a game, it uses multiple gamification features[5]. Content comprises short chunks that build into larger achievements; targets and progress are shown clearly; and "badges" for activities are immediate. Other features common to gamification include a colourful aesthetic, optional notifications, minimal text, animations and social influence. As an often underrecognized but important feature of gamification, Headspace provides extensive yet simple choices and opportunities for user control[5]. While there are few trials of Headspace for psychiatric disorders, it is one of the most downloaded mental well-being apps in the world[2] and has among the highest retention rates of these[8], demonstrating both phenomenal appeal and good "stickiness". There are no direct comparisons to consider how much these are due to gamification, and Headspace also utilizes other features such as a large promotions budget. However, gamification features are integral in this app.

SPARX is an unguided computerized CBT program offered in a game-like format. It makes extensive use of metaphor and story to allow users to discover and rehearse therapeutic content in a playful manner, and then reflect on skills and their use in real life with an animated virtual therapist. Gamification features include narrative, imaginary settings, opportunities to explore, puzzles, reward "mini-games", and playful quizzes. SPARX was not inferior to treatment-as-usual for depressive symptoms in a large trial[9]. Retention rates were good in studies, and adolescents reported that game features were helpful for engagement[9]. However, once implemented outside of research settings in New Zealand, retention has been lower, and adolescents have commented on the need for updates in line with expectations of commercial games[9]. Interestingly, while New Zealand adolescents advised that SPARX is suitable for younger teens, a Japanese version of SPARX has been most widely used by adult men[9].

These examples illustrate that, far from being only for the young, or for non-clinical use, gamified interventions can engage adults and offer evidence-based treatment. As well as these examples, there are many other instances of gamification in digital mental health[3,4,7]. However, the literature is at an early stage. It would be premature to claim major impact or failure for gamification in psychiatry. There are also specific challenges, including high expectations of gaming in accordance with the high budgets involved in many computer games, and, on the other hand, expectations of non-playful interventions for serious needs. While we have mentioned that gamification might support motivation, external rewards can undermine internal motivation if not used carefully[4]. Future research should explore these questions and examine the impact of specific gamification features, make stronger use of gamification theory, and consider audience segmentation and the importance of user preferences[3,4,9].

It is critical to expand scalable approaches to improving mental health. Digital tools offer extraordinary potential for this. However, the appeal and stickiness of digital tools must be addressed. Gamification offers promise for increasing appeal and engagement and should be pursed alongside other opportunities.

Theresa Fleming[1], Marlou Poppelaars[2], Hiran Thabrew[3]
[1]School of Health, Te Herenga Waka, Victoria University of Wellington, Wellington, New Zealand; [2]Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands; [3]Department of Psychological Medicine, University of Auckland, Auckland, New Zealand

1. Torous J, Bucci S, Bell I et al. World Psychiatry 2021;20:318-35.
2. Baumel A, Muench F, Edan S et al. J Med Internet Res 2019;21:e14567.
3. Fleming T, Sutcliffe K, Lucassen M et al. Serious games and gamification in clinical psychology. Reference module in neuroscience and biobehavioral psychology. Amsterdam: Elsevier, 2020.
4. Cheng VWS. Front Psychol 2020;11:3434.
5. Fish MT, Saul AD. Simul Gaming 2019;50:419-35.
6. Six SG, Byrne KA, Tibbett TP et al. JMIR Ment Health 2021;8:e32199.
7. Poppelaars M, Wols A, Lichtwarck-Aschoff A et al. Front Psychol 2018;9:1837.
8. Fleming T, Merry S, Stasiak K et al. JMIR Ment Health 2019;6:e12656.
9. Fleming T, Lucassen M, Stasiak K et al. Adolesc Ment Health 2021;26:92-4.

DOI:10.1002/wps.21041

The future of psychopharmacology: a critical appraisal of ongoing phase 2/3 trials, and of some current trends aiming to de-risk trial programmes of novel agents

Christoph U. Correll[1-4], Marco Solmi[1,5-9], Samuele Cortese[9-13], Maurizio Fava[14], Mikkel Højlund[15,16], Helena C. Kraemer[17], Roger S. McIntyre[18-23], Daniel S. Pine[24], Lon S. Schneider[25], John M. Kane[2-4]

[1]Department of Child and Adolescent Psychiatry, Charité Universitätsmedizin Berlin, Berlin, Germany; [2]Department of Psychiatry, Zucker Hillside Hospital, Northwell Health, Glen Oaks, NY, USA; [3]Department of Psychiatry and Molecular Medicine, Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY, USA; [4]Center for Psychiatric Neuroscience, Feinstein Institute for Medical Research, Manhasset, NY, USA; [5]Department of Psychiatry, University of Ottawa, Ottawa, ON, Canada; [6]Department of Mental Health, Ottawa Hospital, Ottawa, ON, Canada; [7]Ottawa Hospital Research Institute (OHRI) Clinical Epidemiology Program, University of Ottawa, Ottawa, ON, Canada; [8]School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada; [9]Centre for Innovation in Mental Health, School of Psychology, Faculty of Environmental and Life Sciences, University of Southampton, Southampton, UK; [10]Clinical and Experimental Sciences (CNS and Psychiatry), Faculty of Medicine, University of Southampton, Southampton, UK; [11]Solent NHS Trust, Southampton, UK; [12]Division of Psychiatry and Applied Psychology, School of Medicine, University of Nottingham, Nottingham, UK; [13]Hassenfeld Children's Hospital at NYU Langone, New York University Child Study Center, New York, NY, USA; [14]Depression Clinical and Research Program, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; [15]Department of Public Health, Clinical Pharmacology, Pharmacy and Environmental Medicine, University of Southern Denmark, Odense, Denmark; [16]Mental Health Services in the Region of Southern Denmark, Department of Psychiatry Aabenraa, Aabenraa, Denmark; [17]Department of Psychiatry and Behavioral Sciences, Stanford University, Cupertino, CA, USA; [18]Mood Disorders Psychopharmacology Unit, University Health Network, Toronto, ON, Canada; [19]Institute of Medical Science, University of Toronto, Toronto, ON, Canada; [20]Canadian Rapid Treatment Center of Excellence, Mississauga, ON, Canada; [21]Department of Psychiatry, University of Toronto, Toronto, ON, Canada; [22]Department of Pharmacology, University of Toronto, Toronto, ON, Canada; [23]Brain and Cognition Discovery Foundation, Toronto, ON, Canada; [24]Section on Developmental Affective Neuroscience, National Institute of Mental Health, Bethesda, MD, USA; [25]Department of Psychiatry and Behavioral Sciences, and Department of Neurology, Keck School of Medicine, and L. Davis School of Gerontology, University of Southern California, Los Angeles, CA, USA

Despite considerable progress in pharmacotherapy over the past seven decades, many mental disorders remain insufficiently treated. This situation is in part due to the limited knowledge of the pathophysiology of these disorders and the lack of biological markers to stratify and individualize patient selection, but also to a still restricted number of mechanisms of action being targeted in monotherapy or combination/augmentation treatment, as well as to a variety of challenges threatening the successful development and testing of new drugs. In this paper, we first provide an overview of the most promising drugs with innovative mechanisms of action that are undergoing phase 2 or 3 testing for schizophrenia, bipolar disorder, major depressive disorder, anxiety and trauma-related disorders, substance use disorders, and dementia. Promising repurposing of established medications for new psychiatric indications, as well as variations in the modulation of dopamine, noradrenaline and serotonin receptor functioning, are also considered. We then critically discuss the clinical trial parameters that need to be considered in depth when developing and testing new pharmacological agents for the treatment of mental disorders. Hurdles and perils threatening success of new drug development and testing include inadequacy and imprecision of inclusion/exclusion criteria and ratings, sub-optimally suited clinical trial participants, multiple factors contributing to a large/increasing placebo effect, and problems with statistical analyses. This information should be considered in order to de-risk trial programmes of novel agents or known agents for novel psychiatric indications, increasing their chances of success.

Key words: Psychopharmacology, clinical trials, design, methodology, novel mechanisms of action, schizophrenia, bipolar disorder, major depressive disorder, anxiety disorders, trauma-related disorders, substance use disorders, dementia

(World Psychiatry 2023;22:48–74)

The timely as well as effective and safe treatment of mental disorders is a key focus in medicine, due to the early onset of these disorders, and their severity, chronicity and major effects on multiple biopsychosocial aspects of human life[1-4]. Clinicians, patients, family members and the society at large have substantial interest in the availability of new treatment options that have greater, broader or more specific efficacy and similar or enhanced tolerability compared to already available agents, ideally also involving new mechanisms of action that may help personalization of treatment[5-7].

Pharmacological approaches to mental disorders were initially mostly the outcome of observation and serendipitous discoveries, also informed by substances that could alter mental states and lead to addiction. In the 1950s and 1960s, there was a steep increase in the availability of pharmacological agents that were helpful in improving mental health by reducing symptoms of multiple psychiatric disorders. Most of the finer understanding of brain mechanisms involved in mental illness generation was derived from inductive reasoning, i.e., the effect of a medication on the brain was observed, the mechanism of action of the drug was studied in animal and human models, and the insights were used as the basis for hypothesizing biological underpinnings of mental disorders.

In that sense, psychopharmacology is essentially a symptom-based discipline. This approach is further related to the fact that our systems for classifying mental illness consist of patterns of often co-occurring and/or connected symptoms, which are elevated to the status of disorders as long as they lead to distress or dysfunction and are not due to the effects of a substance or a medical condition. This classification is not related to an underlying biology of the identified disorders. Comorbidities are very common and medications often do not work in a substantial number of people with a given diagnosis and/or have pleiotropic and non-specific effects, working for more than one disorder. Recognizing these shortcomings of current nosological systems, alternative approaches are being

proposed[8-10], but are not adopted in the clinical and regulatory classification and drug approval process.

Moreover, the pharmacological nomenclature has remained arcane, being only rarely or incompletely related to the mechanisms of action of medications, as is common in medicine to characterize drug classes. Instead, medications are usually named after their first indication. This has given rise to a terminology that can confuse patients, family members, clinicians and even regulators[11]. For example, the so-called antipsychotics are approved for such diverse indications as schizophrenia, bipolar mania, bipolar depression, major depressive disorder, tic disorder, and irritability associated with autism[12,13]; and have been also found effective for anxiety, insomnia, agitation/aggression, obsessive-compulsive disorder (OCD), and post-traumatic stress disorder (PTSD)[14]. Similarly, the so-called antidepressants have been approved for major depressive disorder, various types of anxiety disorders, and OCD; and are used clinically also for bipolar depression and insomnia, among other conditions[12,13,15,16].

This diagnostically non-specific, pleiotropic use of medication classes is certainly in part due to the complexity and overlap of the biological mechanisms underlying behavioral, emotional and cognitive manifestations. At the same time, medications often do not impact a single biological system, but have a variety of biological effects, that would need to be dissected further and may be dose-dependent. For example, quetiapine, one of the most prescribed so-called antipsychotics, is more frequently administered in combination with other drugs than in monotherapy for psychosis, and is more often used for mood, anxiety and sleep disorders than for psychotic symptoms. The use of quetiapine for such diverging diagnoses and symptoms is linked to the fact that the main pharmacodynamic effect of this medication varies according to the dose at which it is administered[17]. For example, at low doses (25-50 mg/day), it acts as an antihistaminic, which can help treat anxiety, insomnia and agitation/tension. At medium doses (150-300 mg/day), it turns out to have alpha-2 adrenergic receptor blocking and noradrenaline-reuptake inhibiting activity, making it useful as a treatment for major depressive disorder and bipolar depression. At higher doses (450-600 mg/day and above), its postsynaptic dopamine antagonism becomes relevant, making it useful for the treatment of psychosis and mania.

This disorder-driven approach to psychopharmacology is shared by regulatory bodies. Thus, for example, a medication initially marketed for a given disorder may automatically get a black box warning when it becomes indicated for another disorder, even though the safety risk data motivating that warning apply to a pharmacologically entirely different drug class, and no such risk has been described for that medication. This carry-over effect has occurred, for instance, for all dopamine receptor blockers and partial dopamine agonists with respect to the risk of suicide, when they received regulatory approval by the US Food and Drug Administration (FDA) for major depressive disorder, although the relevant (possibly medication-related) data in adolescents and young adults[18,19] were restricted to traditional "antidepressants" that are monoamine reuptake inhibitors or modulators.

The neuroscience-based nomenclature initiative has been to some extent helpful in trying to refine our pharmacological terminology, bringing to bear the knowledge that we have so far in order to classify medication classes and members of each class[20-23].

At the core of state-of-the-art testing of the risks and benefits of a new molecular entity in psychopharmacology are randomized controlled parallel-group clinical trials. However, multiple hurdles in trial design and conduct may interfere with the development of molecular entities showing promise in phase 1 and 2 trials, when they are tested in increasingly large phase 3 trial programmes. Relatively recent failures concerning medications for schizophrenia have included pomaglumetad for total symptoms[24,25], encenicline for cognitive symptoms[26,27], and bitopertin for negative symptoms[28-30]. Similarly, multiple drug development failures on the translational trajectory from phase 1 and 2 into phase 3 trials have involved drugs targeting dementia[31].

Reasons for these failures may be related to the true inefficacy of a drug, its toxicity profile, insufficiently understood dose-response relationships, unknown patient factors, but also the limited knowledge of the biological mechanisms underpinning mental disorders, which prevents the identification of potentially relevant subgroups. An additional factor involved is the increasing placebo response across multiple mental disorders, whose reasons remain insufficiently understood[32-40].

After many decades with few, if any, discoveries of novel effective targets beyond enhancing serotonin and noradrenaline or blocking postsynaptic dopamine transmission for the treatment of mental disorders, some advances have recently occurred. Medications with more recent regulatory approval have targeted the melatonin[41], orexin[42], GABA-A[43,44], opioid[45,46] and N-methyl-D-aspartate (NMDA)[47,48] receptor systems, the vesicular monoamine transporter-2 (VMAT-2) for tardive dyskinesia[49], and inverse agonism of 5-HT2A receptors[50]. Furthermore, there is currently a renaissance of exploiting mechanisms of action of psychedelics, attempting to isolate their beneficial effects without their short- or longer-term risk of brain harm or addictive potential[51-55]. Nonetheless, there is great concern that many, if not most, of the currently studied drugs with new mechanisms of action may not pass through the "valley of death" of their phase 2 and, especially, phase 3 development.

In this paper, we first provide an overview – based on a systematic search in clinicaltrials.gov and clinicaltrialsregister.eu (EudraCT) – of medications with innovative mechanisms of action that are undergoing phase 2 or 3 testing for the treatment of a main mental disorder in adults, such as schizophrenia, bipolar disorder, major depressive disorder, anxiety and trauma-related disorders, substance use disorders, and dementia, highlighting those agents that are seen as having the most promise (as emerging from documented superiority over placebo, magnitude of the observed effect, and demonstration of requirements for safety and tolerability). We then critically discuss the ongoing developments in clinical trial methodology, design and conduct that need to be considered in depth when developing and testing pharmacological agents for the treatment of men-

tal disorders, in order to de-risk trial programmes of novel agents or known agents for novel psychiatric indications.

OVERVIEW OF MEDICATIONS UNDERGOING PHASE 2 AND 3 CLINICAL TRIALS

Schizophrenia

Agents in development for the treatment of schizophrenia target directly or indirectly, among others, the cannabinoid, cholinergic, dopamine, estrogen, GABA, glutamatergic, histamine, inflammatory, immunological, ion channel, melatonin, noradrenaline, opioid, phosphodiesterase, serotonin, sigma, and trace amine associated receptor (TAAR) systems (see Table 1 and supplementary information). Across 176 identified phase 2 or 3 trials, only 12 molecules that were tested in 42 trials have so far outperformed placebo on primary outcomes in 13 positive trials (see Table 1).

For total symptoms of schizophrenia, a 5-week phase 2 trial (NCT03697252) showed that KarXT (containing a fixed combination of the muscarinic M1/M4 agonist xanomeline plus the non-centrally acting anticholinergic trospium chloride), given twice daily, outperformed placebo (effect size = 0.75), without relevant cardiometabolic or neuromotor adverse effects, but with some modest and mostly time-limited anticholinergic adverse events[56,57]. In August 2022, positive topline results for the primary outcome total Positive and Negative Syndrome Scale (PANSS) score (effect size = 0.61) and secondary outcomes have been released for the first of two similarly designed, placebo-controlled phase 3 studies in patients with acutely exacerbated schizophrenia (NCT04659161). The second phase 3 trial of KarXT in monotherapy vs. placebo (NCT04738123), as well as one 6-week trial in patients with residual positive symptoms testing KarXT in an augmentation design (NCT05145413), are ongoing.

Moreover, in a small, 6-week, phase 1B study (which is therefore not included in Table 1), emraclidine, an M4 positive allosteric modulator, also separated from placebo both in the 20 mg bid and 30 mg qd dose arms (NCT04136873). Results are being followed up in two 6-week phase 2 trials testing 10 mg and 30 mg qd (NCT05227690) as well as 15 mg and 30 mg qd (NCT05227703) vs. placebo.

Ulotaront, a TAAR-1 and 5-HT1A agonist, outperformed placebo in a 4-week, phase 2 trial in patients with schizophrenia aged 40 or younger and with no more than two prior lifetime hospitalizations for exacerbation of schizophrenia, without relevant neuromotor or cardiometabolic adverse effect risk (NCT02969382)[58]. Three additional placebo-controlled trials are ongoing (NCT04825860, NCT04072354, NCT04092686), extending the age until 65 years and being less restrictive about prior number of hospitalizations. Additionally, ralmitaront, a TAAR-1 partial agonist, is undergoing phase 2 testing (NCT04512066, NCT03669640).

Brilaroxazine, a D2, D3, D4, 5-HT1A, 5-HT2A partial agonist, and 5-HT2B, 5-HT6, 5-HT7 antagonist, was superior to placebo in a 4-week phase 2 trial (NCT01490086)[59], and a phase 3 trial has recently started (NCT05184335). Two phase 3 trials (NCT03893825, NCT03503318) have been completed for a novel subcutaneous once monthly and every two months injected long-acting formulation of risperidone, TV-46000, confirming the efficacy of other formulations of this drug in the acute treatment and relapse prevention of schizophrenia.

Raloxifene, an estrogen receptor modulator, improved PANSS total, general and negative symptoms in a phase 3 trial in postmenopausal women with schizophrenia (NCT01573637)[60], but another phase 3 trial showed inferior efficacy compared with placebo (NCT01280305)[61]. Melatonin also improved PANSS total symptoms more than placebo in a phase 2 trial (NCT01593774)[62].

For positive symptoms (co-primary outcome), a phase 2 trial (NCT02006628) showed that adjunctive cannabidiol outperformed placebo after six weeks of treatment[63]. While a significant difference was also reported for Clinical Global Impression - Severity (CGI-S), cannabidiol was not superior to placebo regarding total symptoms (co-primary outcome). Finally, estradiol outperformed placebo on PANSS positive symptoms after eight weeks of treatment in a phase 2 trial (NCT03848234)[64].

For negative symptoms of schizophrenia, the 5-HT2A inverse agonist/antagonist pimavanserin (approved for Parkinson's disease psychosis and under review for dementia-related psychosis) had one positive phase 2 study with regards to the primary outcome, Negative Symptom Assessment-16 (NSA-16) total scale change, but without greater improvement versus placebo in CGI-S and other negative symptom assessment scales (NCT02970305)[65].

Targeting schizophrenia patients with residual psychotic symptoms, a phase 3 trial reported no improvement of total symptoms with adjunctive pimavanserin in the entire sample, but there were favorable results in the approximately 80% European subsample, and significant improvements in negative symptoms and CGI-S in the total sample (NCT02970292).

Roluperidone, a 5-HT2A and sigma-2 receptor antagonist, had one successful phase 2 trial (EU2014-004878-42) for negative symptoms[66], albeit in the context of an unusually low placebo response. The subsequent phase 3 trial (NCT03397134) was suggestive of efficacy, but missed statistical significance versus placebo in the intent-to-treat analysis[67]. A potential complication is that this drug has been tested only in monotherapy, i.e., in patients with schizophrenia who were off traditional dopamine receptor blockers or partial agonists, without documentation that it is effective on total and positive symptoms.

Concerning cognitive dysfunction in schizophrenia, a phase 3 clinical trial programme follows up on a successful phase 2 study with BI 425809 (NCT02832037), a glycine transporter-1 inhibitor, that outperformed placebo at week 12 on MATRICS Consensus Cognitive Battery[68], but not on the Schizophrenia Cognition Rating Scale (SCoRS), which measures functional impact of cognitive improvement, a required co-primary endpoint for regulatory approval of agents targeting cognitive dysfunction in schizophrenia.

Regarding the management of adverse events of already approved antipsychotics in schizophrenia, glycopyrrolate (a muscarinic receptor antagonist) improved sialorrhea more than placebo in a phase 2 trial (EU2012-002299-15)[69].

While a number of trials targeting multiple mechanisms of action are ongoing or

Table 1 Medications for schizophrenia with positive results in phase 2 or 3 randomized controlled trials

Drug	Mechanisms of action	Control	Duration (weeks)	Phase	NCT/EudraCT number	Status	Results
BI 425809	Glycine transporter-1 inhibitor	Placebo	26	3	NCT04860830	R	No results available
BI 425809		Placebo	26	3	NCT04846868	R	No results available
BI 425809		Placebo	26	3	NCT04846881	R	No results available
BI 425809		Placebo	12	2	NCT03859973	R	No results available
BI 425809		Placebo	26	3	EU2020-003726-23	O	No results available
BI 425809		Placebo	12	2	NCT02832037	C	Superior on cognition
Brilaroxazine	Dopamine-5-HT partial agonist, 5-HT antagonist	Placebo, Aripiprazole	4	2	NCT01490086	C	Superior (PANSS)
Brilaroxazine		Placebo	4	3	NCT05184335	R	No results available
Cannabidiol	Multiple (among others, binds to CB1/CB2 receptors, activates 5-HT1A receptors, antagonizes alpha-1 adrenergic and mu opioid receptors, inhibits synaptosomal uptake of noradrenaline, dopamine, serotonin and GABA)	Placebo	26	2	NCT02926859	ANR	No results available
Cannabidiol		Placebo, Olanzapine	4	2	NCT02088060	ANR	No results available
Cannabidiol		Placebo	10	2	NCT02504151	ANR	No results available
Cannabidiol		Placebo	8	3	NCT04411225	R	No results available
Cannabidiol		Risperidone	7	2	NCT04105231	R	No results available
Cannabidiol		Placebo	12	2	NCT04421456	R	No results available
Cannabidiol		Placebo	6	2	NCT02006628	C	Superior on PANSS positive, CGI-S
Estradiol	Estrogen receptor agonist	Placebo	8	3	NCT03848234	C	Superior on PANSS positive
Estradiol		Placebo	16	3	NCT04093518	R	No results available
Glycopyrrolate	Muscarinic receptor antagonist	Placebo	1	3	EU2012-002299-15	C	Superior on sialorrhea
Melatonin	Melatonin receptor agonist	Placebo	24	4	NCT01431092	C	Data available for a subsample of 48 participants
Melatonin		Placebo	8	2	NCT01593774	C	Superior on PANSS total
Pimavanserin	5-HT2A inverse agonist/ antagonist	Placebo	26	3	NCT04531982	R	No results available
Pimavanserin		Placebo	6	3	NCT02970292	C	No effect on PANSS total
Pimavanserin		Placebo	26	2	NCT02970305	C	Superior on NSA-16
Pimavanserin		Placebo	26	3	EU2016-003437-18	C	No results available
Raloxifene	Estrogen receptor modulator	Placebo	24	3	NCT01573637	C	Superior on PANSS total, negative, general
Raloxifene		Placebo	12	3	NCT01280305	C	Inferior on PANSS total
Raloxifene		Placebo	12	4	NCT03418831	C	No results available
Raloxifene		Placebo	12	4	NCT02354001	C	No results available
Raloxifene		Placebo	12	4	NCT01481883	R	No results available
Raloxifene		Placebo	12	3	NCT03043820	R	No results available
Roluperidone	5-HT2A and sigma-2 receptor antagonist	Placebo	12	2	EU2014-004878-42	C	Superior on negative symptoms
Roluperidone		Placebo	12	3	NCT03397134 EU2017-003333-29	C	No difference in intention-to-treat analysis, superior on negative symptoms in modified intention-to-treat analysis
TV-46000 (subcutaneous risperidone)	Dopamine antagonist	Placebo	56	3	NCT03893825	C	Superior in acute and long-term treatment
TV-46000 (subcutaneous risperidone)		Placebo	108	3	NCT03503318	C	Superior on relapse prevention

Table 1 Medications for schizophrenia with positive results in phase 2 or 3 randomized controlled trials *(continued)*

Drug	Mechanisms of action	Control	Duration (weeks)	Phase	NCT/EudraCT number	Status	Results
Ulotaront	TAAR-1/5-HT1A agonist	Quetiapine XR	52	3	NCT04115319	R	No results available
Ulotaront		Placebo	4	2	NCT02969382	C	Superior on PANSS total
Ulotaront		Placebo	6	2/3	NCT04825860	R	No results available
Ulotaront		Placebo	5	3	NCT04072354	R	No results available
Ulotaront		Placebo	6	3	NCT04092686	R	No results available
Xanomeline + Trospium Chloride (KarXT)	M1/M4 muscarinic agonist, peripheral muscarinic antagonist	Placebo	5	2	NCT03697252	C	Superior on PANSS total
Xanomeline + Trospium Chloride (KarXT)		Placebo	5	3	NCT04738123	R	No results available
Xanomeline + Trospium Chloride (KarXT)		Placebo	5	3	NCT04659161	C	Superior on PANSS total
Xanomeline + Trospium Chloride (KarXT)		Placebo	6	3	NCT05145413	R	No results available

NCT/EudraCT number – number in clinicaltrials.gov or clinicaltrialsregister.eu, R – recruiting, O – ongoing, C – completed, ANR – active, not recruiting, TAAR-1 – trace amine-associated receptor-1, PANSS – Positive and Negative Syndrome Scale, CGI-S – Clinical Global Impression - Severity, NSA-16 – Negative Symptom Assessment-16. Results without information on statistical significance are classified among "results not available".

have been completed without available results (see supplementary information), the currently most promising targets for schizophrenia appear to be M1/M4 muscarinic receptor agonism, M4 muscarinic positive allosteric agonism, TAAR-1 agonism, and dopamine-serotonin partial agonism/serotonin antagonism. Due to mixed/inconclusive findings, questions remain about 5-HT2A inverse agonism/antagonism for negative and residual psychotic symptoms, and 5-HT2A/sigma-2 antagonism for negative symptoms, as well as about glycine transporter-1 inhibition for improvement of cognitive dysfunction, that is required to also significantly improve functionality to gain regulatory approval.

Bipolar disorder

Agents in development for the treatment of bipolar disorder target directly or indirectly, among others, the cholinergic, dopamine, GABA, glutamatergic, inflammatory, immunological, ion channel, melatonin, neurotrophic, noradrenaline, and serotonin systems (see Table 2 and supplementary information). Across 38 identified trials, only six molecules that were tested in 11 trials outperformed placebo on primary outcomes in six positive trials (see Table 2).

For bipolar depression, N-acetyl cysteine (a glutathione precursor) plus acetylsalicylic acid, added to treatment-as-usual, outperformed placebo regarding response in one phase 2 trial (NCT01797575)[70]. Furthermore, non-racemic amisulpride (SEP-4199) was superior to placebo at 6 weeks on the Montgomery-Asberg Depression Rating Scale (MADRS) in the US, European Union and Japanese cohorts, at doses of 200 or 400 mg/day[71,72]. Adjunctive armodafinil, an R-enantiomer of modafinil, was associated with a significantly greater reduction in the 30-item Inventory of Depressive Symptomatology, Clinician Rated (IDS-C) total score at week 8[73] in one phase 3 trial vs. placebo (NCT01072929), but two other phase 3 trials (NCT01072630 and NCT01305408) did not confirm this superiority[74,75].

D-cycloserine (an NMDA antagonist) plus lurasidone outperformed lurasidone plus placebo after an initial ketamine infusion in reducing depressive symptoms in severely depressed patients with bipolar disorder (NCT02974010)[76]. Moreover, adjunctive infliximab – a tumor necrosis factor-alpha (TNF-α) inhibitor – was superior to placebo regarding depressive symptoms in a phase 2 trial (NCT02363738), yet with no difference regarding treatment response[77-79]. Interestingly, secondary analyses suggested higher efficacy in subjects with childhood maltreatment. Ketamine outperformed placebo in a phase 2 trial targeting suicidal ideation (NCT01944293).

We did not identify any positive randomized controlled trial (RCT) for treatment of acute mania or for the maintenance treatment of bipolar disorder.

While a number of trials targeting multiple mechanisms of action are ongoing or have been completed without available results (see supplementary information), the currently most promising targets for bipolar depression are dopamine antagonism plus 5-HT7 antagonism, non-steroidal anti-inflammatory action plus glutathione precursor activity, NMDA receptor antagonism, and TNF-α inhibition. Notably, neither bipolar mania nor bipolar disorder maintenance are currently relevant targets in drug development, and the most promising agents for bipolar depression are all repurposed from different existing indications.

Major depressive disorder

Agents in development for the treatment of major depressive disorder target directly or indirectly, among others, the cannabinoid, cholinergic, dopamine, estrogen, GABA, glutamatergic, inflammatory, immunological, ion channel, neurotrophic, noradrenaline, opioid, peroxisome proliferator-activated receptor, serotonin, sigma, TAAR, and substance P systems (see Table 3 and supple-

Table 2 Medications for bipolar depression with positive results in phase 2 or 3 randomized controlled trials

Drug	Mechanisms of action	Control	Duration (weeks)	Phase	NCT/EudraCT number	Status	Results
N-acetyl cysteine + Acetylsalicylic acid	Glutathione precursor + NSAID	Placebo	16	2	NCT01797575	C	Superior on response
Amisulpride, non-racemic	Dopamine/5-HT7 antagonist	Placebo	6	2	NCT03543410	C	Superior on depressive symptoms
Armodafinil	Sympathomimetic	Placebo	8	3	NCT01072630	C	No difference
Armodafinil		Placebo	8	3	NCT01072929	C	Superior on depressive symptoms
Armodafinil		Placebo	8	3	NCT01305408	C	No difference
D-cycloserine + Lurasidone	NMDA antagonist + dopamine antagonist	Lurasidone + Placebo	6	2	NCT02974010	C	Superior on depressive symptoms
Infliximab	TNF-α inhibitor	Placebo	12	2	NCT02363738	C	Superior on depressive symptoms
Ketamine	NMDA antagonist	Midazolam	28	3	NCT04939649	R	No results available
Ketamine		Placebo	2	2	NCT05004896	NYR	No results available
Ketamine		Midazolam	2	2	EU2016-002068-14	C	No results available
Ketamine		Midazolam	1 day	2	NCT01944293	C	Superior on suicidal ideation

NCT/EudraCT number – number in clinicaltrials.gov or clinicaltrialsregister.eu, R – recruiting, C – completed, NYR – not yet recruiting, NSAID – non-steroidal anti-inflammatory drug, NMDA – N-methyl-D-aspartate, TNF-α – tumor necrosis factor alpha. Results without information on statistical significance are classified among "results not available".

mentary information). Across 177 identified trials, 19 molecules that were tested in 43 trials outperformed placebo on primary outcomes in 19 positive trials (see Table 3).

Cariprazine, a D3-preferring D3/D2 partial dopamine agonist with antagonist activity at 5-HT2B and 5-HT2A receptors, is currently under FDA review as augmentation in major depressive disorder, following a positive phase 3 trial (NCT03738215) and one partially positive phase 2 trial (at 2-4.5 mg/day, but not at 1-2 mg/day) (NCT01469377)[80], alongside a negative trial (NCT03739203). Lurasidone, a 5-HT2A-D2 antagonist with 5-HT7 antagonism, was superior to placebo in a phase 3 trial of subjects with major depressive disorder and mixed features (NCT01421134)[81].

The extended release (ER) formulation of levomilnacipran, a serotonin-noradrenaline reuptake inhibitor, outperformed placebo in a phase 3 trial (NCT01377194)[82], although the switch to levomilnacipran ER was not superior to quetiapine plus antidepressants in another phase 3 trial (NCT 02720198). Pimavanserin, a 5-HT2A antagonist/inverse agonist, had a positive phase 2 sequential parallel comparison design study (positive in stage 1+2 and 1, but not in stage 2) as augmentation in major depressive disorder (NCT03018340)[83], followed by a negative standard phase 3 study (NCT03968159) compared to placebo.

With the FDA approval of intranasal esketamine[84] and the widespread off-label use of racemic ketamine, both intravenously and intranasally, for resistant depression[85,86], the field of psychopharmacology has seen a renewed focus on the development of antidepressant therapies that modulate the glutamatergic system.

One such agent is AXS-05, the combination of dextromethorphan with low-dose bupropion, whose pharmacological actions are non-competitive NMDA receptor antagonism, sigma-1 receptor agonism, nicotinic acetylcholine receptor antagonism, and inhibition of serotonin, noradrenaline and dopamine transporters. In two phase 2 trials, AXS-05 was superior to low-dose bupropion[87] (NCT03595579) or to placebo (NCT04019704) on the MADRS at week 6, leading to FDA approval for major depressive disorder in August 2022. For treatment-resistant depression, AXS-05 showed in a one-year study significantly delayed time to relapse (primary outcome) and decreased relapse rate (secondary outcome) (NCT04608396); however, it did not separate from bupropion 150 mg/day in a 12-week study (NCT02741791).

A second anti-glutamatergic agent is esmethadone, an NMDA receptor antagonist with very weak opioid mu agonism, which is being developed as an augmenting agent in treatment-resistant depression, following a positive phase 2 trial (NCT03051256)[88]. The phase 3 programme is ongoing, with three 4-week placebo-controlled studies (NCT04855747, NCT05081167, NCT04688164). A single dose of rapastinel, a NMDA partial agonist, was superior to placebo, when given at 5 or 10 mg, but not 1 mg, in a phase 2 trial (NCT01234558)[89], but three phase 3 trials were negative (NCT02951988, NCT02943564, NCT02943577).

There has also been significant interest in GABAergic modulation for the treatment of depression. Following FDA approval of the intravenous GABA-A receptor positive allosteric modulator brexanolone in postpartum depression[90,91], the orally administered zuranolone, which is also a neuroactive steroid binding to GABA-A receptors, is being developed for both postpartum depression and major depressive disorder. Zuranolone had a positive phase 2 study in severe postpartum depression, despite a large placebo response (NCT02978326)[92]. A second trial for postpartum depression is

Table 3 Medications for major depressive disorder with positive results in phase 2 or 3 randomized controlled trials

Drug	Mechanisms of action	Control	Duration (weeks)	Phase	NCT number	Status	Results
Ayahuasca	5-HT multimodal modulator, TAAR-1 and sigma-1 agonist	Placebo	1	2	NCT02914769	C	Superior on HAM-D
Botulinum toxin type A neurotoxin complex	Acetylcholine release inhibitor	Placebo	12	2	NCT01392963	C	Superior on HAM-D
Buprenorphine + Samidorphan + Antidepressant	Kappa opioid agonist + mu opioid antagonist	Placebo + Antidepressant	4	2	NCT01500200	C	Superior on HAM-D (only 2 + 2 mg/day)
Buprenorphine + Samidorphan + Antidepressant		Placebo + Antidepressant	6	3	NCT02218008	C	Superior on MADRS
Buprenorphine + Samidorphan + Antidepressant		Placebo + Antidepressant	6	3	NCT03188185	C	No difference
Buprenorphine + Samidorphan + Antidepressant		Placebo + Antidepressant	6	3	NCT02158546	C	No difference
Buprenorphine + Samidorphan + Antidepressant		Placebo + Antidepressant	5	3	NCT02158533	C	No difference
Dextromethorphan + Bupropion (AXS-05)	NMDA antagonist, sigma-1 agonist, nicotinic acetylcholine receptor antagonist, 5-HT/noradrenaline/dopamine reuptake inhibitor	Bupropion SR	6	2	NCT04971291	R	No results available
Dextromethorphan + Bupropion (AXS-05)		Bupropion	12	3	NCT02741791	C	No superiority for treatment-resistant depression
Dextromethorphan + Bupropion (AXS-05)		Placebo	52	2	NCT04608396	C	Delayed time to relapse
Cariprazine + Antidepressant	Dopamine D3/D2 partial agonist, serotonin antagonist	Placebo + Antidepressant	8	2	NCT01469377	C	Superior on MADRS at week 8 (only 2-4.5 mg/day)
Cariprazine + Antidepressant		Placebo + Antidepressant	6	3	NCT03738215	C	Superior at week 6
Cariprazine + Antidepressant		Placebo + Antidepressant	6	3	NCT03739203	C	No difference
Esmethadone + Antidepressant	NMDA antagonist	Placebo + Antidepressant	3	2	NCT03051256	C	Superior on MADRS at week 2
Esmethadone + Antidepressant		Placebo + Antidepressant	4	3	NCT04855747	R	No results available
Esmethadone + Antidepressant		Placebo + Antidepressant	4	3	NCT05081167	R	No results available
Esmethadone + Antidepressant		Placebo+ Antidepressant	4	3	NCT04688164	R	No results available
Estradiol + Progesterone	Estrogen receptor agonist, progesterone receptor agonist	Placebo	52	2/3	NCT01308814	C	Superior on CES-D
Ezogabine	Opening of neuronal voltage activated potassium channels	Placebo	5	2	NCT03043560	C	Superior on MADRS
Levomilnacipran ER	5-HT/noradrenaline reuptake inhibitor	Quetiapine + Antidepressant	8	3	NCT02720198	C	No difference
Levomilnacipran ER		Placebo	8	3	NCT01377194	C	Superior on MADRS
Lurasidone	5-HT7, 5-HT2A and dopamine antagonist	Placebo	6	3	NCT01421134	C	Superior on MADRS
Metformin + Fluoxetine	AMP-activated protein kinase	Placebo + Fluoxetine	12	1/2	NCT04088448	C	Superior on HAM-D
Naltrexone + Antidepressant	Opioid receptor antagonist	Placebo + Antidepressant	3	2	NCT01874951	C	Superior on MADRS but not on HAM-D
Nitrous Oxide	Inhalation anesthetic	Placebo	1	2	NCT03283670	C	Superior on HAM-D
Nitrous Oxide		Placebo	1	2	NCT02139540	C	Superior on depressive symptoms at 24 hours
Nitrous Oxide		Placebo	2	2	NCT03932825	C	No results available
Nitrous Oxide		Placebo	4	2	NCT03869736	NA	No results available

Table 3 Medications for major depressive disorder with positive results in phase 2 or 3 randomized controlled trials *(continued)*

Drug	Mechanisms of action	Control	Duration (weeks)	Phase	NCT number	Status	Results
Pimavanserin + Antidepressant	5-HT2A inverse agonist/ antagonist	Placebo + Antidepressant	5	2	NCT03018340	C	Superior on HAM-D (stage 1 and 1+2, not stage 2)
Pimavanserin + Antidepressant		Placebo + Antidepressant	5	3	NCT03968159	C	No difference
Pioglitazone + Citalopram + Chlordiazepoxide	PPARγ agonist	Placebo + Citalopram + Chlordiazepoxide	6	2/3	NCT01109030	C	Superior on response (HAM-D)
Psilocybin	5-HT1A/5-HT2A agonist	Waitlist	8	2	NCT03181529	C	Superior on GRID-HAM-D
Psilocybin		Escitalopram	6	2	NCT03429075	C	No difference
Psilocybin		Placebo	5	2	NCT03715127	O	No results available
Psilocybin		Placebo	8	2	NCT04989972	O	No results available
Psilocybin		Ketamine	26	2	NCT03380442	O	No results available
Psilocybin		Placebo	4	2	NCT04620759	O	No results available
Psilocybin		Niacin	1	2	NCT04630964	O	No results available
Psilocybin		Niacin	7	2	NCT03866174	O	No results available
Psilocybin + Psychological therapy		Placebo + Psychological therapy	3	2	NCT04959253	O	No results available
Psilocybin		Placebo	4	2	NCT05259943	O	No results available
Psilocybin + Psychological therapy		Nicotinamide + Psychological therapy	6	2	NCT04670081	O	No results available
Rapastinel + Antidepressant	NMDA partial agonist	Placebo + Antidepressant	3	3	NCT02932943	C	No difference
Rapastinel		Placebo	1 dose	2	NCT01234558	C	Superior (5-10 mg, not 1 mg)
Rapastinel		Placebo	52	3	NCT02951988	C	No difference
Rapastinel + Antidepressant		Placebo + Antidepressant	6	2	NCT01684163	C	No results available
Rapastinel		Placebo	3	3	NCT02943564	C	No difference
Rapastinel		Placebo	3	3	NCT02943577	C	No difference
Zuranolone (30 mg/day)	GABA-A receptor positive allosteric modulator	Placebo	7	3	NCT02978326	C	Superior for postpartum depression on HAM-D at day 15
Zuranolone		Placebo	2	3	NCT04442503	NYR	No results for postpartum depression available
Zuranolone (30 mg/day)		Placebo	2	2	NCT03000530	C	Superior for major depression on HAM-D at day 15
Zuranolone (20 mg/day and 30 mg/day)		Placebo	2	3	NCT03672175	C	No superiority on HAM-D at day 15
Zuranolone (50 mg/day)		Placebo	2	3	NCT04442490	C	Superior for major depression on HAM-D at day 15
Zuranolone (50 mg/day) + Antidepressant		Placebo + Antidepressant	2	3	NCT04476030	C	Superior for major depression on HAM-D at day 3 (primary endpoint), but not day 15

NCT number – number in clinicaltrials.gov, R – recruiting, C – completed, O – ongoing, NYR – not yet recruiting, NA – not available, NMDA – N-methyl-D-aspartate, PPARγ – peroxisome proliferator-activated receptor gamma, TAAR-1 – trace amine-associated receptor-1, HAM-D – Hamilton Depression Rating Scale, MADRS – Montgomery-Åsberg Depression Rating Scale, CES-D – Center for Epidemiological Studies-Depression Scale. Results without information on statistical significance are classified among "results not available".

awaiting results (NCT04442503).

In patients with major depressive disorder, one study of zuranolone at 30 mg/day (NCT0300530) met the primary endpoint on the Hamilton Depression Rating Scale (HAM-D) on day 15[93]. Another monotherapy study of the drug at 50 mg/day (NCT04 442490) also met the primary endpoint of superiority vs. placebo on the HAM-D at day 15. However, high placebo response accounted for a negative study at day 15 for zuranolone 20 mg/day and 30 mg/day, despite superiority over placebo on the HAM-D in the 30 mg/day arm at days 3, 8 and 12 (NCT03672175). In a phase 3 trial (NCT04476030), zuranolone 50 mg/day co-initiated with a standard antidepressant was superior to placebo on HAM-D total score at day 3 (primary endpoint), and throughout the 2-week treatment period (key secondary endpoint), but not at day 15, confirming an effect in speeding up of efficacy.

Other mechanisms of action are also being pursued. For example, pioglitazone, an agonist of the peroxisome proliferator-activated receptor gamma, plus citalopram plus chlordiazepoxide was superior to placebo in a phase 2/3 study (NCT01109030) regarding treatment response based on HAM-D scores[94]. Naltrexone, an opioid receptor antagonist, plus antidepressants was superior to placebo plus antidepressants in a phase 2 trial in preventing relapse or symptom recurrence on the MADRS, but not the HAM-D (NCT01874951)[95].

The combination of buprenorphine, a kappa opioid agonist, with the opioid mu antagonist samidorphan as adjunctive treatment in major depressive disorder was superior to placebo in two trials (phase 2: NCT01500200; phase 3: NCT02218008)[96], but not in three other phase 3 trials (NCT0 3188185, NCT02158546, NCT02158533) [96,97], without significant separation of buprenorphine alone from placebo in a meta-analysis[98].

Ezogabine, which induces the opening of neuronal voltage activated potassium channels, was superior to placebo on the MADRS in a phase 2 trial (NCT03043560)[99]. Botulinum toxin type A neurotoxin complex, an acetylcholine release inhibitor, was superior to placebo in a phase 2 trial (NCT0139 2963)[100]. The anaesthetic nitrous oxide was superior to placebo at 24 hours in a phase 2 study (NCT02139540), and at 2 hours, 24 hours, and 1 week in another phase 2 trial (NCT03283670)[101].

Psychedelics are also being investigated increasingly, with positive findings in phase 2 trials of Ayahuasca (5-HT2A partial agonism, affinity for multiple other 5-HT receptors, TAAR-1 agonism, sigma-1 agonism) (NCT02914769)[102] and psilocybin (5-HT2A agonism) (NCT03181529)[103]. Psilocybin was also found to be not inferior to escitalopram in a phase 2 trial (NCT03429075)[104].

The combination of metformin (glucose-lowering, insulin-sensitizing) and fluoxetine (selective serotonin reuptake inhibitor) was superior to placebo plus fluoxetine on the HAM-D in a phase 1/2 trial (NCT04088448) [105]. Finally, transdermal estradiol plus intermittent micronized progesterone (NCT01308 814) was more efficacious than placebo in preventing the development of clinically significant depressive symptoms among initially euthymic peri-menopausal and early post-menopausal women in a phase 2/3 study[106].

While a number of trials targeting multiple mechanisms of action are ongoing or have been completed without available results (see supplementary information), the currently most promising targets for major depressive disorder appear to be D3/D2 partial agonism with 5-HT2A/B antagonism, D2/5-HT2A/5-HT7 antagonism, 5-HT2A antagonism/inverse agonism, NMDA receptor antagonism and partial agonism, sigma-1 receptor agonism, nicotinic acetylcholine receptor antagonism, GABA-A receptor positive allosteric modulation, peroxisome proliferator-activated receptor gamma agonism, opening of neuronal voltage activated potassium channels, acetylcholine release inhibition, and 5-HT2A agonism.

Anxiety and trauma-related disorders

Agents in development for the treatment of anxiety and trauma-related disorders target directly or indirectly, among others, the cannabinoid, cholinergic, dopamine, GABA, glucocorticoid, glutamatergic, melatonin, noradrenaline, oxytocin, serotonin, and substance P systems (see Table 4 and supplementary information). Across 98 identified trials, only nine molecules that were tested in 31 trials outperformed placebo on primary outcomes in 18 trials (see Table 4).

In PTSD, intranasal oxytocin was more effective than placebo on amygdala connectivity in a phase 2 trial (EU2012-001288-58), and 3,4-methylenedioxy-methamphetamine (MDMA)-assisted psychotherapy (via release of serotonin and noradrenaline) was superior to placebo on characteristic symptoms in four phase 2 trials (NCT00090064, NCT01211405, NCT01793610, NCT00353 938) and one phase 3 trial (NCT03537 014)[107-114], although in one trial (NCT017936 10) the superiority was not observed in intent-to-treat analysis.

In panic disorder, d-cycloserine (NMDA co-agonist) as augmentation of exposure therapy outperformed placebo on neurocognitive processing in a phase 2 trial (NCT 01680107)[115]. In social anxiety disorder, one phase 2 trial showed that d-cycloserine as augmentation of cognitive behavioral therapy (CBT) outperformed placebo (NCT 02066792)[116-119], although two other studies were negative (NCT00633984, NCT00128 401)[120-122].

In generalized anxiety disorder, ABIO 08/01 (a selective inhibitor of GABA- and glutamate-gated chloride channels) outperformed placebo on CGI in a phase 3 trial (EU2006-003643-23). Agomelatine (melatonin receptor agonist) was superior to placebo on relapse rate in one phase 3 trial (EU2006-005674-47), and on anxiety symptoms in two phase 3 trials (EU2004-002577-23, EU2009-013789-17). Pregabalin (voltage-gated calcium channel modulator) was more efficacious than placebo on anxiety symptoms in two phase 3 trials (EU2006-006339-31, EU2004-001500-13). Quetiapine extended-release (histamine antagonist, alpha-2 antagonist, noradrenaline reuptake inhibitor) was superior to placebo in two phase 3 trials on anxiety symptoms (EU2005-005054-46) and relapse rate (EU2005-005055-18). Finally, SR58611A (selective beta-3 adrenoceptor agonist) reduced anxiety symptoms more than placebo in a phase 3 trial (NCT00266747), and vortioxetine (multimodal serotonergic modulator) prevented relapse in one phase 3 trial (EU2008-001673-15).

Table 4 Medications for anxiety and trauma-related disorders with positive results in phase 2 or 3 randomized controlled trials

Drug	Mechanisms of action	Control	Duration	Phase	NCT/EudraCT number	Status	Results
Post-traumatic stress disorder (PTSD)							
Intranasal oxytocin	Oxytocin receptor agonist	Placebo	12	2	NCT04523922	R	Results not available
Intranasal oxytocin		Placebo	10	2	NCT04228289	R	Results not available
Intranasal oxytocin		Placebo	6	2	EU2012-003072-39	R	Results not available
Intranasal oxytocin		Placebo	1 dose	2	EU2012-001288-58	C	Superior effect on amygdala connectivity
MDMA	5-HT, dopamine, noradrenaline releaser	Placebo	8	2	NCT00090064	C	Superior on PTSD symptoms and response
MDMA		Placebo	4	2	NCT01211405	C	Superior on PTSD symptoms
MDMA		Placebo	4	2	NCT01793610	C	Superior on PTSD symptoms per-protocol, not significant in intention-to-treat
MDMA		Placebo	3	2	NCT00353938	C	Superior on PTSD symptoms
MDMA		Placebo	18	3	NCT03537014	C	Superior on PTSD symptoms
MDMA		Placebo	18	3	NCT04077437	R	Results not available
Panic disorder							
D-cycloserine	NMDA receptor agonist	Placebo	1 dose	2	NCT 01680107	C	Superior effect on both threat bias and amygdala response
D-cycloserine		Placebo	NA	2	EU2010-021198-35	C	Results not available
D-cycloserine		Placebo	56	2	EU2011-001398-19	C	Results not available
Social anxiety disorder							
D-cycloserine	NMDA receptor agonist	Placebo	12	3	NCT02066792	C	Superior on anxiety symptoms
D-cycloserine		Placebo	13	3	NCT00633984	C	No difference
D-cycloserine		Placebo	12	2	NCT00515879	C	Results not available
D-cycloserine		Placebo	12	2	NCT00128401	C	No difference
Generalized anxiety disorder							
ABIO 08/01	Inhibition of GABA- and glutamate-gated chloride channels	Placebo	8	3	EU2006-003643-23	C	Superior on CGI
Agomelatine	Melatonin receptor agonist	Placebo	26	3	EU2006-005674-47	C	Superior on relapse rate
Agomelatine		Placebo	12	3	EU2004-002577-23	C	Superior on anxiety symptoms
Agomelatine		Citalopram	12	2	EU2012-003699-37	C	Not inferior on anxiety symptoms
Agomelatine		Placebo	12	3	EU2009-013789-17	C	Superior on anxiety symptoms
Pregabalin	Voltage-gated calcium channel inhibitor	Placebo	8	3	EU2006-006339-31	C	Superior on anxiety symptoms
Pregabalin		Placebo	8	3	EU2004-001500-13	C	Superior to placebo on anxiety symptoms
Quetiapine fumarate	Histamine, dopamine, 5-HT, noradrenaline multimodal agent	Placebo	8	3	EU2005-005054-46	C	Superior on anxiety symptoms
Quetiapine fumarate		Placebo	52	3	EU2005-005055-18	C	Superior on relapse rate
SR58611A	Noradrenergic agonist	Placebo	10	3	NCT00252343	C	Results not available
SR58611A		Placebo	8	3	NCT00266747 EU2005-003181-41	C	Superior on anxiety symptoms
Vortioxetine	5-HT multimodal agent	Placebo	24	3	EU2008-001673-15	C	Superior on relapse rate

NCT/EudraCT number – number in clinicaltrials.gov or clinicaltrialsregister.eu, R – recruiting, C – completed, NA – not available, MDMA – 3,4-methylenedioxy-methamphetamine, NMDA – N-methyl-D-aspartate, CGI – Clinical Global Impression. Results without information on statistical significance are classified among "results not available".

Notably, no promising treatment was identified for OCD.

While a number of trials targeting multiple mechanisms of action are ongoing or have been completed without available results (see supplementary information), the currently most promising targets for anxiety and trauma-related disorders appear to be serotonin release (MDMA) for PTSD, and glutamate agonism for panic and social anxiety disorder. For generalized anxiety disorder, several candidate mechanisms have been identified, including GABA- and glutamate-gated chloride channel inhibition, melatonin receptor agonism, voltage-gated calcium channel modulation, histamine antagonism, alpha-2 antagonism, noradrenaline reuptake inhibition, selective beta-3 adrenoceptor agonism, and multimodal serotoninergic modulation. This promise reflects the capacity of at least some of these mechanisms to impact extinction-related processes.

Substance use disorders

Agents in development for the treatment of substance use disorders target directly or indirectly, among others, the cannabinoid, cholinergic, dopamine, GABA, glucocorticoid, glutamatergic, histaminergic, inflammatory, insulin, ion channel, melatonin, neurokinin, noradrenaline, opioid, orexin, oxytocin, phosphodiesterase, peroxisome proliferator-activated receptor, serotonin, and vasopressin systems (see Table 5 and supplementary information). Across 185 identified trials, ten molecules that were tested in 17 trials outperformed the control condition on primary outcomes in 12 positive trials (see Table 5).

Many agents outperforming placebo in phase 2/3 clinical trials are repurposed medications already approved for another indication. For alcohol use disorder, these include baclofen (GABA agonist), with one positive phase 3 trial (NCT01711125)[123] on time to lapse and relapse and percentage of abstinent participants; gabapentin (voltage-gated calcium channel modulator) in one phase 2 trial (NCT02349477)[124] on "proportion with heavy drinking"; ibudilast (phosphodiesterase 4 inhibitor and toll-like receptor-4 antagonist, used in the treatment of asthma) in one phase 2 trial (NCT 03489850)[125] again on "proportion with heavy drinking"; and ketamine (NMDA antagonist) in one phase 2 trial (NCT0264931) [126] regarding days of abstinence.

For methamphetamine use disorder, agents with positive placebo-controlled phase 2 trials include mirtazapine (alpha-2-adrenergic, histamine-1, 5-HT2A/C and 5-HT3 antagonist) (NCT01888835)[127], and the combination of naltrexone (opioid antagonist) and extended-release bupropion (noradrenaline-dopamine reuptake inhibitor, nicotinic receptor antagonist, non-selective serotonin reuptake inhibitor and sigma-1 receptor agonist) (NCT03078075)[128], both on the number of substance-positive urine samples.

In amphetamine use disorder, sustained-release methylphenidate (noradrenaline and dopamine reuptake inhibitor) reduced the number of substance-positive urine samples vs. placebo among dependent individuals with comorbid attention-deficit/hyperactivity disorder in a phase 2 trial.

For cocaine use disorder, drugs outperforming controls include AFQ056 (metabotropic glutamate receptor antagonist) on the proportion of cocaine use days in a phase 2 trial (NCT03242928); ketamine (NMDA antagonist) on motivation to quit cocaine and on cue-induced craving in a phase 2 trial (NCT01790490)[129]; and zonisamide (voltage-sensitive sodium channel blocker and allosteric GABA receptor agonist) on Visual Analog Questionnaire in a phase 1/2 trial (NCT01137890),

For nicotine use disorder, the combination of zonisamide plus varenicline was superior on self-reported smoking and nicotine withdrawal, but not on biochemically verified smoking, in a phase 1/2 trial (NCT01685996)[130]. For opioid use disorder, positive findings are available for cortisol on craving in users with low, but not medium or high, daily heroin intake in a phase 2 trial (NCT01718964)[131].

While a number of trials targeting multiple mechanisms of action are ongoing or have been completed without available results (see supplementary information), the currently most promising targets for substance use disorders appear to be calcium channel modulation, GABA agonism, phosphodiesterase 4 inhibition, toll-like receptor 4 antagonism and glutamate antagonism for alcohol use disorder; opioid antagonism, multimodal adrenergic and serotonergic modulation, and noradrenaline/dopamine reuptake inhibition for amphetamine/methamphetamine use disorder; glutamate antagonism and sodium channel blockade for cocaine use disorder; sodium channel blockade for nicotine use disorder; and glucocorticoid receptor agonism for opioid use disorder. However, positive results have mainly involved medications already marketed for other disorders, while novel mechanisms of action have yielded much less positive results, despite strong ongoing efforts.

Dementia

Agents in development for the treatment of dementia-spectrum disorders target directly or indirectly, among others, the cholinergic, dopamine, GABA, glucocorticoid, glutamatergic, histaminergic, immunological, inflammatory, insulin, ion channel, neuroprotection, phosphodiesterase, peroxisome proliferator-activated receptor, serotonin, and sigma systems; and additionally include vaccines against beta-amyloid or tau protein, mesenchymal stem cells, and antibodies (see Table 6 and supplementary information). Across 265 identified trials, only 14 molecules that were tested in 27 trials outperformed placebo on primary outcomes in 15 trials (see Table 6).

Among trials targeting cognition or disease-modifying markers, positive phase 2 trials included those investigating acitretin (retinoid X receptor agonist) (NCT01078168), insulin glulisine (insulin signaling inhibitor) (NCT01436045), neflamapimod (MAP kinase inhibitor) (NCT04001517), ORM-12741 (selective antagonist of alpha-2C adrenoceptors) (NCT01324518)[132], sargramostim (granulocyte-macrophage colony-stimulating factor) (NCT01409915)[133], and rasagiline (monoamine oxidase-B inhibitor) (NCT02359552)[134].

Among trials aiming to improve behavioral and psychiatric symptoms in people with dementia, brexpiprazole, a dopamine partial agonist (NCT01862640, phase 3)[135]; dextromethorphan/quinidine, a sigma-1 agonist/NMDA antagonist/multimodal agent

Table 5 Medications for substance use disorders with positive results in phase 2 or 3 randomized controlled trials

Drug	Mechanisms of action	Control	Duration (weeks)	Phase	NCT/EudraCT number	Status	Results
Alcohol use disorder							
Baclofen	GABA agonist	Diazepam	1	3	NCT03293017	R	Results not available
Baclofen		Placebo	12	3	NCT01711125	C	Superior on time to lapse and relapse and percentage abstinent
Gabapentin	Voltage-gated calcium channel modulator	Placebo	24	2	NCT02349477	C	Superior on proportion with heavy drinking
Gabapentin		Placebo	9	2	NCT03205423	ANR	Results not available
Gabapentin XR		Placebo	25	2	NCT02252536	C	Results not available
Ibudilast	Phosphodiesterase 4 inhibitor and toll-like receptor-4 antagonist	Placebo	2	2	NCT03489850	C	Superior on proportion with heavy drinking
Ibudilast		Placebo	12	2	NCT03594435	R	Results not available
Ketamine	NMDA antagonist	Placebo	24	2	NCT02649231	C	Superior on days abstinent
Amphetamine/methamphetamine use disorder							
Mirtazapine	Alpha-2 adrenergic, histamine-1, 5-HT2A/C and 5-HT3 antagonist	Placebo	24	2	NCT01888835	C	Superior on substance-positive urine samples
Mirtazapine		Placebo	18	3	NCT02541526	NA	Results not available
Naltrexone + Bupropion ER	Opioid receptor antagonist + noradrenaline/dopamine reuptake inhibitor	Placebo	12	3	NCT03078075	C	Superior on substance-positive urine samples
Sustained-Release Methylphenidate	Noradrenaline/dopamine reuptake inhibitor	Placebo	24	2	EU2006-002249-35	C	Superior on substance-positive urine samples
Cocaine use disorder							
AFQ056	Metabotropic glutamate receptor antagonist	Placebo	14	2	NCT03242928	C	Superior (proportion of cocaine use days)
Ketamine	NMDA antagonist	Lorazepam	1 day	2	NCT01790490	C	Superior on motivation to quit cocaine and on cue-induced craving
Zonisamide	Voltage-gated sodium channel blockade, allosteric GABA receptor agonism	Placebo	5	1/2	NCT01137890	C	Superior on Visual Analog Questionnaire
Nicotine use disorder							
Zonisamide + Varenicline	Voltage-gated sodium channel blockade, allosteric GABA receptor agonism	Placebo	10	1/2	NCT01685996	C	Superior on self-reported smoking, nicotine withdrawal, but not on biochemically verified smoking
Opioid use disorder							
Cortisol	Glucocorticoid receptor agonist	Placebo	1	2	NCT01718964	C	Superior on craving in users with low daily heroin intake

NCT/EudraCT number – number in clinicaltrials.gov or clinicaltrialsregister.eu, R – recruiting, C – completed, ANR – active, not recruiting, NA – not available, NMDA – N-methyl-D-aspartate. Results without information on statistical significance are classified among "results not available".

(NCT01584440, phase 2)[136]; and the CB1/2 partial agonist nabilone (NCT02351882, phase 2/3)[137] each improved agitation. Additionally, AVP-786 (deuterated form of dextromethorphan/quinidine) improved agitation in one phase 3 trial (NCT02442765), but not in another one (NCT02442778)[138]. Furthermore, two orexin receptor 1 and 2 antagonists – lemborexant (NCT03001557, phase 2)[139] and suvorexant (NCT02750306, phase 3)[140] – improved restlessness and sleep, respectively.

AXS-05, the combination of dextromethorphan with low-dose bupropion – whose pharmacological actions are non-competitive NMDA receptor antagonism, sigma-1 receptor agonism, nicotinic acetylcholine receptor antagonism, and inhibition of serotonin, noradrenaline and dopamine transporters – was found superior to placebo on agitation in a phase 2/3 trial (NCT03226522)[141], with another trial ongoing (NCT04797715).

Pimavanserin, a 5-HT2A receptor antagonist/inverse agonist, with lesser activity as a 5-HT2C antagonist/inverse agonist, outperformed placebo for relapse of de-

Table 6 Medications for dementia with positive results in phase 2 or 3 randomized controlled trials

Drug	Mechanisms of action	Control	Duration (weeks)	Phase	NCT number	Status	Results
Acitretin	Retinoid X receptor agonist	Placebo	4	2	NCT01078168	C	Superior on cerebrospinal fluid soluble alpha-cleaved amyloid precursor protein concentration
Insulin glulisine	Insulin receptor agonist	Saline	0.14	2	NCT01436045	C	Superior on cognitive performance
Neflamapimod	MAP kinase inhibitor	Low dose	12	2	NCT02423122	C	Results not available
Neflamapimod		Low dose	12	2	NCT02423200	C	Results not available
Neflamapimod		Placebo	24	2	NCT03402659	C	Results not available
Neflamapimod		Placebo	13	2	NCT03435861	R	Results not available
Neflamapimod		Placebo	16	2	NCT04001517	C	Superior on neuropsychological symptoms
ORM-12741	Alpha-2C adrenoceptor antagonist	Placebo	12	2	NCT01324518	C	Superior on cognition
ORM-12741		Placebo	12	2	NCT02471196	C	Results not available
Rasagiline	MAO-B inhibitor	Placebo	24	2	NCT02359552	C	Superior on FDG-PET measures and quality of life
Sargramostim	Granulocyte-macrophage colony-stimulating factor	Placebo	20	2	NCT01409915	C	Superior on MMSE
Sargramostim		Saline	30	2	NCT04902703	NYR	Results not available
AVP-786	NMDA antagonist, sigma-1 receptor agonist	Placebo	12	3	NCT02442778	C	Not superior on agitation
AVP-786		Placebo	12	3	NCT02442765	C	Superior on agitation
AVP-786		Placebo	12	3	NCT03393520	O	Results not available
Dextromethorphan + Bupropion (AXS-05)	NMDA antagonist, sigma-1 agonist, nicotinic acetylcholine receptor antagonist, serotonin/ noradrenaline/dopamine reuptake inhibitor	Buproprion + Placebo	5	2/3	NCT03226522	C	Superior for agitation
Dextromethorphan + Bupropion (AXS-05)		Placebo	26	3	NCT04797715	O	No results available
Brexpiprazole	Dopamine partial agonist	Placebo	12	3	NCT01922258	C	No difference
Brexpiprazole		Placebo	12	3	NCT01862640	C	Superior in improving agitation
Dextromethorphan/ quinidine	NMDA antagonist, sigma-1 receptor agonist	Placebo	6	3	NCT03854019	R	Results not available
Dextromethorphan/ quinidine		Placebo	10	2	NCT01584440	C	Superior on aggression and agitation
Lemborexant	Orexin receptor antagonist	Placebo	4	2	NCT03001557	C	Superior on restlessness
Nabilone	Cannabinoid receptor partial agonist	Placebo	14	2/3	NCT02351882	C	Superior on agitation
Nabilone		Placebo	8	3	NCT04516057	R	Results not available
Pimavanserin	5-HT inverse agonist/ antagonist	Placebo	6	2	NCT02035553	C	Superior on psychotic symptoms
Pimavanserin		Placebo	26	3	NCT04797715	C	Superior on relapse of psychosis
Suvorexant	Orexin receptor antagonist	Placebo	4	3	NCT02750306	C	Superior on total sleep time

NCT number – number in clinicaltrials.gov, R – recruiting, C – completed, O – ongoing, NYR – not yet recruiting, NMDA – N-methyl-D-aspartate, MAO – monoamine oxidase, FDG-PET – 18F-fluorodeoxyglucose-positron emission tomography, MMSE – Mini Mental State Examination. Results without information on statistical significance are classified among "results not available".

mentia-related psychosis in one phase 2 (NCT02035553)[142,143] and one phase 3 trial (NCT03325556)[144].

While a number of trials targeting multiple mechanisms of action are ongoing or have been completed without available results (see supplementary information), the currently most promising targets for dementia appear to be retinoid X receptor antagonism, insulin signaling inhibition, MAP kinase inhibition, selective antagonism of alpha-2C adrenoceptors, and granulocyte-macrophage colony-stimulation. Dopamine partial agonism, sigma-1 agonism/NMDA antagonism, and CB1/2 partial agonism appear to be promising mechanisms to improve agitation, and orexin receptor inhibition to improve restlessness and sleep. For dementia-related psychosis, 5-HT2A inverse agonism/antagonism has shown promising results.

However, it is difficult to predict the most promising pharmacological targets for the treatment of the core features of dementia,

and in particular of Alzheimer's disease. Although a substantial proportion of ongoing trials test anti-amyloid and, more recently, anti-tau treatments, all phase 2 and 3 trials in this area have not shown statistical significance on their primary outcomes, except for one phase 3 trial, albeit only in sub-analyses, leading to the controversial approval of aducanumab[145]. Therefore, there is scant available evidence to suggest that the ongoing trials of anti-amyloid and anti-tau treatments will be successful. Anti-inflammatory, metabolic, neuroprotective and cholinergic targets are all viable, but have not been substantially researched.

TRENDS AIMED TO DE-RISK TRIAL PROGRAMMES OF NOVEL AGENTS

The previous overview of the currently active phase 2 and 3 clinical trials of new pharmacotherapies for the main mental disorders indicates that a large number of chemical entities and potentially useful mechanisms of action are undergoing testing. This large activity and investment are motivated and justified by the frequency and impact of the targeted mental health conditions.

However, many, if not most, of these programmes will not yield an approved medication that can be used in clinical care. Why is this so? What must we learn and consider and what can be done to minimize the failure rate? What follows is a critical discussion of the basic tenants, challenges, opportunities and potential solutions with regards to clinical trial methodology, conduct and interpretation. This analysis should help inform future psychopharmacological research with the aim to de-risk trial programmes of novel agents or of known agents for novel psychiatric indications, increasing their chances of success.

Validity and power of clinical trials

Over the past 70 years, psychopharmacology trials have evolved considerably[146]. The RCT has become the cornerstone of clinical research aimed at obtaining regulatory approval for pharmacological agents. It is meant to provide consumers (clinicians, policy makers, patients, families, other researchers) with an accurate assessment of the efficacy/effectiveness and safety of a treatment in a population of patients at risk for or with a disorder.

Since a misleading answer may cause harm, the prime consideration in RCTs is *validity*, i.e., minimizing the probability of a misleading endorsement of an ineffective or unsafe treatment. The usual criterion is that a treatment endorsement must be true "beyond reasonable doubt", with less than a 5% chance of being wrong. However, consumers also have a major stake in rapid identification of safe and effective treatments, as do researchers who conduct RCTs and their funders. Thus, *power* is also important, i.e., the probability of endorsement if the treatment is indeed effective and safe enough in that population to warrant clinical use.

The foundation on which every RCT is based is *a priori* exploration. This process includes a review of the research literature concerning the disorder or target symptom of interest, those liable to that disorder, treatments already available and their effectiveness and safety. It includes relevant results of studies on animals, pre-post or case-control studies on patients, and *post-hoc* exploration of previously performed relevant RCTs. Finally, pilot studies may be performed to assess the feasibility or viability of the strategies considered for the proposed RCT. Important information gleaned from pilot studies include target engagement (if a biological effect is hypothesized via specific mechanisms), patient selection and possibly patient enrichment for the studied mechanism or increase in treatment effect, optimal trial duration, treatment doses, need for dose titration, selection of assessments with maximum precision and sensitivity to change, and potentially required stratification of factors that may affect treatment efficacy or safety and that need to be balanced between treatment groups. The strongest the rationale for the RCT, the more de-risked the trial will be.

This sequential process is necessary for three reasons. First, it allows the formulation of the *a priori* hypothesis, i.e., the statement of what it is exactly hoped the RCT will prove (recorded in RCT registration), that, if true, would lead to regulatory drug approval and advance clinical decision-making. Second, it is unethical to randomize patients unless the RCT researchers are in "clinical equipoise", i.e., there must be a rationale and empirical justification for thinking that the hypothesis may be true and important, but also reasonable doubt as to whether it is true or not. Ethical issues stem primarily from a concern about putting the burden of participation on patients in an RCT with little hope of advancing clinical knowledge, either because the hypothesis is unlikely to be true or because it has already been shown to be true without reasonable doubt. Another reason for the clinical equipoise is methodological in nature. There are scores of decisions that researchers must make in the conduct of an RCT. If they already "know" the "right" answer, they are likely (consciously or unconsciously) to bias decisions in the direction of their "right" answer, increasing the risk of an invalid RCT. Third, the best choice for every one of those scores of decisions depends on what is known from *a priori* exploration. The more the information from careful exploration guides the RCT design, the greater the validity and power of that RCT.

Adaptive trial designs

Several aspects of the trial design can affect the chances of finding significant differences between active and control arm. Traditional non-adaptive trial designs that do not account for evidence generated by the initial stages of the trial, and apply a one-design-fits-all-trial-stages approach, miss the low hanging fruit of adapting randomization and analytic plans based on accruing data generated by the trial itself[147]. By contrast, trials should be "adaptive by design" rather than being characterized by *post-hoc* protocol deviations[147,148]. Early learning stage trials (e.g., minimally effective or toxicity dose) are typically necessary before confirmatory trials, that are instead needed for drug approval from regulatory agencies. The earlier trials need stronger control for type II error (false negatives), and less so for type I errors (false positive), which are instead crucial in phase 2 and 3 trials.

One aspect that can be adapted in terms of design is drug dose. Typically, drug dose is set *a priori*, and tested in different arms, with many patients exposed to drug doses that are not effective, and not necessarily safe. Being able to identify the optimal dose of a medication as soon as possible in an RCT is important, because it could minimize exposure to medication doses that are not effective and potentially not safe, reduce RCT duration, and decrease costs. The continual reassessment method is a Bayesian approach leveraging dose-response curves to identify the maximum tolerated dose (MTD), allowing to promptly set dose around MDT during early stages of trial. MTD design is frequently used in oncology and neurology (in particular in studies on stroke), but it can be adapted to needs of any field[149,150]. The need of identifying MTD, as opposed to *a priori* estimating it, has the additional benefit of avoiding expensive and frequently underpowered trials with multiple arms with different doses. However, there are additional challenges when dose-response-based adaptive designs are implemented in efficacy and approval-aiming trials, given that frequently a dose range, rather than a single dose, more appropriately meets real-world patients' needs.

A second aspect that can be adapted is randomization. While randomization accounts for allocation bias with large sample size, it does not warrant balance in arm assignment across different levels of variables that are potentially influencing safety or efficacy. Hence, potential unbalanced distribution of moderators/mediators of the outcome of interest can affect the whole trial success. To overcome this limitation, covariate adaptive randomization can be applied, which randomizes allocation within matched levels of putative prognostic factors[151,152]. Additional randomization adaptive designs exist, including response adaptive randomization design, or Bayesian adaptive randomization, which however are more prone to type I error[152,153].

One further potentially adaptive trial key element is the sample size[154]. Sample size needs to be as large as possible to warrant enough statistical power to avoid type II error, and has to account for attrition rates, but also has to consider associated costs and duration, which linearly increase with the number of people to be recruited. While there is a type I error risk when using treatment-arm information to recalculate sample size, a masked (or unmasked) internal pilot method that only uses first-stage nuisance parameters can be used in phase 2 and 3 trials.

A fourth trial aspect that can be adapted by design is narrowing population characteristics, to identify subgroups of patients likely benefitting from a treatment. While including selected participants based on specific and not necessarily frequent characteristics goes in the opposite direction of inclusivity and representativeness of trial population, this so-called "enrichment" design has great value in late learning stages, consistent with the concept of precision medicine. The main downfall of enrichment design is that it yields poorly generalizable findings, and there are also concerns about their replicability in real-world confirmatory pragmatic trials, with the risk of type I error[155]. Trials already tend to select partially representative samples[156], on whom then a "super selection" would be operated. Hence, enrichment trial designs tend to be restricted to pharmacogenetic studies[157].

However, enriched sample selection can also be useful for proof of concept and fast-fail trials whereby data are used to make a decision as to whether and how or in whom to continue the drug development process of a given molecule. Successful applications of this approach have included the testing of the TAAR-1 agonist ulotaront in patients ≤40 years old and with no more than two hospitalizations for an exacerbation of schizophrenia, i.e. patients with less dopamine system alterations due to prior treatment and/or the underlying illness (see the previous overview of clinical trials on schizophrenia).

It is unclear, however, to what degree effect size and sample size calculations need to be adjusted when expanding the population to be more inclusive and less enriched. *Post-hoc* analyses of a phase 2 placebo-controlled trial in Alzheimer's dementia-related psychosis (see the previous overview of clinical trials on dementia) found that response to pimavanserin was enhanced in patients with greater baseline psychosis scores[143]. On the other hand, for Parkinson's disease-related psychosis, response to pimavanserin was greater in patients with greater cognitive impairment[158]. Similarly, *post-hoc* analyses of phase 2 trials of BI 425809, a glycine transporter inhibitor under investigation for cognitive dysfunction in schizophrenia, indicated greater response to drug in patients receiving not more than one concurrent antipsychotic, with more negative symptoms and not receiving concurrent benzodiazepines, and with the 10 mg dose in females and in patients aged 38 years or younger, a schizophrenia illness duration of 5-10 years, and worse baseline cognition[68]. Such data create decision points as to whether a trial programme should always target the entire population with a given illness, where the effect size may be diluted, or whether it would not be safer and, ultimately, more cost-effective to obtain approval for a more restricted subsample with the greatest chance of success. If data indicate viability of the treatment for the entire or a more expanded patient sample, such trials could be performed afterwards.

Moreover, enrichment designs can base their randomization on previous response, as occurs in trials conducted in stabilized patients who are randomized to continuation of study drug or a switch to placebo. Duration and degree of stability and related placebo relapse rates are important considerations when designing such trials, as shorter durations and less complete remission increase the likelihood of relapse, particularly in the placebo arm. However, one also needs to guard against spurious relapses due to rebound and withdrawal phenomena upon rapid drug discontinuation[159], which naturally occur less readily the longer the half-life of a given medication is[160]. Furthermore, in bipolar disorder, illness polarity of the pre-stabilization illness phase is largely predictive of the polarity of the next episode[161], which needs to be considered when designing relapse prevention trials. Although such enrichment has been criticized as a limitation[162], it matches and informs clinical care where those patients are continued on maintenance therapy who have responded to and tolerate the medication.

In addition to the adaptive randomiza-

tion outlined above, an additional strategy for randomization of patients is having a lead-in phase with single-blind placebo, open-label medication or double-blind placebo, basing randomization on response during this lead-in phase. In the placebo run-in stage, patients are treated with placebo, and then only those not responding to placebo are randomized to either placebo or active treatment. This design has been implemented in augmentation studies of antidepressants with second-generation antipsychotics for patients with major depression and suboptimal response to antidepressants[163], in which those improving too much during the single-blind dose optimization phase were excluded from the randomization.

While a large number of trials adopted the single-blind placebo lead-in period as a form of full enrichment of the trial in placebo non-responders, this enrichment has failed to show benefits, as suggested by a meta-analysis of 101 antidepressant trials[164] and recently replicated in a meta-analysis of 347 antidepressant trials, of which 174 used a single-blind placebo run-in period[165]. Single-blind placebo and open-label medication lead-in phases are inferior to other enrichment study designs, such as sequential parallel design[166], and have longer duration and higher costs. Accounting for costs, sample size, and duration of trials, the sequential parallel design may to be more effective for phase 3 trials aiming to regulatory approval[166].

As we have seen in the previous overview of clinical trials on major depressive disorder, sequential parallel comparison is a study design that attempts to overcome limitations of placebo lead-in stages[167-171]. Trials are structured in two stages, and can be conducted with one randomization, if the trial has two arms, or two randomizations if three arms are used (one active, two placebo). Participants are first randomized to placebo (stage 1). Then, non-responders to placebo are re-randomized again to the two treatment options (stage 2), in case of two arms trials. If a three arms trial is conducted (one active arm, two placebo arms), placebo non-responders of both placebo arms are assigned to active treatment, or placebo. Data are analyzed from the first randomization, as well as from the second randomization[172], and they are pooled in the same analysis generating one p value. It has been estimated that with this design it is possible to keep the same level of power conducting trials with 20% to 50% fewer individuals[173].

Finally, "adaptive seamless designs" are trial designs that attempt to conduct one multi-phase trial, as opposed to multiple separate learning and confirmatory trials. This design can reduce the time from phase 1 to phase 3 trials aiming to regulatory approval, implementing continuous recruitment, with intense monitoring and data analysis that can inform adaptive dose, randomization, and sample size. However, there are concerns regarding the risk of type I error in this type of design[174].

Despite adaptive designs, trials often fail. The worst-case scenario, which is far from rare, is recruiting a quite large amount of participants, e.g. 500 patients, exposing them to experimental medications, with potential safety issues and important costs, but ultimately observing no significant differences between medication and placebo. Stopping for futility is an important design that can terminate trials prematurely as soon as there is evidence of no significant effect of the interventions versus the control[175]. Several methods have been proposed to *a priori* define optimal futility thresholds, that can be applied to different study designs, including sequential trials with one or more endpoints[176,177]. Stopping for futility trials based on issues with the drug, selected doses, target population or assessments, allows to terminate trials early that are bound to ultimately fail, protecting many patients from potential adverse events of experimental medications, and saving cost and time in case the failed trial informs an improved study design and/or trial conduct[178].

A recent study investigating the potential of adaptive design trials has been submitted to the European Medicines Agency (EMA). Out of 59 adaptive design trials, 30 actually started, 23 were concluded, nine had a significant treatment effect, and four led to a market authorization[175]. Importantly, only 18 trials actually implemented the adaptive elements, which might suggest challenges in implementation of these elements. On the other hand, of these 18 trials, 11 were concluded, and six had significant findings, which points to the potential of adaptive designs[175]. Most frequently adapted elements were dose selection, sample size re-assessment, and stopping for futility[175].

Placebo response and drug-placebo difference

While the ingredients driving placebo effect can be studied and have the potential to identify safe therapeutic elements that can be exported into clinical care[35], high placebo response is a plague that affects RCTs across different mental disorders[32,38,39]. In fact, it has been suggested that some major pharmaceutical companies have diminished their investment in developing medications for mental disorders because of the challenges in signal detection due to higher than expected placebo responses.

Many regulatory agencies (such as the FDA and the EMA) as well as researchers have taken the position that to assess the efficacy of a new treatment for many mental disorders is not possible without a placebo-controlled design. Needless to say, this guidance has had enormous impact on drug development. Consequently, every psychotropic medication that has been approved for the treatment of a mental disorder in either the US or Europe in the past 30 years has been assessed in placebo-controlled clinical trials.

This practice has been challenged by the increasing reluctance of clinicians[179] and patients[180,181] to participate in such studies. In addition, ethical committees in many countries are making it increasingly difficult to conduct placebo-controlled clinical trials. Of course, when these studies are allowed, risk minimization procedures must be in place. At the same time, studies in recent years have found large dropout rates in trials utilizing placebo controls[182], as well as a decrease of the placebo-drug difference[183-186], largely driven by increasing placebo effects without similar degrees of increased drug effects.

The placebo response has increased over a period of many years in conditions such as depression, while the drug response has not[187]. In an analysis that included 167 dou-

ble-blind RCTs with 28,102 (mainly chronic) participants, it was reported that, of the response predictors analyzed, 16 trial characteristics changed over the decades[188]. However, in a multivariable meta-regression, only industry sponsorship and increasing placebo response were significant moderators of effect sizes. Drug response remained stable over time.

The magnitude of placebo effect is larger in trials on depressive disorder, bipolar depression and mania, and smaller in trials on schizophrenia[38,39]. Nevertheless, placebo effect has been increasing not only in depression[38] but also in schizophrenia over the past 24 years[189], and is a major obstacle for developing novel medications[32]. Indeed, placebo response is particularly high in trials sponsored by the industry[38]. For example, analyses of schizophrenia trials indicated an increase in total psychopathology improvement over 45 years of 12.3 points for placebo, while the increase was of merely 1.2 points for antipsychotic agents[188]. Similarly concerning increases in placebo response in regulatory schizophrenia trials have been reported by the FDA, indicating that dropout rates also increased in parallel, with greater dropout rates in US-based studies[190].

Having a large placebo response fatally reduces the chances of finding significant differences with the experimental arm. In pharmacological clinical trials of depression, it has been shown that critical placebo response rates are 30% and 40% for monotherapy and augmentation, respectively[191]. Above these thresholds, chances of positive trials dramatically worsen[191].

Trial design, treatment, population and study conduct characteristics that are associated with placebo effects have been extensively studied, and several variables have been identified as being consistently associated with increased drug-placebo difference across different mental disorders. These factors should be considered carefully when designing trials aiming to increase the likelihood of success, i.e., separation from placebo. For example, an open-label lead-in phase before double-blind randomization increases placebo effect[38]. A second factor is poor recruitment with invalid baseline assessment and caseness ascertainment. On the other hand, more severe symptoms at baseline are associated with lower placebo response and greater drug-placebo difference in trials testing antidepressants for depressive disorders[192] as well as in schizophrenia trials, independent of year of the study[32]. However, when aiming for adequately high baseline symptom severity, one needs to consider artificial baseline symptom severity inflation due to wash-out or rebound phenomena, or to rater bias aiming to include patients above a certain minimum illness severity[189,193,194].

Greater improvement versus placebo in acutely exacerbated and more severe cases may be achieved more quickly, allowing for shorter trials to separate from placebo[195,196]. On the other hand, separation from placebo regarding negative symptoms, remission of symptoms or functional recovery may require longer trial designs. Therefore, the targeted outcome needs to be taken into consideration when setting symptom severity and trial duration parameters for trials.

Since some factors that increase the placebo response may also increase response to the experimental arm, ultimately having no net effect on the chances of a trial success, or may even increase drug response to a greater degree, it is most important to assess factors from the viewpoint of decreasing or increasing the drug-placebo difference. The largest evidence synthesis to date has shown that factors moderating larger drug-placebo differences in schizophrenia trials were smaller sample size, less study sites, less active study arms, more patients randomized to placebo, use of the Brief Psychiatric Rating Scale (BPRS) instead of the later introduced PANSS, longer wash-out period, longer study duration, shorter duration of illness, and younger age[188,197]. In multivariable meta-regression analyses, the only remaining predictors of greater drug-placebo difference included lower placebo response and non-industry sponsorship, which is associated with a lower likelihood of having trial design features that have been associated with greater placebo effects[197]. The fact that placebo response is inflated when randomizing more patients to the active arm and less to the placebo arm, as shown in depression[198] and schizophrenia[193], is probably due to expectations of improvement[172].

Population, recruitment

The results of every clinical trial apply to the population represented by the sample, not beyond. For instance, the results of an RCT conducted in patients with early-stage Alzheimer's disease do not necessarily apply to the prevention of that disease in at-risk individuals or those with minimal cognitive impairment, or to those at middle or late stages of the disease. For ethical reasons, one cannot include those unwilling to consent to participate, or patients who are likely to be harmed by participation. Otherwise, to which population the RCT researchers intend their conclusions to apply determines inclusion/exclusion criteria, clearly stated and consistently applied.

Moreover, the results of any RCT do not necessarily apply to every subgroup of the population sampled. If a treatment is shown highly effective in the population sampled, there may yet be a minority subgroup in which the treatment is ineffective or toxic. If an RCT detects little or no treatment versus control difference, the population may split into two subgroups, in one of which treatment is more effective and safe, while in the other control is more effective and safe, cancelling each other in the total population[200].

Patients included in trials for schizophrenia are usually not representative of the real-world population seen in everyday clinical practice. Moreover, trial and population characteristics have changed over time[188]. For instance, patients with schizophrenia that are typically eligible in trials have less physical comorbidities, less psychiatric comorbidities, and less suicidal behaviors[156]. Overall, only one patient out of five real-world patients with schizophrenia would be eligible to be recruited in a randomized controlled trial[156].

Such limited representativeness of phase 2 and 3, placebo-controlled trials in the field of schizophrenia applies also to other conditions, including mood disorders[201] and substance use disorders, due to similarly restricted inclusion criteria and also to the fact that patients need to be capable of giving informed consent. This limited representativeness puts emphasis on the importance of well-designed phase 4 studies that aim to test not if, but in whom and under which circumstances a medication

works. It would be helpful if certain regulatory minimal standards and requirements for phase 4 studies could be attached to approval of a new medication. While current post-approval requirements are generally restricted to additional indications (e.g., relapse prevention trials, pediatric trials) or safety assessments/risk mitigation measures, it would be desirable and welcome if a set of standards for phase 4 trials aiming at testing generalizability or utility in certain patient subgroups could be developed and applied.

Another relevant problem is inflation of symptoms at baseline. This can derive from several factors. First, symptoms do vary through the natural course of a disease, and can be reactive to stressful stimuli, such as routine disruption or anticipation of novel scenarios. Participating in a clinical trial can certainly come with stress, and so at the baseline assessment a person might show inflated symptoms, that can then regress to the mean once the trial environment and visits have become the new "normal". Another explanation can be the need of sites to recruit patients, that can produce, even not deliberately, higher symptoms ratings at baseline.

Several strategies can be implemented to optimize patient representativeness, and reduce symptom inflation at baseline. First, to reduce the risk of including "professional" trial participants, chronically unstable instead of acutely exacerbated patients, or those with unclear diagnosis and treatment history, it may be advisable to require medical records documenting at least the recent past in those not recruited from regular clinical care settings. Second, relaxing to some degree inclusion criteria, without increasing risk to study participants or the integrity of the study, by allowing participants with a certain set of physical or psychiatric comorbidities, would make recruitment easier, and the trial more pragmatic and clinically useful, potentially decrease placebo response, and allow greater adherence to equity, diversity and inclusion principles[202-205].

Retention is also part of recruitment, i.e., the continual "recruitment" of patients into staying in the study. Retention is crucial to minimize loss of data, that may actually be missing not at random, and to retain sufficient statistical power needed to test the hypothesis. Of note, exit strategies and lined trial phases may affect retention vs. dropout from the trial. For example, if exit strategies are too lenient or have too much appeal (e.g., open extension study with free treatment), more patients than necessary may drop out. If, on the other hand, exit strategies are too strict, patients may be kept in the study longer than they should. Thus, it is important to balance the desire for low dropout with need for patient safety by permitting more rescue strategies within the study that are transient and/or do not compromise the outcome. However, one may want to limit rewarding dropout and roll-over options into next/additional study phases.

Sites

Trials are typically conducted across multiple sites, to allow timely recruitment of sufficiently large samples. However, having a high number of sites does not come without downfalls. First, sites are frequently incentivized to recruit, and have pressure to recruit, which can lead to inclusion of inappropriate patients with regards to diagnosis, duration of exacerbation, or baseline severity. The more sites participate in a trial, the higher the heterogeneity, the higher the chance of poor quality of trial procedure compliance, including randomization, blinding and ratings, and the harder the quality control.

Dropping sites with poor recruitment early, as well those sites showing abnormal placebo response, can mitigate the impact of this heterogeneity. Second, sites should be certified, re-certified, and strictly monitored, with rater retraining being offered or raters being dropped in case of signs of inconsistent ratings. Third, since the number of sites moderates larger placebo response, having fewer highly efficient and high-quality sites as opposed to many poorly efficient sites is preferable. Moreover, in situations where multiple trials with multiple molecules are being conducted at similar times, competition over eligible patients can be a problem. In such situations, it is possible that patients required for trials with more restrictive criteria regarding illness duration or severity, comorbidities or comedications are steered preferentially toward those trials, so that some of such patients are removed from the other trials.

Lacking objective "laboratory" tests and biomarkers, we rely on the participant's subjective report, and on the training of assessors as well as their reliability with other assessors in the same trial. Given the number of sites often involved in such trials, how realistic is it to expect true inter-rater reliability to be established and maintained? Yet, inter-rater reliability contributes to statistical power.

Reliability training is almost always performed only on the ratings of interviews conducted by an expert with a model patient, thereby creating an ideal situation that allows for time-efficient rater training. The skill to elicit the information that is to be rated is left out, which can create serious issues with the actual elicitation of valid data. Thus, raters should also be trained and assessed in the elicitation, not only the rating procedures. Furthermore, as there can be rater drift over time, trainings need to be repeated throughout often long trial programmes.

Centralized raters were introduced with the goal of addressing these issues, by utilizing live, two-way videos to vastly reduce the number of required raters and enable ongoing calibration of reliability[206,207]. In addition, providing such external assessment and adjudication of patient eligibility is intended to help reduce misaligned incentives in determining patient eligibility and the phenomenon of baseline inflation[208]. Although such methods can provide advantages, there are limitations as well, including the lack of information gathered in a direct encounter.

The introduction of new technologies holds enormous promise for making such processes more reliable, continuous, applicable in the real world, and cost-effective. For example, language processing and speech analysis[209,210] and analyses of facial expression[211] could be very informative in conditions such as schizophrenia, mania and depression, or even in such domains as agitation and negative symptoms. At the same time, ecological momentary assessment can provide repeated sampling of subjects' current behaviors and experi-

ences in real time, in their natural environments[212,213]. Such a strategy can minimize recall bias and maximize ecological validity. The use of smartphones and wearable devices can provide objective information on geolocation, activity levels, frequency and timing of social interactions, sleep and other measures of interest to clinical trialists[214], including medication assumption[215,216].

The integration of digital phenotyping, as well as symptom efficacy and tolerability surveillance using passively collected data, have been underexploited in both the selection of adequate patients as well as the ongoing assessment of outcomes throughout clinical trials and drug discovery and development in psychiatry. These modern technologies provide unprecedented opportunities and need to be explored as supportive, key secondary, or even primary outcomes for regulatory approval trial programmes. Moreover, as patient-reported outcomes as well as functional endpoints gain traction, digital assessments are going to provide more continuous, reliable and real-world data that can be used to assess the value of a new treatment versus the appropriate control condition.

Assessment and outcomes

Raters should administer scales and measures that are clinically relevant, that are meaningful for the patient, that are not too time consuming, and that are broadly used in the field (also to allow evidence synthesis efforts). Special attention should be given to the time of the assessment, in particular – but not only – with cognitive symptoms, due to diurnal variation of the performance[217].

Assessment should be ideally repeated over time, to feed analyses with richer data. For example, to compare treatment vs. control on change in severity over eight weeks, one could measure only the endpoint, or the change in severity between baseline and the endpoint, or the slope of severity over the eight weeks, or one could dichotomize any of these possibilities, which would all be valid choices. Using the endpoint or pre-post change is generally not the best choice, as, with dropout, the endpoint is the time point most likely to be missing. Instead, the slope (say, over weeks 0, 1, 4, 8) is a better choice, since this is a linear combination of the repeated severity measures, which increases the reliability of the outcome measure (hence power). The availability of repeated measures over time also improves imputation, better protecting validity. However, requiring measures, say, daily over eight weeks, rather than only at four time points, may erase such advantages by encouraging dropout and missing data. A balance between the burden on patients and the needs of the research must always be considered and tailored to the research question at hand.

More than one outcome in a trial is desirable, as one outcome only can hardly provide a comprehensive clinical picture, yet adjusting for multiple comparisons in the statistical analyses is needed in case that more than one primary outcome is being assessed or in case that inferential statistical testing is desired even of key secondary outcomes. For secondary and exploratory, hypothesis-generating outcomes and those requiring a lot of multidimensional data, such as for functioning, modern tools including digital phenotyping and ecological momentary assessment can be of great value and should be progressively introduced in assessment of trials[218-228]. Digital phenotyping and ecological momentary assessments can be repeated multiple times, and can be even continuous in case of passive monitoring. To what degree interactive digital phenotyping may affect placebo response is still unclear, and whether a digital outcome parameter could become a primary outcome leading to approval of a medicine will need to be seen, but is not beyond the realms of feasibility and validity. Additionally, monitoring of physiologic parameters is a potential candidate tool to facilitate measurement of objective response, biomarkers of subgroups with better response, or target engagement.

Beyond secondary and exploratory outcomes that can be manifold but should be assessed with minimal patient time and burden, the most salient problem, however, is multiplicity for the primary outcome measures in an RCT. The goal of an RCT is to recommend *one* treatment over the other in the population sampled: *one* decision. Having multiple primary outcome measures that give conflicting answers undermines the purpose of the RCT. With one primary outcome, the chance of a false positive with usual approaches is less than 5%. With two independent primary outcomes, the chance of one or more false positives is 10%; with three it is 14%, ever increasing the chance of a misleading conclusion. If there is adjustment for multiple testing, using a significance level lower enough for each outcome, so that the overall chance of a false positive result is less than 5%, there is a loss of power, a greater risk of a failed RCT, and still, conflicting results on the multiple tests.

An RCT should have *one and only one* primary outcome measure, but that may be a composite measure. Ideally, with that measure presented for two patients in the population, clinicians should be able to unequivocally recognize which (if either) had the better clinical outcome. For example, the decrease of symptoms over treatment might be an acceptable outcome measure. However, if patients develop serious health problems due to treatment or control, that is not a sufficient primary outcome measure. Ideally, the appropriate outcome measure should reflect a benefit-to-harm balance. If there are several independent benefits and several independent harms of concern, the outcome of treatment is the cumulative effect on the patient of whatever the benefits and harms experienced[229]. Benefits and harms ideally should somehow be considered jointly, with the effect of treatment indicated by the total effect on the patient, not the separate effects on multiple outcome measures[230]. By the same token, if symptom severity is measured weekly over, say, eight weeks of treatment, the impact of treatment should not be separately assessed at each week, but some composite measure (e.g., the trend of the severity over time) should be used.

Finally, dichotomization of an ordinal outcome is always a poor choice. For example, if "success" were defined by a ≥50% decrease in symptoms over the eight weeks, a patient with a 51% decrease in symptoms has the identical outcome to another with a 100% decrease, while a patient with a 49% decrease is considered the same as one with 0% decrease or an increase. Moreover,

two patients, one with 49% and one with 51% decrease, are considered as different from each other as one with 0% and another with 100% decrease. Consequently, there is a significant risk for misclassification and a major loss of power with dichotomization[231]; sample sizes may have to be doubled or tripled to have the same power as that from using the ordinal or continuous outcome. To make matters worse, different choices of cut-point may change the conclusions. The "costs of dichotomization" have long been recognized[232], but are often ignored. However, it is possible to turn a dichotomized outcome, such as response or relapse, into a scaled outcome, by estimating the time to an event. Although this approach increases the statistical power, nevertheless, the decision about the specific definition and cut-points involved in the definition of the categorical outcome remain.

Statistical analyses

The success of a trial, and approval of a medication to treat a given disease, also largely depend on the results of the statistical analyses. These analyses, if wrong, even in presence of a sound design, can jeopardize a large amount of work and investments. Hence, adopting appropriate statistical approaches that minimize type I and II error chances is paramount.

One of the aspects in statistical analyses is how they are adjusted for multiple testing. One commonly used method is the most conservative Bonferroni correction, that divides the alpha=0.05 by the number of statistical tests. However, a number of related and different methods exist that should be considered[233]. Such methods also include hierarchical testing in case multiple secondary outcomes are subjected to inferential statistics, whereby outcomes are ordered based on importance or likelihood of success and then each tested at $p<0.05$, stopping all further testing once the next *a priori* selected outcome does not reach that statistical threshold.

Another important aspect in statistical analyses is how covariates are handled. Baseline factors that identify subgroups in which treatment effects are different are "moderators of treatment outcome" in that population[234]. What the results of an RCT demonstrate is what would happen if everyone in the population sampled were given treatment rather than control. If there are moderators known *a priori*, that affects sampling decisions. For example, if it is already known from previous research that a treatment is effective only for women and not for men, further research on that treatment would focus on women. If there is only suggestive evidence that sex might moderate treatment outcome, the RCT might be stratified by sex, with adequate representation of males and females, to test the *a priori* hypothesis that sex moderates treatment outcome and to estimate separate effect sizes for women and for men.

Some researchers would throw sex in as a covariate in a linear model "just in case". If sex is irrelevant to the outcome, the treatment effect tested and estimated is exactly the same one as when the covariate is not included, but with a loss of power and precision. Conversely, if sex moderates treatment outcome, and the interaction term is omitted (as it often is), the effect size tested and estimated is uninterpretable. Only if it is known *a priori* that the treatment vs. control effect is the same for males and females, is the treatment effect size meaningful, representing the common effect size for males and females in that population.

The situation worsens when there are multiple covariates entered into a linear model "just in case", that are correlated with each other (collinear), and the interactions of each covariate with the treatment or with each other are incorrectly assumed to be zero, or it is incorrectly assumed that each has a linear effect on the outcome. If any of these assumptions is wrong, the RCT validity and power will be compromised. Yet, many published RCTs enter multiple covariates into their models without a rationale or justification, under a misapprehension that "controlling for" factors by adding in covariates "just in case" improves RCT results. Instead, each covariate to be used in a RCT analysis should be explicitly mentioned in the *a priori* hypothesis and registration, and the rationale and justification for each should be presented in both the proposal and the resulting paper. How covariates are to be included must be specified and justified in the analysis plan, and the sample size increased to accommodate the consequent loss of power.

Another important aspect of statistical analyses is imputation. Imputation is needed to conduct intention-to-treat or modified intent-to-treat analyses where patients are included who have treatment exposure and at least one post-baseline assessment. Intention-to-treat analyses are more representative of the overall efficacy/acceptability ratio of an experimental treatment, as opposed to "completer" analyses that are conducted on selected "ideal" patients who likely benefitted the most from that medication. In fact, completer analyses violate the randomization principle and are to be avoided.

Various imputation methods exist to handle missing data. The simplest method is last-observation-carried-forward. However, this method assumes no further change after dropout and disadvantages the group in which there is earlier and more discontinuation in terms of efficacy, but also reduces the time for cumulative adverse effects in that study arm. A now frequently used alternative is the mixed model for repeated measures (MMRM), a popular choice for randomized trials with longitudinal continuous outcomes. In MMRM analyses, the results from patients staying in the study longer are used to model the estimated change after study discontinuation based on trajectories of patients with similar initial symptom change. However, as patients completing trials on placebo may be systematically different from those who do not, especially if they drop out for inefficacy, MMRM models may overestimate placebo effects, which may be another reason for increasing placebo effects in more recent years, when MMRM analyses have become the standard data method in RCTs.

Another potentially important issue is whether the assumption that data are missing at random, which underlie all standard data analytic techniques, is true. Given that efficacy and tolerability differences between study arms may significantly affect missingness of data, especially in longer-term studies with higher dropout rates, non-random missingness can significantly affect the results. Thus, it is important to check if data are in fact missing at ran-

dom and to employ different data analytic techniques if this assumption is violated, such as selection models or pattern mixture models[235-237], which is rarely done, but which can affect the results and interpretation of the study.

DISCUSSION

Clinical trials are the cornerstone of current evidence-based medicine. The field has evolved, and increasingly complex as well as simplified clinical trial designs have been developed. Designs range from effectiveness trials with maximized internal validity but limited external generalizability, to large simple trials that maximize external validity but have reduced precision. In the case of non-randomized trials, large nationwide database studies can aid hypothesis generation, but are insufficient to allow making causal inferences. Data analytics have equally evolved and are now very sophisticated, and it has become increasingly important to choose the most appropriate statistical analysis plan for a given trial design, research question and attempt at minimizing type I and/or type II error.

In drug development and for regulatory approval purposes, randomized, placebo-controlled, parallel-design trials are the main vehicle. They include placebo-controlled trials for the approval of acute treatments as well as placebo substitution trials for the approval of maintenance interventions. Increasingly, an active control (not comparison) arm is included in order to test the integrity of the study, which enables to distinguish between negative trials (the established medication does separate from placebo, while the experimental drug does not) from failed trials (neither the experimental nor the established medication separate from placebo). Moreover, comparison with an established "common comparator", either as part of the placebo-controlled phase 3 trial programme or of phase 4 studies, will gain traction to go beyond common symptom and adverse effect outcomes to include also quality of life and/or functional endpoints, on which the new medication can demonstrate statistically and clinically relevant advantages. Indeed, patient-reported subjective well-being and quality of life, caregiver/observer reports and functional outcomes, which may be captured more objectively and comprehensively in the living world environment via digital assessments, have become increasingly relevant.

However, in mental health, novel psychopharmacological mechanisms of action that effectively and safely treat common and often severely impairing mental disorders have remained extremely scarce, and many initially promising trial programmes ultimately failed. Clinical trials in psychiatric disorders have been challenged by issues around recruitment of a sufficiently large and representative sample of patients, within a reasonable amount of time, fulfilling strict inclusion criteria to answer a given question. However, sample sizes have increased, especially in phase 3 trials, due to a disproportionate increase in placebo response with relatively little increase in drug response over the past few decades.

When targeting outcomes beyond symptoms, including quality of life and functionality in multiple relevant domains – self-care, social interactions, leisure time activities, and educational/work performance – medications mostly "only" prepare the brains of people with mental disorders to have the potential to function better, without putting their increased or restituted "capacity" into action. In order to translate the improved symptomatic status into action and also improve measurable "performance", designs that combine drugs with psychosocial interventions may need to be considered more, especially when targeting complex cognitive, behavioral and functional outcomes. As a matter of fact, when seeking approval for the pharmacological treatment of cognition in schizophrenia, a functional co-primary outcome is required demonstrating that the statistically significantly improvement in cognitive performance has real-world impact on behavior and functioning.

The rapid evolution of widely available and scalable digital technology holds enormous promise to enhance the precision and granularity as well as the temporal coverage of the assessment of symptoms and behavior in people before and during treatment with a tested pharmacological entity or its control. Such digital phenotyping can be helpful to measure symptoms more comprehensively and with more precision and ecological validity, including their variability over time and in relationship to internal and external contexts. Moreover, digital tools can provide more reliably and objectively assessments of cognitive, academic, behavioral and social functioning. Inasmuch as passive instead of interactive digital monitoring in applied, concerns about increased placebo effects via digital engagement should be mitigated.

The overview of ongoing phase 2 and 3 trials that we present in this paper has some limitations. First, although we attempted to be inclusive in the identification of pharmacological agents with novel mechanisms of action, or already known agents targeting a currently unapproved mental condition, we may have missed some agents. The exclusion of eligible agents may have been due to our restricting the search to the US and European clinical trials registers, so that agents and trial programmes not registered yet may have been missed. Moreover, there may be trial programmes and agents in other than the US and European trial registries that we did not survey. Additionally, some agents that might have been approved for another condition or age group may have been classified as phase 4 trials and missed. Furthermore, as the field of psychopharmacology is a highly dynamic and evolving one, new agents and targets may have been identified since our last search date. Second, we may have listed drugs and targets that have since been dropped and trial programmes that have been discontinued. However, as clinical trial registries are updated on a voluntary basis, this information may have been actually not available. On-time updating of the records by sponsors would be desirable. Third, although we attempted to classify the mechanisms of action of emerging and newly tested psychopharmacological agents, for some of them insufficient information was available, so that they may not have been classifiable or may even be (partially) incorrectly classified. Hence, as further information about the specific mechanisms of action of individual pharmacological treatments emerge, our classifications may need to be updated or corrected.

In conclusion, the development and approval process for new pharmacological

agents that target medical conditions is complex, and this complexity and the related perils of failure may be even enhanced when targeting mental disorders. The information contained in this paper aims to provide practical knowledge on issues related to clinical trial methodology and implementation that need to be considered and weighed, with their relative pros and cons, serving as a roadmap that targets successful approval of new agents for the treatment of mental disorders.

Additionally, in taking stock of the current drug development targets and related mechanisms of action aimed at the treatment of the main mental disorders in adults, we aimed to provide an overview of the most promising molecules that the field should observe, learn from and, possibly, pursue further, should specific agents under development successfully progress through their phase 2 and 3 programs and, ultimately, lead to regulatory approval.

It is hoped that, in ten years from now, multiple new drug targets will become available, ideally for each of the reviewed main mental disorders, allowing clinicians to improve outcomes of many patients who are currently still only sub-optimally managed with the currently available agents, so that not only impact on symptoms and tolerability are increased, but also subjective well-being, quality of life and social functioning can be improved more and in sustainable ways.

ACKNOWLEDGEMENTS

C.U. Correll, M. Solmi and S. Cortese contributed equally to this work. Supplementary information on the study is available at https://osf.io/ys9pr/?view_only=ed9fae2fffc44daeafff5f56a5f3e1ff.

REFERENCES

1. Dragioti E, Radua J, Solmi M et al. Global population attributable fraction of potentially modifiable risk factors for mental disorders: a meta-umbrella systematic review. Mol Psychiatry 2022; doi: 10.1038/s41380-022-01586-8.
2. Fusar-Poli P, Correll C, Arango C et al. Preventive psychiatry: a blueprint for improving the mental health of young people. World Psychiatry 2021;20:200-21.
3. Salazar de Pablo G, De Micheli A, Solmi M et al. Universal and selective interventions to prevent poor mental health outcomes in young people: systematic review and meta-analysis. Harv Rev Psychiatry 2021;29:196-215.
4. Solmi M, Dragioti E, Arango C et al. Risk and protective factors for mental disorders with onset in childhood/adolescence: an umbrella review of published meta-analyses of observational longitudinal studies. Neurosci Biobehav Rev 2021;120:565-73.
5. Maj M, Stein DJ, Parker G et al. The clinical characterization of the adult patient with depression aimed at personalization of management. World Psychiatry 2020;19:269-93.
6. Maj M, van Os J, De Hert M et al. The clinical characterization of the patient with primary psychosis aimed at personalization of management. World Psychiatry 2021;20:4-33.
7. McIntyre R, Alda M, Baldessarini R et al. The clinical characterization of the adult patient with bipolar disorder aimed at personalization of management. World Psychiatry 2022;21:364-87.
8. Kotov R, Jonas KG, Carpenter WT et al. Validity and utility of Hierarchical Taxonomy of Psychopathology (HiTOP): I. Psychosis superspectrum. World Psychiatry 2020;19:151-72.
9. Krueger RF, Hobbs KA, Conway CC et al. Validity and utility of Hierarchical Taxonomy of Psychopathology (HiTOP): II. Externalizing superspectrum. World Psychiatry 2021;20:171-93.
10. Watson D, Levin-Aspenson HF, Waszczuk MA et al. Validity and utility of Hierarchical Taxonomy of Psychopathology (HiTOP): III. Emotional dysfunction superspectrum. World Psychiatry 2022;21:26-54.
11. Zohar J, Allgulander C. Antipsychotics in anxiety disorders: an oxymoron or a reflection of nonadequate nomenclature? Eur Neuropsychopharmacol 2011;21:427-8.
12. Huhn M, Tardy M, Spineli LM et al. Efficacy of pharmacotherapy and psychotherapy for adult psychiatric disorders: a systematic overview of meta-analyses. JAMA Psychiatry 2014;71:706-15.
13. Correll CU, Cortese S, Croatto G et al. Efficacy and acceptability of pharmacological, psychosocial, and brain stimulation interventions in children and adolescents with mental disorders: an umbrella review. World Psychiatry 2021;20:244-75.
14. John M. Eisenberg Center for Clinical Decisions and Communications Science. Off-label use of atypical antipsychotics: an update. In: Comparative effectiveness review summary guides for clinicians. Rockville: US Agency for Healthcare Research and Quality, 2007.
15. Williams T, Stein DJ, Ipser J. A systematic review of network meta-analyses for pharmacological treatment of common mental disorders. Evid Based Ment Health 2018;21:7-11.
16. Locher C, Koechlin H, Zion SR et al. Efficacy and safety of SSRIs, SNRIs, and placebo in common psychiatric disorders: a comprehensive meta-analysis in children and adolescents. JAMA Psychiatry 2017;74:1011-20.
17. Jensen NH, Rodriguiz RM, Caron MG et al. N-desalkylquetiapine, a potent norepinephrine reuptake inhibitor and partial 5-HT1A agonist, as a putative mediator of quetiapine's antidepressant activity. Neuropsychopharmacology 2008;33:2303-12.
18. Boaden K, Tomlinson A, Cortese S et al. Antidepressants in children and adolescents: meta-review of efficacy, tolerability and suicidality in acute treatment. Front Psychiatry 2020;11:717.
19. Kaminski JA, Bschor T. Antidepressants and suicidality: a re-analysis of the re-analysis. J Affect Disord 2020;266:95-9.
20. Zohar J, Levy DM. Neuroscience-based nomenclature of psychotropics: progress report. Eur Neuropsychopharmacol 2022;57:36-8.
21. Zohar J, Stahl S, Moller HJ et al. A review of the current nomenclature for psychotropic agents and an introduction to the Neuroscience-based Nomenclature. Eur Neuropsychopharmacol 2015;25:2318-25.
22. Zohar J, Kasper S. Neuroscience-based Nomenclature (NbN): a call for action. World J Biol Psychiatry 2016;17:318-20.
23. Sultan RS, Correll CU, Zohar J et al. What's in a name? Moving to neuroscience-based nomenclature in pediatric psychopharmacology. J Am Acad Child Adolesc Psychiatry 2018;57:719-21.
24. Downing AM, Kinon BJ, Millen BA et al. A double-blind, placebo-controlled comparator study of LY2140023 monohydrate in patients with schizophrenia. BMC Psychiatry 2014;14:351.
25. Stauffer VL, Millen BA, Andersen S et al. Pomaglumetad methionil: no significant difference as an adjunctive treatment for patients with prominent negative symptoms of schizophrenia compared to placebo. Schizophr Res 2013;150:434-41.
26. Recio-Barbero M, Segarra R, Zabala A et al. Cognitive enhancers in schizophrenia: a systematic review and meta-analysis of alpha-7 nicotinic acetylcholine receptor agonists for cognitive deficits and negative symptoms. Front Psychiatry 2021;12:631589.
27. Lewis AS, van Schalkwyk GI, Bloch MH. Alpha-7 nicotinic agonists for cognitive deficits in neuropsychiatric disorders: a translational meta-analysis of rodent and human studies. Prog Neuropsychopharmacol Biol Psychiatry 2017;75:45-53.
28. Bugarski-Kirola D, Blaettler T, Arango C et al. Bitopertin in negative symptoms of schizophrenia – results from the phase III FlashLyte and DayLyte studies. Biol Psychiatry 2017;82:8-16.
29. Hirayasu Y, Sato S-I, Shuto N et al. Efficacy and safety of bitopertin in patients with schizophrenia and predominant negative symptoms: subgroup analysis of Japanese patients from the global randomized phase 2 trial. Psychiatry Investig 2017;14:63-73.
30. Bugarski-Kirola D, Iwata N, Sameljak S et al. Efficacy and safety of adjunctive bitopertin versus placebo in patients with suboptimally controlled symptoms of schizophrenia treated with antipsychotics: results from three phase 3, randomised, double-blind, parallel-group, placebo-controlled, studies in the SearchLyte clinical trial programme. Lancet Psychiatry 2016;3:1115-28.
31. Verma A, Kumar Waiker D, Bhardwaj B et al. The molecular mechanism, targets, and novel molecules in the treatment of Alzheimer's disease. Bioorg Chem 2022;119:105562.
32. Gopalakrishnan M, Zhu H, Farchione TR et al. The trend of increasing placebo response and decreasing treatment effect in schizophrenia trials continues: an update from the US Food and Drug Administration. J Clin Psychiatry 2020;81:19r12960.
33. Parker G, Ricciardi T, Hadzi-Pavlovic D. Placebo response rates in trials of antidepressant drugs in adults with clinical depression: increasing, decreasing, constant or all of the above? J Affect Disord 2020;271:139-44.
34. Siafis S, Çıray O, Schneider-Thoma J et al. Placebo response in pharmacological and dietary supplement trials of autism spectrum disorder (ASD): systematic review and meta-regression analysis.

Mol Autism 2020;11:66.
35. Faraone SV, Newcorn JH, Cipriani A et al. Placebo and nocebo responses in randomised, controlled trials of medications for ADHD: a systematic review and meta-analysis. Mol Psychiatry 2022; 27:212-9.
36. Scherrer B, Guiraud J, Addolorato G et al. Baseline severity and the prediction of placebo response in clinical trials for alcohol dependence: a meta-regression analysis to develop an enrichment strategy. Alcohol Clin Exp Res 2021;45:1722-34.
37. Nasir M, Li F, Courley S et al. Meta-analysis: pediatric placebo response in depression trials does not replicate in anxiety and obsessive-compulsive disorder trials. J Child Adolesc Psychopharmacol 2021;31:670-84.
38. Jones BDM, Razza LB, Weissman CR et al. Magnitude of the placebo response across treatment modalities used for treatment-resistant depression in adults: a systematic review and meta-analysis. JAMA Netw Open 2021;4:e2125531.
39. Cao B, Liu YS, Selvitella A et al. Differential power of placebo across major psychiatric disorders: a preliminary meta-analysis and machine learning study. Sci Rep 2021;11:21301.
40. Ahmadzad-Asl M, Davoudi F, Mohamadi S et al. Systematic review and meta-analysis of the placebo effect in panic disorder: implications for research and clinical practice. Aust N Z J Psychiatry 2022;56:1130-41.
41. Cipriani A, Furukawa TA, Salanti G et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. Lancet 2018; 391:1357-66.
42. Xue T, Wu X, Chen S et al. The efficacy and safety of dual orexin receptor antagonists in primary insomnia: a systematic review and network meta-analysis. Sleep Med Rev 2022;61:101573.
43. Zheng W, Cai D-B, Zheng W et al. Brexanolone for postpartum depression: a meta-analysis of randomized controlled studies. Psychiatry Res 2019;279:83-9.
44. Walkery A, Leader LD, Cooke E et al. Review of allopregnanolone agonist therapy for the treatment of depressive disorders. Drug Des Devel Ther 2021;15:3017-26.
45. Correll CU, Newcomer JW, Silverman B et al. Effects of olanzapine combined with samidorphan on weight gain in schizophrenia: a 24-week phase 3 study. Am J Psychiatry 2020;177:1168-78.
46. Martin WF, Correll CU, Weiden PJ et al. Mitigation of olanzapine-induced weight gain with samidorphan, an opioid antagonist: a randomized double-blind phase 2 study in patients with schizophrenia. Am J Psychiatry 2019;176:457-67.
47. Jawad MY, Di Vincenzo JD, Ceban F et al. The efficacy and safety of adjunctive intranasal esketamine treatment in major depressive disorder: a systematic review and meta-analysis. Expert Opin Drug Saf 2022;21:841-52.
48. Bahji A, Zarate CA, Vazquez GH. Efficacy and safety of racemic ketamine and esketamine for depression: a systematic review and meta-analysis. Expert Opin Drug Saf 2022;21:853-66.
49. Solmi M, Pigato G, Kane JM et al. Treatment of tardive dyskinesia with VMAT-2 inhibitors: a systematic review and meta-analysis of randomized controlled trials. Drug Des Devel Ther 2018;12:1215-38.
50. Mansuri Z, Reddy A, Vadukapuram R et al. Pimavanserin in the treatment of Parkinson's disease psychosis: meta-analysis and meta-regression of randomized clinical trials. Innov Clin Neurosci 2022;19:46-51.
51. Strickland JC, Johnson MW. Human behavioral pharmacology of psychedelics. Adv Pharmacol 2022;93:105-32.
52. Andersen KAA, Carhart-Harris R, Nutt DJ et al. Therapeutic effects of classic serotonergic psychedelics: a systematic review of modern-era clinical studies. Acta Psychiatr Scand 2021;143: 101-18.
53. Rucker JJH, Iliff J, Nutt DJ. Psychiatry & the psychedelic drugs. Past, present & future. Neuropharmacology 2018;142:200-18.
54. Luoma JB, Chwyl C, Bathje GJ et al. A meta-analysis of placebo-controlled trials of psychedelic-assisted therapy. J Psychoactive Drugs 2020;52:289-99.
55. Romeo B, Karila L, Martelli C et al. Efficacy of psychedelic treatments on depressive symptoms: a meta-analysis. J Psychopharmacol 2020;34:1079-85.
56. Brannan SK, Sawchak S, Miller AC et al. Muscarinic cholinergic receptor agonist and peripheral antagonist for schizophrenia. N Engl J Med 2021; 384:717-26.
57. Targum SD, Murphy C, Breier A et al. Site-independent confirmation of primary site-based PANSS ratings in a schizophrenia trial. J Psychiatr Res 2021;144:241-6.
58. Koblan KS, Kent J, Hopkins SC et al. A non-D2-receptor-binding drug for the treatment of schizophrenia. N Engl J Med 2020;382:1497-506.
59. Cantillon M, Prakash A, Alexander A et al. Dopamine serotonin stabilizer RP5063: a randomized, double-blind, placebo-controlled multicenter trial of safety and efficacy in exacerbation of schizophrenia or schizoaffective disorder. Schizophr Res 2017;189:126-33.
60. Usall J, Huerta-Ramos E, Labad J et al. Raloxifene as an adjunctive treatment for postmenopausal women with schizophrenia: a 24-week double-blind, randomized, parallel, placebo-controlled trial. Schizophr Bull 2016;42:309-17.
61. Weiser M, Levi L, Burshtein S et al. Raloxifene plus antipsychotics versus placebo plus antipsychotics in severely ill decompensated postmenopausal women with schizophrenia or schizoaffective disorder: a randomized controlled trial. J Clin Psychiatry 2017;78:e758-65.
62. Modabbernia A, Heidari P, Soleimani R et al. Melatonin for prevention of metabolic side-effects of olanzapine in patients with first-episode schizophrenia: randomized double-blind placebo-controlled study. J Psychiatr Res 2014;53:133-40.
63. McGuire P, Robson P, Cubala WJ et al. Cannabidiol (CBD) as an adjunctive therapy in schizophrenia: a multicenter randomized controlled trial. Am J Psychiatry 2018;175:225-31.
64. Weiser M, Levi L, Zamora D et al. Effect of adjunctive estradiol on schizophrenia among women of childbearing age: a randomized clinical trial. JAMA Psychiatry 2019;76:1009-17.
65. Bugarski-Kirola D, Arango C, Fava M et al. Pimavanserin for negative symptoms of schizophrenia: results from the ADVANCE phase 2 randomised, placebo-controlled trial in North America and Europe. Lancet Psychiatry 2022;9:46-58.
66. Davidson M, Saoud J, Staner C et al. Efficacy and safety of MIN-101: a 12-week randomized, double-blind, placebo-controlled trial of a new drug in development for the treatment of negative symptoms in schizophrenia. Am J Psychiatry 2017;174: 1195-202.
67. Davidson M, Saoud J, Staner C et al. Efficacy and safety of roluperidone for the treatment of negative symptoms of schizophrenia. Schizophr Bull 2022;48:609-19.
68. Fleischhacker WW, Podhorna J, Gröschl M et al. Efficacy and safety of the novel glycine transporter inhibitor BI 425809 once daily in patients with schizophrenia: a double-blind, randomised, placebo-controlled phase 2 study. Lancet Psychiatry 2021;8:191-201.
69. Man WH, Colen-de Koning J, Schulte P et al. Clozapine-induced hypersalivation: the association between quantification, perceived burden and treatment satisfaction reported by patients. Ther Adv Psychopharmacol 2017;7:209-10.
70. Bauer IE, Green C, Colpo GD et al. A double-blind, randomized, placebo-controlled study of aspirin and N-acetylcysteine as adjunctive treatments for bipolar depression. J Clin Psychiatry 2018;80: 18m12200.
71. Loebel A, Koblan KS, Tsai J et al. A randomized, double-blind, placebo-controlled proof-of-concept trial to evaluate the efficacy and safety of non-racemic amisulpride (SEP-4199) for the treatment of bipolar I depression. J Affect Disord 2022;296:549-58.
72. Hopkins SC, Wilkinson S, Corriveau TJ et al. Discovery of nonracemic amisulpride to maximize benefit/risk of 5-HT7 and D2 receptor antagonism for the treatment of mood disorders. Clin Pharmacol Ther 2021;110:808-15.
73. Calabrese JR, Frye MA, Yang R et al. Efficacy and safety of adjunctive armodafinil in adults with major depressive episodes associated with bipolar I disorder: a randomized, double-blind, placebo-controlled, multicenter trial. J Clin Psychiatry 2014;75:1054-61.
74. Ketter TA, Yang R, Frye MA. Adjunctive armodafinil for major depressive episodes associated with bipolar I disorder. J Affect Disord 2015;181: 87-91.
75. Frye MA, Amchin J, Bauer M et al. Randomized, placebo-controlled, adjunctive study of armodafinil for bipolar I depression: implications of novel drug design and heterogeneity of concurrent bipolar maintenance treatments. Int J Bipolar Disord 2015;3:34.
76. Henter ID, Park LT, Zarate CA Jr. Novel glutamatergic modulators for the treatment of mood disorders: current status. CNS Drugs 2021;35:527-43.
77. McIntyre RS, Subramaniapillai M, Lee Y et al. Efficacy of adjunctive infliximab vs placebo in the treatment of adults with bipolar I/II depression: a randomized clinical trial. JAMA Psychiatry 2019; 76:783-90.
78. Lee Y, Mansur RB, Brietzke E et al. Efficacy of adjunctive infliximab vs. placebo in the treatment of anhedonia in bipolar I/II depression. Brain Behav Immun 2020;88:631-9.
79. Lee Y, Subramaniapillai M, Brietzke E et al. Anticytokine agents for anhedonia: targeting inflammation and the immune system to treat dimensional disturbances in depression. Ther Adv Psychopharmacol 2018;8:337-48.
80. Durgam S, Earley W, Guo H et al. Efficacy and safety of adjunctive cariprazine in inadequate responders to antidepressants: a randomized, double-blind, placebo-controlled study in adult patients with major depressive disorder. J Clin Psychiatry 2016;77:371-8.
81. Clayton AH, Tsai J, Mao Y et al. Effect of lurasidone

on sexual function in major depressive disorder patients with subthreshold hypomanic symptoms (mixed features): results from a placebo-controlled trial. J Clin Psychiatry 2018;79:18m12132.
82. Bakish D, Bose A, Gommoll C et al. Levomilnacipran ER 40 mg and 80 mg in patients with major depressive disorder: a phase III, randomized, double-blind, fixed-dose, placebo-controlled study. J Psychiatry Neurosci 2014;39:40-9.
83. Fava M, Dirks B, Freeman MP et al. A phase 2, randomized, double-blind, placebo-controlled study of adjunctive pimavanserin in patients with major depressive disorder and an inadequate response to therapy (CLARITY). J Clin Psychiatry 2019;80:19m12928.
84. Papakostas GI, Salloum NC, Hock RS et al. Efficacy of esketamine augmentation in major depressive disorder: a meta-analysis. J Clin Psychiatry 2020;81:19r12889.
85. McIntyre RS, Carvalho IP, Lui LMW et al. The effect of intravenous, intranasal, and oral ketamine in mood disorders: a meta-analysis. J Affect Disord 2020;276:576-84.
86. Fava M, Freeman MP, Flynn M et al. Double-blind, placebo-controlled, dose-ranging trial of intravenous ketamine as adjunctive therapy in treatment-resistant depression (TRD). Mol Psychiatry 2020;25:1592-603.
87. Tabuteau H, Jones A, Anderson A et al. Effect of AXS-05 (dextromethorphan-bupropion) in major depressive disorder: a randomized double-blind controlled trial. Am J Psychiatry 2022;179:490-9.
88. Fava M, Stahl S, Pani L et al. REL-1017 (esmethadone) as adjunctive treatment in patients with major depressive disorder: a phase 2a randomized double-blind trial. Am J Psychiatry 2021;179:122-31.
89. Preskorn S, Macaluso M, Mehra V et al. Randomized proof of concept trial of GLYX-13, an N-methyl-D-aspartate receptor glycine site partial agonist, in major depressive disorder nonresponsive to a previous antidepressant agent. J Psychiatr Pract 2015;21:140-9.
90. Meltzer-Brody S, Colquhoun H, Riesenberg R et al. Brexanolone injection in post-partum depression: two multicentre, double-blind, randomised, placebo-controlled, phase 3 trials. Lancet 2018; 392:1058-70.
91. Kanes S, Colquhoun H, Gunduz-Bruce H et al. Brexanolone (SAGE-547 injection) in post-partum depression: a randomised controlled trial. Lancet 2017;390:480-9.
92. Deligiannidis KM, Meltzer-Brody S, Gunduz-Bruce H et al. Effect of zuranolone vs placebo in postpartum depression: a randomized clinical trial. JAMA Psychiatry 2021;78:951-9.
93. Gunduz-Bruce H, Silber C, Kaul I et al. Trial of SAGE-217 in patients with major depressive disorder. N Engl J Med 2019;381:903-11.
94. Sepanjnia K, Modabbernia A, Ashrafi M et al. Pioglitazone adjunctive therapy for moderate-to-severe major depressive disorder: randomized double-blind placebo-controlled trial. Neuropsychopharmacology 2012;37:2093-100.
95. Mischoulon D, Hylek L, Yeung AS et al. Randomized, proof-of-concept trial of low dose naltrexone for patients with breakthrough symptoms of major depressive disorder on antidepressants. J Affect Disord 2017;208:6-14.
96. Fava M, Thase ME, Trivedi MH et al. Opioid system modulation with buprenorphine/samidorphan combination for major depressive disorder: two randomized controlled studies. Mol Psychiatry 2020;25:1580-91.
97. Zajecka JM, Stanford AD, Memisoglu A et al. Buprenorphine/samidorphan combination for the adjunctive treatment of major depressive disorder: results of a phase III clinical trial (FORWARD-3). Neuropsychiatr Dis Treat 2019;15:795-808.
98. Dinoff A, Lynch ST, Sekhri N et al. A meta-analysis of the potential antidepressant effects of buprenorphine versus placebo as an adjunctive pharmacotherapy for treatment-resistant depression. J Affect Disord 2020;271:91-9.
99. Costi S, Morris LS, Kirkwood KA et al. Impact of the KCNQ2/3 channel opener ezogabine on reward circuit activity and clinical symptoms in depression: results from a randomized controlled trial. Am J Psychiatry 2021;178:437-46.
100. Magid M, Reichenberg JS, Poth PE et al. Treatment of major depressive disorder using botulinum toxin A: a 24-week randomized, double-blind, placebo-controlled study. J Clin Psychiatry 2014;75:837-44.
101. Nagele P, Palanca BJ, Gott B et al. A phase 2 trial of inhaled nitrous oxide for treatment-resistant major depression. Sci Transl Med 2021;13:eabe1376.
102. Palhano-Fontes F, Barreto D, Onias H et al. Rapid antidepressant effects of the psychedelic ayahuasca in treatment-resistant depression: a randomized placebo-controlled trial. Psychol Med 2019;49:655-63.
103. Davis AK, Barrett FS, May DG et al. Effects of psilocybin-assisted therapy on major depressive disorder: a randomized clinical trial. JAMA Psychiatry 2021;78:481-9.
104. Carhart-Harris R, Giribaldi B, Watts R et al. Trial of psilocybin versus escitalopram for depression. N Engl J Med 2021;384:1402-11.
105. Abdallah MS, Mosalam EM, Zidan AAA et al. The antidiabetic metformin as an adjunct to antidepressants in patients with major depressive disorder: a proof-of-concept, randomized, double-blind, placebo-controlled trial. Neurotherapeutics 2020;17:1897-906.
106. Gordon JL, Rubinow DR, Eisenlohr-Moul TA et al. Efficacy of transdermal estradiol and micronized progesterone in the prevention of depressive symptoms in the menopause transition: a randomized clinical trial. JAMA Psychiatry 2018; 75:149-57.
107. Mitchell JM, Bogenschutz M, Lilienstein A et al. MDMA-assisted therapy for severe PTSD: a randomized, double-blind, placebo-controlled phase 3 study. Nat Med 2021;27:1025-33.
108. Mithoefer MC, Feduccia AA, Jerome L et al. MDMA-assisted psychotherapy for treatment of PTSD: study design and rationale for phase 3 trials based on pooled analysis of six phase 2 randomized controlled trials. Psychopharmacology 2019;236:2735-45.
109. Jerome L, Feduccia AA, Wang JB et al. Long-term follow-up outcomes of MDMA-assisted psychotherapy for treatment of PTSD: a longitudinal pooled analysis of six phase 2 trials. Psychopharmacology 2020;237:2485-97.
110. Feduccia AA, Jerome L, Yazar-Klosinski B et al. Breakthrough for trauma treatment: safety and efficacy of MDMA-assisted psychotherapy compared to paroxetine and sertraline. Front Psychiatry 2019;10:650.
111. Mithoefer MC, Mithoefer AT, Feduccia AA et al. 3,4-methylenedioxymethamphetamine (MDMA)-assisted psychotherapy for post-traumatic stress disorder in military veterans, firefighters, and police officers: a randomised, double-blind, dose-response, phase 2 clinical trial. Lancet Psychiatry 2018;5:486-97.
112. Mithoefer MC, Wagner MT, Mithoefer AT et al. The safety and efficacy of {+/-}3,4-methylenedioxymethamphetamine-assisted psychotherapy in subjects with chronic, treatment-resistant posttraumatic stress disorder: the first randomized controlled pilot study. J Psychopharmacol 2011; 25:439-52.
113. Mithoefer MC, Wagner MT, Mithoefer AT et al. Durability of improvement in post-traumatic stress disorder symptoms and absence of harmful effects or drug dependency after 3,4-methylenedioxymethamphetamine-assisted psychotherapy: a prospective long-term follow-up study. J Psychopharmacol 2013;27:28-39.
114. Oehen P, Traber R, Widmer V et al. A randomized, controlled pilot study of MDMA (±3,4-methylenedioxymethamphetamine)-assisted psychotherapy for treatment of resistant, chronic posttraumatic stress disorder (PTSD). J Psychopharmacol 2012;27:40-52.
115. Reinecke A, Nickless A, Browning M et al. Neurocognitive processes in d-cycloserine augmented single-session exposure therapy for anxiety: a randomized placebo-controlled trial. Behav Res Ther 2020;129:103607.
116. Smits JAJ, Pollack MH, Rosenfield D et al. Dose timing of d-cycloserine to augment exposure therapy for social anxiety disorder: a randomized clinical trial. JAMA Netw Open 2020;3:e206777.
117. Dutcher CD, Dowd SM, Zalta AK et al. Sleep quality and outcome of exposure therapy in adults with social anxiety disorder. Depress Anxiety 2021;38: 1182-90.
118. Hofmann SG, Papini S, Carpenter JK et al. Effect of d-cycloserine on fear extinction training in adults with social anxiety disorder. PLoS One 2019;14:e0223729.
119. Hofmann SG, Carpenter JK, Otto MW et al. Dose timing of D-cycloserine to augment cognitive behavioral therapy for social anxiety: study design and rationale. Contemp Clin Trials 2015;43:223-30.
120. Smits JAJ, Rosenfield D, Otto MW et al. D-cycloserine enhancement of exposure therapy for social anxiety disorder depends on the success of exposure sessions. J Psychiatr Res 2013;47:1455-61.
121. Hofmann SG, Smits JAJ, Rosenfield D et al. D-cycloserine as an augmentation strategy with cognitive-behavioral therapy for social anxiety disorder. Am J Psychiatry 2013;170:751-8.
122. Zalta AK, Dowd S, Rosenfield D et al. Sleep quality predicts treatment outcome in CBT for social anxiety disorder. Depress Anxiety 2013;30:1114-20.
123. Morley KC, Baillie A, Fraser I et al. Baclofen in the treatment of alcohol dependence with or without liver disease: multisite, randomised, double-blind, placebo-controlled trial. Br J Psychiatry 2018;212: 362-9.
124. Anton RF, Latham P, Voronin K et al. Efficacy of gabapentin for the treatment of alcohol use disorder in patients with alcohol withdrawal symptoms: a randomized clinical trial. JAMA Intern Med 2020;180:728-36.
125. Burnette EM, Ray LA, Irwin MR et al. Ibudilast attenuates alcohol cue-elicited frontostriatal functional connectivity in alcohol use disorder. Alcohol Clin Exp Res 2021;45:2017-28.
126. Grabski M, McAndrew A, Lawn W et al. Adjunc-

tive ketamine with relapse prevention-based psychological therapy in the treatment of alcohol use disorder. Am J Psychiatry 2022;179:152-62.
127. Coffin PO, Santos G-M, Hern J et al. Effects of mirtazapine for methamphetamine use disorder among cisgender men and transgender women who have sex with men: a placebo-controlled randomized clinical trial. JAMA Psychiatry 2020; 77:246-55.
128. Trivedi MH, Walker R, Ling W et al. Bupropion and naltrexone in methamphetamine use disorder. N Engl J Med 2021;384:140-53.
129. Dakwar E, Levin F, Foltin RW et al. The effects of subanesthetic ketamine infusions on motivation to quit and cue-induced craving in cocaine-dependent research volunteers. Biol Psychiatry 2014;76:40-6.
130. Dunn KE, Marcus TF, Pharm CK et al. Zonisamide reduces withdrawal symptoms but does not enhance varenicline-induced smoking cessation. Nicotine Tob Res 2016;18:1171-9.
131. Walter M, Bentz D, Schicktanz N et al. Effects of cortisol administration on craving in heroin addicts. Transl Psychiatry 2015;5:e610.
132. Rinne JO, Wesnes K, Cummings JL et al. Tolerability of ORM-12741 and effects on episodic memory in patients with Alzheimer's disease. Alzheimers Dement 2016;3:1-9.
133. Potter H, Woodcock JH, Boyd TD et al. Safety and efficacy of sargramostim (GM-CSF) in the treatment of Alzheimer's disease. Alzheimers Dement 2021;7:e12158.
134. Matthews DC, Ritter A, Thomas RG et al. Rasagiline effects on glucose metabolism, cognition, and tau in Alzheimer's dementia. Alzheimers Dement 2021;7:e12106.
135. Grossberg GT, Kohegyi E, Mergel V et al. Efficacy and safety of brexpiprazole for the treatment of agitation in Alzheimer's dementia: two 12-week, randomized, double-blind, placebo-controlled trials. Am J Geriatr Psychiatry 2020;28:383-400.
136. Cummings JL, Lyketsos CG, Peskind ER et al. Effect of dextromethorphan-quinidine on agitation in patients with Alzheimer disease dementia: a randomized clinical trial. JAMA 2015;314:1242-54.
137. Herrmann N, Ruthirakuhan M, Gallagher D et al. Randomized placebo-controlled trial of nabilone for agitation in Alzheimer's disease. Am J Geriatr Psychiatry 2019;27:1161-73.
138. Khoury R, Marx C, Mirgati S et al. AVP-786 as a promising treatment option for Alzheimer's disease including agitation. Expert Opin Pharmacother 2021;22:783-95.
139. Moline M, Thein S, Bsharat M et al. Safety and efficacy of lemborexant in patients with irregular sleep-wake rhythm disorder and Alzheimer's disease dementia: results from a phase 2 randomized clinical trial. J Prev Alzheimers Dis 2021;8:7-18.
140. Herring WJ, Ceesay P, Snyder E et al. Polysomnographic assessment of suvorexant in patients with probable Alzheimer's disease dementia and insomnia: a randomized trial. Alzheimers Dement 2020;16:541-51.
141. O'Gorman C, Jones A, Cummings JL et al. Efficacy and safety of AXS-05, a novel, oral, NMDA-receptor antagonist with multimodal activity, in agitation associated with Alzheimer's disease: results from ADVANCE-1, a phase 2/3, double-blind, active and placebo-controlled trial. Alzheimers Dement 2020;16:e047684.
142. Ballard C, Banister C, Khan Z et al. Evaluation of the safety, tolerability, and efficacy of pimavanserin versus placebo in patients with Alzheimer's disease psychosis: a phase 2, randomised, placebo-controlled, double-blind study. Lancet Neurol 2018;17:213-22.
143. Ballard C, Youakim JM, Coate B et al. Pimavanserin in Alzheimer's disease psychosis: efficacy in patients with more pronounced psychotic symptoms. J Prev Alzheimers Dis 2019;6:27-33.
144. Tariot PN, Cummings JL, Soto-Martin ME et al. Trial of pimavanserin in dementia-related psychosis. N Engl J Med 2021;385:309-19.
145. Dunn B, Stein P, Cavazzoni P. Approval of aducanumab for Alzheimer disease – the FDA's perspective. JAMA Intern Med 2021;181:1276-8.
146. Leon AC. Evolution of psychopharmacology trial design and analysis: six decades in the making. J Clin Psychiatry 2011;72:331-40.
147. Kairalla JA, Coffey CS, Thomann MA et al. Adaptive trial designs: a review of barriers and opportunities. Trials 2012;13:145.
148. Vandemeulebroecke M. Group sequential and adaptive designs – a review of basic concepts and points of discussion. Biom J 2008;50:541-57.
149. Le Tourneau C, Lee JJ, Siu LL. Dose escalation methods in phase I cancer clinical trials. J Natl Cancer Inst 2009;101:708-20.
150. Cheung K, Kaufmann P. Efficiency perspectives on adaptive designs in stroke clinical trials. Stroke 2011;42:2990-4.
151. Jensen RK, Leboeuf-Yde C, Wedderkopp N et al. Rest versus exercise as treatment for patients with low back pain and Modic changes. a randomized controlled clinical trial. BMC Med 2012;10:22.
152. Rosenberger WF, Sverdlov O, Hu F. Adaptive randomization for clinical trials. J Biopharm Stat 2012;22:719-36.
153. Thall PF, Fox P, Wathen J. Statistical controversies in clinical research: scientific and ethical problems with adaptive randomization in comparative clinical trials. Ann Oncol 2015;26:1621-8.
154. Proschan MA. Sample size re-estimation in clinical trials. Biom J 2009;51:348-57.
155. Wong ICK, Banaschewski T, Buitelaar J et al. Emerging challenges in pharmacotherapy research on attention-deficit hyperactivity disorder – outcome measures beyond symptom control and clinical trials. Lancet Psychiatry 2019;6:528-37.
156. Taipale H, Schneider-Thoma J, Pinzón-Espinosa J et al. Representation and outcomes of individuals with schizophrenia seen in everyday practice who are ineligible for randomized clinical trials. JAMA Psychiatry 2022;79:210-8.
157. Van Der Baan FH, Knol MJ, Klungel OH et al. Potential of adaptive clinical trial designs in pharmacogenetic research. Pharmacogenomics 2012;13:571-8.
158. Espay AJ, Guskey MT, Norton JC et al. Pimavanserin for Parkinson's Disease psychosis: effects stratified by baseline cognition and use of cognitive-enhancing medications. Mov Disord 2018;33:1769-76.
159. Horowitz MA, Macaulay A, Taylor D. Limitations in research on maintenance treatment for individuals with schizophrenia. JAMA Psychiatry 2022;79:83-5.
160. Correll CU, Jain R, Meyer JM et al. Relationship between the timing of relapse and plasma drug levels following discontinuation of cariprazine treatment in patients with schizophrenia: indirect comparison with other second-generation antipsychotics after treatment discontinuation. Neuropsychiatr Dis Treat 2019;15:2537-50.
161. Carvalho AF, Quevedo J, McIntyre RS et al. Treatment implications of predominant polarity and the polarity index: a comprehensive review. Int J Neuropsychopharmacol 2015;18:pyu079.
162. Nestsiarovich A, Gaudiot CES, Baldessarini RJ et al. Preventing new episodes of bipolar disorder in adults: systematic review and meta-analysis of randomized controlled trials. Eur Neuropsychopharmacol 2022;54:75-89.
163. Kishimoto T, Hagi K, Kurokawa S et al. Efficacy and safety/tolerability of antipsychotics in the treatment of adult patients with major depressive disorder: a systematic review and meta-analysis. Psychol Med 2022; doi: 10.1017/S0033291722000745.
164. Trivedi MH, Rush J. Does a placebo run-in or a placebo treatment cell affect the efficacy of antidepressant medications? Neuropsychopharmacology 1994;11:33-43.
165. Scott AJ, Sharpe L, Quinn V et al. Association of single-blind placebo run-in periods with the placebo response in randomized clinical trials of antidepressants: a systematic review and meta-analysis. JAMA Psychiatry 2022;79:42-9.
166. Salloum NC, Fava M, Ball S et al. Success and efficiency of phase 2/3 adjunctive trials for MDD funded by industry: a systematic review. Mol Psychiatry 2020;25:1967-74.
167. Papakostas GI, Vitolo OV, Ishak WW et al. A 12-week, randomized, double-blind, placebo-controlled, sequential parallel comparison trial of ziprasidone as monotherapy for major depressive disorder. J Clin Psychiatry 2012;73:1541-7.
168. Papakostas GI, Shelton RC, Zajecka JM et al. L-methylfolate as adjunctive therapy for SSRI-resistant major depression: results of two randomized, double-blind, parallel-sequential trials. Am J Psychiatry 2012;169:1267-74.
169. Fava M, Mischoulon D, Iosifescu D et al. A double-blind, placebo-controlled study of aripiprazole adjunctive to antidepressant therapy among depressed outpatients with inadequate response to prior antidepressant therapy (ADAPT-A Study). Psychother Psychosom 2012;81:87-97.
170. Fava M, Evins AE, Dorer DJ et al. The problem of the placebo response in clinical trials for psychiatric disorders: culprits, possible remedies, and a novel study design approach. Psychother Psychosom 2003;72:115-27.
171. Fava M. Implications of a biosignature study of the placebo response in major depressive disorder. JAMA Psychiatry 2015;72:1073-4.
172. Baer L, Ivanova A. When should the sequential parallel comparison design be used in clinical trials? Clin Investig 2013;3:823-33.
173. Ivanova A, Qaqish B, Schoenfeld DA. Optimality, sample size, and power calculations for the sequential parallel comparison design. Stat Med 2011;30:2793-803.
174. Stallard N, Todd S. Seamless phase II/III designs. Stat Methods Med Res 2011;20:623-34.
175. Collignon O, Koenig F, Koch A et al. Adaptive designs in clinical trials: from scientific advice to marketing authorisation to the European Medicine Agency. Trials 2018;19:642.
176. Schüler S, Kieser M, Rauch G. Choice of futility boundaries for group sequential designs with two endpoints. BMC Med Res Methodol 2017;17:119.
177. Asakura K, Hamasaki T, Evans SR. Interim evaluation of efficacy or futility in group-sequential trials with multiple co-primary endpoints. Biom J 2017;59:703-31.
178. Shen J, Preskorn S, Dragalin V et al. How adaptive

trial designs can increase efficiency in psychiatric drug development: a case study. Innov Clin Neurosci 2011;8:26-34.
179. Fleischhacker WW, Burns T, European Group For Research In Schizophrenia. Feasibility of placebo-controlled clinical trials of antipsychotic compounds in Europe. Psychopharmacology 2002; 162:82-4.
180. Hummer M, Holzmeister R, Kemmler G et al. Attitudes of patients with schizophrenia toward placebo-controlled clinical trials. J Clin Psychiatry 2003;64:277-81.
181. Roberts LW. The ethical basis of psychiatric research: conceptual issues and empirical findings. Compr Psychiatry 1998;39:99-110.
182. Kemmler G, Hummer M, Widschwendter C et al. Dropout rates in placebo-controlled and active-control clinical trials of antipsychotic drugs: a meta-analysis. Arch Gen Psychiatry 2005;62:1305-12.
183. Calabrese JR, Pikalov A, Streicher C et al. Lurasidone in combination with lithium or valproate for the maintenance treatment of bipolar I disorder. Eur Neuropsychopharmacol 2017;27:865-76.
184. Keck PE Jr, Welge JA, Strakowski SM et al. Placebo effect in randomized, controlled maintenance studies of patients with bipolar disorder. Biol Psychiatry 2000;47:756-61.
185. Kemp AS, Schooler NR, Kalali AH et al. What is causing the reduced drug-placebo difference in recent schizophrenia clinical trials and what can be done about it? Schizophr Bull 2010;36:504-9.
186. Loebel A, Cucchiaro J, Siu C et al. Signal detection in clinical trials: a post-study survey of schizophrenia trial sites. Presented at the Autumn Conference of the International Society for CNS Clinical Trials and Methodology, Baltimore, October 2010.
187. Fava M. The role of regulators, investigators, and patient participants in the rise of the placebo response in major depressive disorder. World Psychiatry 2015;14:307-8.
188. Leucht S, Leucht C, Huhn M et al. Sixty years of placebo-controlled antipsychotic drug trials in acute schizophrenia: systematic review, Bayesian meta-analysis, and meta-regression of efficacy predictors. Am J Psychiatry 2017;174:927-42.
189. Rutherford BR, Pott E, Tandler JM et al. Placebo response in antipsychotic clinical trials: a meta-analysis. JAMA Psychiatry 2014;71:1409-21.
190. Fournier JC, DeRubeis RJ, Hollon SD et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. JAMA 2010;303:47-53.
191. Iovieno N, Papakostas GI. Correlation between different levels of placebo response rate and clinical trial outcome in major depressive disorder: a meta-analysis. J Clin Psychiatry 2012;73:1300-6.
192. Kirsch I, Deacon BJ, Huedo-Medina TB et al. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. PLoS Med 2008;5:e45.
193. Agid O, Siu CO, Potkin SG et al. Meta-regression analysis of placebo response in antipsychotic trials, 1970-2010. Am J Psychiatry 2013;170:1335-44.
194. Taiminen T, Syvälahti E, Saarijärvi S et al. Prediction of positive placebo response among chronic schizophrenic outpatients. J Nerv Ment Dis 1996;184:109-13.
195. Tedeschini E, Fava M, Papakostas GI. Placebo-controlled, antidepressant clinical trials cannot be shortened to less than 4 weeks' duration: a pooled analysis of randomized clinical trials employing a diagnostic odds ratio-based approach. J Clin Psychiatry 2011;72:98-103.
196. Younis IR, Gopalakrishnan M, Mathis M et al. Association of end point definition and randomized clinical trial duration in clinical trials of schizophrenia medications. JAMA Psychiatry 2020;77:1064-71.
197. Leucht S, Chaimani A, Mavridis D et al. Disconnection of drug-response and placebo-response in acute-phase antipsychotic drug trials on schizophrenia? Meta-regression analysis. Neuropsychopharmacology 2019;44:1955-66.
198. Weimer K, Colloca L, Enck P. Placebo effects in psychiatry: mediators and moderators. Lancet Psychiatry 2015;2:246-57.
199. Leucht S, Crippa A, Siafis S et al. Dose-response meta-analysis of antipsychotic drugs for acute schizophrenia. Am J Psychiatry 2019;177:342-53.
200. Wallace ML, Frank E, Kraemer HC. A novel approach for developing and interpreting treatment moderator profiles in randomized clinical trials. JAMA Psychiatry 2013;70:1241-7.
201. Zimmerman M, Clark HL, Multach MD et al. Have treatment studies of depression become even less generalizable? A review of the inclusion and exclusion criteria used in placebo-controlled antidepressant efficacy trials published during the past 20 years. Mayo Clin Proc 2015;90:1180-6.
202. Nazha B, Mishra M, Pentz R et al. Enrollment of racial minorities in clinical trials: old problem assumes new urgency in the age of immunotherapy. Am Soc Clin Oncol Educ B 2019;39:3-10.
203. McGuire TG, Miranda J. New evidence regarding racial and ethnic disparities in mental health: policy implications. Health Aff 2008;27:393-403.
204. Clark LT, Watkins L, Piña IL et al. Increasing diversity in clinical trials: overcoming critical barriers. Curr Probl Cardiol 2019;44:148-72.
205. Coakley M, Fadiran EO, Parrish LJ et al. Dialogues on diversifying clinical trials: successful strategies for engaging women and minorities in clinical trials. J Womens Health 2012;21:713-6.
206. Shen J, Kobak KA, Zhao Y et al. Use of remote centralized raters via live 2-way video in a multicenter clinical trial for schizophrenia. J Clin Psychopharmacol 2008;28:691-3.
207. Sharp IR, Kobak KA, Osman DA. The use of videoconferencing with patients with psychosis: a review of the literature. Ann Gen Psychiatry 2011; 10:14.
208. Freeman MP, Pooley J, Flynn MJ et al. Guarding the gate: remote structured assessments to enhance enrollment precision in depression trials. J Clin Psychopharmacol 2017;37:176-81.
209. Corcoran CM, Cecchi GA. Using language processing and speech analysis for the identification of psychosis and other disorders. Biol Psychiatry Cogn Neurosci Neuroimaging 2020;5:770-9.
210. Parola A, Simonsen A, Bliksted V et al. Voice patterns in schizophrenia: a systematic review and Bayesian meta-analysis. Schizophr Res 2020; 216:24-40.
211. Liu D, Cheng D, Houle TT et al. Machine learning methods for automatic pain assessment using facial expression information: protocol for a systematic review and meta-analysis. Medicine 2018;97:e13421.
212. McIntyre RS, Lee Y, Rong C et al. Ecological momentary assessment of depressive symptoms using the mind.me application: convergence with the Patient Health Questionnaire-9 (PHQ-9). J Psychiatr Res 2021;135:311-7.
213. Yim SJ, Lui LMW, Lee Y et al. The utility of smartphone-based, ecological momentary assessment for depressive symptoms. J Affect Disord 2020; 274:602-9.
214. He-Yueya J, Buck B, Campbell A et al. Assessing the relationship between routine and schizophrenia symptoms with passively sensed measures of behavioral stability. NPJ Schizophr 2020;6:35.
215. Kane JM, Perlis RH, DiCarlo LA et al. First experience with a wireless system incorporating physiologic assessments and direct confirmation of digital tablet ingestions in ambulatory patients with schizophrenia or bipolar disorder. J Clin Psychiatry 2013;74:e533-40.
216. Bain EE, Shafner L, Walling DP et al. Use of a novel artificial intelligence platform on mobile devices to assess dosing compliance in a phase 2 clinical trial in subjects with schizophrenia. JMIR mHealth uHealth 2017;5:e18.
217. Hufford MR, Davis VG, Hilt D et al. Circadian rhythms in cognitive functioning among patients with schizophrenia: impact on signal detection in clinical trials of potential pro-cognitive therapies. Schizophr Res 2014;159:205-10.
218. Torous J, Bucci S, Bell IH et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021;20:318-35.
219. Montag C, Elhai JD, Dagum P. On blurry boundaries when defining digital biomarkers: how much biology needs to be in a digital biomarker? Front Psychiatry 2021;12:740292.
220. Marsch LA, Chen C-H, Adams SR et al. The feasibility and utility of harnessing digital health to understand clinical trajectories in medication treatment for opioid use disorder: D-TECT study design and methodological considerations. Front Psychiatry 2022;13:871916.
221. Abdul Rashid NA, Martanto W, Yang Z et al. Evaluating the utility of digital phenotyping to predict health outcomes in schizophrenia: protocol for the HOPE-S observational study. BMJ Open 2021;11:e046552.
222. Narkhede SM, Luther L, Raugh IM et al. Machine learning identifies digital phenotyping measures most relevant to negative symptoms in psychotic disorders: implications for clinical trials. Schizophr Bull 2022;48:425-36.
223. Jacobson NC, Bhattacharya S. Digital biomarkers of anxiety disorder symptom changes: personalized deep learning models using smartphone sensors accurately predict anxiety symptoms from ecological momentary assessments. Behav Res Ther 2022;149:104013.
224. Tseng Y-C, Lin EC, Wu CH et al. Associations among smartphone app-based measurements of mood, sleep and activity in bipolar disorder. Psychiatry Res 2022;310:114425.
225. Carlson S, Kim H, Devanand DP et al. Novel approaches to measuring neurocognitive functions in Alzheimer's disease clinical trials. Curr Opin Neurol 2022;35:240-8.
226. Ranjan T, Melcher J, Keshavan M et al. Longitudinal symptom changes and association with home time in people with schizophrenia: an observational digital phenotyping study. Schizophr Res 2022;243:64-9.
227. Cowan T, Cohen AS, Raugh IM et al. Ambulatory audio and video recording for digital phenotyping in schizophrenia: adherence & data usability. Psychiatry Res 2022;311:114485.
228. Kamath J, Leon Barriera R, Jain N et al. Digital phenotyping in depression diagnostics: integrating psychiatric and engineering perspectives. World J Psychiatry 2022;12:393-409.

229. Kraemer HC, Frank E. Evaluation of comparative treatment trials: assessing clinical benefits and risks for patients, rather than statistical effects on measures. JAMA 2010;304:683-4.
230. Kraemer HC, Frank E, Kupfer DJ. How to assess the clinical impact of treatments on patients, rather than the statistical impact of treatments on measures. Int J Methods Psychiatr Res 2011;20: 63-72.
231. Kraemer HC, Blasey C. How many subjects?: Statistical power analysis in research, 2nd ed. London: Sage Publications, 2016.
232. Senn S. Disappointing dichotomies. Pharm Stat 2003;2:239-40.
233. Neuhäuser M. How to deal with multiple endpoints in clinical trials. Fundam Clin Pharmacol 2006;20:515-23.
234. Kraemer HC. Messages for clinicians: moderators and mediators of treatment outcome in randomized clinical trials. Am J Psychiatry 2016; 173:672-9.
235. Heckman JJ. Sample selection bias as a specification error. Econometrica 1979;47:153-61.
236. Little RJA. Pattern-mixture models for multivariate incomplete data. J Am Stat Assoc 1993;88:125-34.
237. Vonesh EF, Greene T, Schluchter MD. Shared parameter models for the joint analysis of longitudinal data and event times. Stat Med 2006;25:143-63.

DOI:10.1002/wps.21056

All levels of the translational spectrum must be targeted to advance psychopharmacology and improve patient outcomes

Correll et al[1] correctly state that many psychiatric disorders remain insufficiently treated despite advances in psychopharmacology, and attribute this to the limited knowledge of pathophysiology of these disorders, the lack of biological markers precluding tailored treatment selection, the few mechanistic targets for treatment development, and the challenges with clinical trial design and conduct. Here I address the chasms at the various levels of the translational spectrum that should be targeted through innovations in order to advance psychopharmacology and improve outcomes for patients.

Drug discovery in psychiatry has been mostly driven by the pharmaceutical industry. The discovery of selective serotonin reuptake inhibitors and second-generation antipsychotics ushered in a "new era" of psychopharmacology in 1980s and 1990s. However, these drugs and their modifications, while claiming to provide better safety and tolerability, primarily targeted monoaminergic systems, similar to tricyclic antidepressants and first-generation antipsychotics. Any attempts to develop new drugs with novel targets, such as metabotropic glutamate receptors, CRF1 receptors, and tackykinin NK1 receptors, were met with failures.

As the pipeline for drug development in psychiatry was drying out, many major pharmaceutical companies announced ceasing further investments in this area, citing "very low probability and disproportionately high cost for attaining success"[2]. Indeed, it takes nearly nine years to bring a psychotropic drug to the market, and the likelihood of drug approval in psychiatry – which includes success in all phases of development leading to regulatory approval – is only 6.2%, which is the lowest amongst non-oncology diseases[3]. Thus, novel strategies to enhance success of drug discovery in psychiatry are urgently needed.

Pre-clinical assays – such as forced swim test and chronic mild stress, as well as stimulant induced locomotor activity and reduced prepulse inhibition – have been used to screen drugs for prediction of antidepressant and antipsychotic activity, along with positron emission tomography (PET) studies in humans to estimate receptor occupancy in order to determine appropriate dosing for therapeutic efficacy. These strategies have worked well in general for drugs that targeted the monoaminergic systems. However, drugs with actions on novel targets (such as NK1 receptors, CRF1 receptors and glutamatergic system), while demonstrating activity in some preclinical assays, did not succeed in phase 3 clinical trials. The general consensus is that newer pre-clinical tests that have better construct and predictive validity are urgently needed.

Attempts to improve construct validity by developing mouse models with knockout of genes implicated in schizophrenia have not proven to be helpful in consistently detecting drugs with antipsychotic activity[4]. Whether CRISPR-based gene editing to create knockout animal models might be more useful remains to be seen. Similarly, human induced pluripotent stem cells and brain organoids are being used to screen drugs for their effects in disease relevant cells, but their full potential is yet to be documented.

Phenotypic screening has been more successful than target-based approaches for drug development in central nervous system disorders. To this end, PsychoGenics has developed a phenotypic drug discovery platform called SmartCube, which uses a target-agnostic approach to screen compounds. This automated testing platform, through its customized hardware, presents a sequence of challenges to a mouse, collects massive amounts of data points, and uses proprietary machine learning algorithms to detect the potential for efficacy of compounds. SEP-363856 (ulotaront) was developed using this platform; it has trace amine-associated receptor 1 (TAAR-1) and serotonin 5-HT1A receptor agonistic properties, and has shown efficacy in a phase 2 clinical trial for schizophrenia[5]. The results of the phase 3 trials for this drug, and the efficacy of other compounds identified using this platform for other indications, will indicate whether it represents a significant advance over the previous models.

The success rate in phase 2 trials for drugs tested for psychiatric disorders is only 24%, which is the lowest among 14 disease areas[3]. Further, many psychotropic drugs that succeed in phase 2 fail in phase 3 trials. Correll et al[1] outline various reasons for such outcomes and suggest use of adaptive trial designs and strategies for minimizing placebo response to reduce the risk of failure.

Given that a high placebo response is a major contributor to failed trials, setting *a priori* a threshold for excluding all patients from centers with an improbable placebo response might be worth considering. In addition, academia must work in close collaboration with the industry to develop innovations in trial designs, and conduct in-depth analyses to take lessons from failed trials which will inform further drug development. For instance, the first trial of cariprazine for bipolar depression[6] failed due to a high placebo response rate of 60%. Knowledge from this and other trials was used to design subsequent phase 2/3 studies, all of which were positive, leading to cariprazine's approval by the US Food and Drug Administration (FDA)[7]. Despite a signal for efficacy in *post-hoc* analyses, a similar strategy was not pursued for agomelatine, which also had a 60% placebo response rate in a bipolar depression trial[8]. This illustrates the impact of business decisions by the industry on drug development in psychiatry.

While development of new drugs with novel mechanisms of action would be a welcome addition to the therapeutic armamentarium, there are limitations to the generalizability of data from randomized placebo-controlled trials. Real-world data coming from a variety of sources must be gathered in order to understand the effectiveness of treatments and tailor them to the needs of each individual. Most currently approved treatments for various psychiatric indications work for about 50% of patients, but there is little information to guide clinicians with regards to what treatment is most likely to work for which patient, and, if the first treatment is ineffective, what is the next most appropriate intervention.

Thus, there is an urgent need to incorporate approved treatments into real-world clinical practice protocols/algorithms, similar to cancer treatment protocols, to generate evidence and move the field towards precision psychiatry. Such efforts could be further bolstered by using learning health care systems in clinical practice settings and collecting data that could be analyzed for discovery of biomarkers that predict response to each treatment.

Moving along the translational spectrum, patients need to access care, and evidence-based treatments need to be used appropriately by clinicians. Although several evidence-based treatment options exist for some psychiatric disorders, such as major depressive disorder, unfortunately only 8% to 33% of patients with this disorder use mental health services, and only 3% to 23% receive minimally adequate treatment[9]. Further, even in developed countries such as the UK, adherence to evidence-based care pathways for treatment of depression is poor, with many patients not receiving guideline-concordant care. In order to address this translational chasm, governments must invest funds to bolster mental health services and support education aimed at addressing stigma. Moreover, health care organizations must make every effort to establish an infrastructure that promotes and supports evidence-based practices to optimize outcomes.

In conclusion, innovations need to occur at all levels of the translational spectrum to advance psychopharmacology and improve patient outcomes.

Lakshmi N. Yatham
University of British Columbia, Vancouver, BC, Canada

1. Correll CU, Solmi M, Cortese S et al. World Psychiatry 2023;22:48-74.
2. Miller G. Science 2010;329:502-4.
3. Thomas DW, Burns J, Audette J et al. Clinical development success rates 2006-2015. www.amplion.com.
4. Pratt J, Winchester C, Dawson N et al. Nat Rev Drug Discov 2012;11:560-79.
5. Koblan KS, Kent J, Hopkins SC et al. N Engl J Med 2020;382:1497-506.
6. Ahuja S, Bose A, Lu K et al. A multicenter, randomized, double-blind trial to evaluate the effect of cariprazine in bipolar depression. Presented at the Conference of the International Society for CNS Clinical Trials and Methodology, Amelia Island, October 2011.
7. Durgam S, Earley W, Lipschitz A et al. Am J Psychiatry 2016;173:271-81.
8. Yatham LN, Vieta E, Goodwin GM et al. Br J Psychiatry 2016;208:78-86.
9. Moitra M, Santomauro D, Collins PY et al. PLoS Med 2022;19:e1003901.

DOI:10.1002/wps.21060

Key considerations for clinical trials in psychopharmacology

The thoughtful review by Correll et al[1] explores the status of drugs for mental disorders with new mechanisms of action currently in testing, and details obstacles to developing such medications. The authors examined established clinical registries and identified ongoing clinical trials of agents that showed the most promise "as emerging from documented superiority over placebo, magnitude of the observed effects, and demonstration of requirements for safety and tolerability". In aggregate, the list of agents is quite encouraging. The paper, however, does not cover negative trials, although the field can learn much from well-conducted trials of drugs that did not separate from placebo; such studies can rule out a specific target, thereby potentially eliminating the unnecessary pursuit of a pathway unlikely to be fruitful.

The most useful part of the paper is the discussion by this group of well-known investigators of ongoing developments in clinical trial methodology, design and conduct that should be carefully considered when developing and testing pharmacological agents for the treatment of mental disorders. These recommendations, which could be used to de-risk trial programs of novel or repurposed agents, are state-of-the-art and should, if possible, be incorporated as much as possible into planned future trials. While all of these suggestions are very thoughtful, I particularly wish to expand upon two: the importance of early phase 2 proof-of-concept studies and of identifying a treatment's precise mechanism of action.

A key and largely unaddressed issue in clinical trials is the ever-increasing placebo-response rates and the resulting diminishment of drug-placebo differences in efficacy over time. As Correll et al point out, solutions such as increasing sample size and adding more study sites have not improved our ability to discern drug efficacy versus placebo, though they have increased the cost of conducting such studies.

In this context, although adequately powered phase 2 and 3 studies are certainly necessary at some point, the importance of smaller, well-controlled and well-conducted phase 2A studies should not be minimized. Such studies have the potential to identify an important efficacy signal that would then allow investigators to move forward more confidently with larger and more costly phase 2 studies. As a key example, one of the pivotal studies in the US Food and Drug Administration (FDA)'s approval of valproate for mania included 36 participants (N=17 valproate, N=19 placebo)[2]. A more recent example concerns the approval of brexanolone for postpartum depression: one of the first reports was a case series of only four women[3], and a subsequent small randomized trial had only 21 participants with postpartum depression (N=10 brexanolone, N=11 placebo)[4]. Ketamine provides another key example: the initial study investigating racemic ketamine's antidepressant effects was a small, controlled trial of seven participants with major depression, followed by a second study of 17 participants with treatment-resistant depression[5,6]. Despite their small size, these two studies were influential in the development and ultimate FDA approval of esketamine for treatment-resistant depression.

These examples underscore how astute clinical observation and small, well-designed, proof-of-concept studies provide a useful strategy for de-risking any novel agent's path to approval. Findings from small early trials can inform go/no-go decisions regarding whether to move forward with larger, well-powered phase 2 studies with effect sizes large enough to survive the elevated placebo rates associated with moving from experimental settings to real-world studies. This approach is of considerable in-

terest to a clinical neuroscience industry that seeks to de-risk failures occurring during phases 2 and 3[7]. In addition, early proof-of-concept studies help identify critical feasibility, safety and design issues before jumping into larger and costlier phase 2 and 3 studies.

Correll et al correctly identify the considerable discrepancy between indication-based nomenclature and the clinical use of psychotropics. They further note that pharmacological nomenclature is arcane and does not completely relate to mechanisms of action. Important recent efforts have led to the creation of a neuroscience-based nomenclature for psychotropics[8]. Multiple international societies and scientific organizations have joined these efforts. Likewise, journals, book publishers and academic curricula have begun to refer to psychotropic medications based on their presumed mechanisms of action. Such important efforts are likely to facilitate scientific communication and move drug development forward. Nevertheless, our knowledge of drug mechanisms is still in its infancy, and nomenclature is likely to change with new insights or findings. In other words, any given medication's presumed mechanism of action is a rapidly evolving concept.

Ketamine provides a salient example. Specifically, ketamine is an N-methyl-D-aspartate receptor (NMDAR) antagonist. While this mechanism is relevant to ketamine's anesthetic properties, the degree to which it underlies its antidepressant properties is a topic of much debate, with evidence on both sides. This question is vital because, if NMDAR antagonism does not underlie ketamine's antidepressant effects, then the field – which seeks to develop a safer alternative to ketamine – should cease chasing a target unlikely to be relevant.

Indeed, multiple NMDAR antagonists have demonstrated no antidepressant efficacy in treatment-resistant depression[9], though some such agents remain in play, including the recently approved AXS-05 (dextromethorphan+bupropion). Although its maker has described NMDAR antagonism as AXS-05's primary mechanism of action, it should be noted that this drug is also a sigma-1 agonist, a nicotinic acetylcholine receptor antagonist, and a serotonin/noradrenaline/dopamine reuptake inhibitor. To date, no significant studies have explored which of these mechanisms might be the most relevant. Because AXS-05 is distinct from most currently available antidepressants, exploring its relevant mechanisms of action may provide novel targets to pursue in clinical trials.

An important limitation to progress in this area, however, is that the field has few ways to identify more precise, mechanistically-relevant biomarkers, although some promising ones are currently under investigation. To date, many of our proposed therapeutic targets were identified via *in vitro* or *in vivo* non-human assays, so our ability to assess whether a suspected mechanism of action is relevant or not remains limited. For example, no suitable positron emission tomography (PET) ligands are yet available to study potential NMDAR antagonists, even though two NMDAR antagonists, esketamine and AXS-05, are FDA-approved to treat depression.

In conclusion, Correll et al's review thoughtfully addresses some of the pitfalls associated with current methods for developing pharmacological treatments with a novel mechanism of action. The solutions that the authors propose are likely to increase the availability of novel treatments for our patients, some of which will hopefully be more effective than available ones. Nevertheless, despite the new targets in the pipeline, it should be noted that, with a few key exceptions (ketamine, brexanolone), no new treatment developed in the past several decades for any psychiatric condition has proven significantly superior to existing treatments in the sense of being disease-modifying. In this context, reverse engineering of the new treatments that are identified as unique in some aspects, such as ketamine – that is, using them as tools to better understand the cellular and molecular mechanisms of the specific disorder under study – might offer the opportunity to develop more effective next-generation treatments. Indeed, such work is already underway.

Carlos A. Zarate Jr
National Institute of Mental Health, Bethesda, MD, USA

Funding for this work was provided by the Intramural Research Program at the US National Institute of Mental Health, National Institutes of Health (NIH) (ZIAMH002857). The views expressed do not necessarily reflect those of the NIH, the US Department of Health and Human Services, or the US Government.

1. Correll CU, Solmi M, Cortese S et al. World Psychiatry 2023;22:48-74.
2. Pope HG Jr, McElroy SL, Keck PE Jr et al. Arch Gen Psychiatry 1991;48:62-8.
3. Kanes SJ, Colquhoun H, Doherty J et al. Hum Psychopharmacol 2017;32:e2576.
4. Kanes S, Colquhoun H, Gunduz-Bruce H et al. Lancet 2017;390:480-9.
5. Berman RM, Cappiello A, Anand A et al. Biol Psychiatry 2000;47:351-4.
6. Zarate CA Jr, Singh JB, Carlson PJ et al. Arch Gen Psychiatry 2006;63:856-64.
7. Hutson PH, Clark JA, Cross AJ. Annu Rev Pharmacol Toxicol 2017;57:171-87.
8. Zohar J, Levy DM. Eur Neuropsychopharmacol 2022;57:36-8.
9. Newport DJ, Carpenter LL, McDonald WM et al. Am J Psychiatry 2015;172:950-66.

DOI:10.1002/wps.21061

Real changes can enhance information yield on novel psychopharmacologic agents

The excellent review of current efforts and issues in the field of psychopharmacology produced by Correll et al[1] does not, unfortunately, provide much that would convince skeptical decision makers that the future of psychopharmacology will look that much different from the past. I write from the perspective of selection of compounds and mechanisms for clinical development as well as of implementation of clinical studies across phases 1-3, both from the industry and the US National Institute of Mental Health (NIMH) vantage points. As both a past decision maker and a current advisor, I will focus on what I believe has greatest

promise for the future of psychopharmacology over the next five to ten years.

Three thematic areas are implicit in Correll et al's review: a) what have we learned that is most useful in terms of design and implementation of clinical trials which herald a better future?; b) what should we do to de-risk both compound selection and dose setting for clinical trials that will improve productivity in terms of knowledge gained as well as advancing compounds?; c) what impact is likely to derive from emerging technologies provided by such US National Institutes of Health (NIH)-funded efforts as the Brain Initiative[2], and from utilization of remote technologies to passively and actively monitor participants in studies?

In the review, what seems to be the basis for identifying promising compounds is that there is a positive phase 2 study. Given the history of positive phase 2 studies that do not lead to successful phase 3 development, most decision makers would not see that these are any more promising than those that have failed in the past. What would be more convincing is evidence of what we have learned that can make future phase 2 trials more informative and predictive. For instance, an analysis showing that use of adaptive designs resulted in more efficient and successful drug development programs, or even a *post-hoc* analysis showing some common flaws in failed phase 3 programs that would allow focus on one or a limited number of variables that could be better managed.

One trial design element that is cited as having been shown not to work, based on meta-analyses of trials dating back to 1994 and recently confirmed, the single-blind lead-in, provides an excellent example of how advances can be made when data are shared. The field might be able to align on eliminating other wasteful practices if there were some way to share relevant data from as many as possible well-powered trials conducted over the last decade, whether or not they resulted in approval by the US Food and Drug Administration (FDA). Such an effort could include NIH-funded studies as well. One current effort to generate support for this kind of data sharing is provided by a panel on this topic scheduled for an upcoming meeting of the International Society for CNS Clinical Trials and Methodology.

The point made in the review that "the strongest the rationale for the randomized controlled trial (RCT), the more de-risked the trial will be"[1] raises the question of what constitutes a strong rationale, given a history of rationales – such as the one for targeting amyloid in Alzheimer's disease – not so far delivering after cumulative investments in the billions of dollars.

Although questions remain, I believe that having solid information on the relationship between a dose of a potential new drug and the degree to which it interacts with its primary site(s) of action in the brain and can be linked to downstream changes in brain function will allow future clinical trials to be better interpreted. One would lower risk of failure by avoiding compounds without robust translational pharmacokinetic-pharmacodynamic (PK/PD) brain effect data. Indeed, a recent analysis of industry success rates of compounds that had full target engagement packages across therapeutic areas reported that 12 of 14 yielded positive proof-of-concept studies, with eight advancing to phase 3, versus none of 12 compounds for which evidence of target engagement was weak or missing[3].

As a corollary, since animals do not provide true models of syndromal clinical brain conditions (except perhaps drug dependence), the future is likely to use evidence of effects on some domains in an animal assay that might be translated into humans for either a broadly defined syndromal disorder or a domain of function, as a core part of building the rationale for advancing a mechanism and/or compound. Such is the potential benefit of building out the Research Domain Criteria approach[4].

As an example, the so-called Fast-Fail approach piloted by the NIMH[5], which complements approaches being taken with industry to generate rationales to pursue a domain such as cognitive impairment in schizophrenia, has shown promise. A specific kappa opiate receptor antagonist, for which brain receptor occupancy data were available, was investigated in terms of potential for the domain of anhedonia. The drug was shown to positively affect a reward task-associated functional magnetic resonance imaging (fMRI) signal and to specifically improve severity of apathy in a group of individuals with DSM depression and anxiety spectrum disorders[6]. This finding was seen as de-risking future studies, and led to large pharma investment in a phase 2 trial followed by a just initiated phase 3 program (NCT03559192 and NCT05518149).

This approach goes beyond examples of selecting subsets of a DSM diagnosed group, such as failure to respond to standard treatments, or restricting subjects to those below a certain age and fewer hospitalizations, as noted for the positive phase 2 trial of ulotaront in schizophrenia. For novel mechanisms, as part of a de-risking strategy, one should first show whether any effect can be detected on some domain of function. Then, one should decide what syndromal disorder(s) might best benefit from the compound.

This domain approach might also help de-risk compounds with three or more pharmacological mechanisms that might be affected in humans, which are problematic in terms of demonstrating target engagement across dose ranges. A functional brain measure that translates from animals to humans, or even one with some degree of "face validity" in humans, can be applied to any molecule, whatever its mix of known mechanisms, or even initially unknown mechanisms. For compounds such as ulotaront, a promising antipsychotic discovered with a phenotypic assay battery (Smart Cube)[7], a functional brain measure can potentially be used to set doses in humans prior to identification of molecular mechanisms and development of specific target engagement tools. Assessment of brain function prior to clinical testing is likely to become more and more part of psychopharmacology.

The utility of emerging methods, such as differentiating pluripotent cells from individuals into a neuronal type in which compounds can be tested prior to be administered, to see if some functional effect detectable *in vitro* predicts activity in humans, remains to be demonstrated. Nonetheless, if early reports of predicting aspects of lithium response in cells from bipolar patients[8] generalize to drug response predictions, this approach may become an important addition to the future of psychopharmacology.

Similarly, by then we should have enough experience to know if remote measures that can be gathered passively on a device or those resulting from approaches such as eco-

logical momentary assessments are more sensitive in terms of picking up systematic drug effects than traditional types of clinical measures. It seems likely that at least some of these assessments will reveal drug effects on one or another variable that we do not currently capture with existing methods.

In summary, beyond what is recommended by Correll et al's review, I predict that the near future of psychopharmacology will include a greater emphasis on target engagement PK/PD studies that can be translated from animals to humans, a focus on functional domains as a core part of building the rationale for advancing a mechanism or a compound, and the development of means for all interested parties to have access to relevant data to decide on design elements that influence signal detection in a trial.

William Z. Potter
Independent Expert, Philadelphia, PA, USA

1. Correll CU, Solmi M, Cortese S et al. World Psychiatry 2023;22:48-74.
2. Mott MC, Gordon JA, Koroshetz WJ. PLoS Biol 2018;16:e3000066.
3. Morgan P, Van Der Graaf PH, Arrowsmith J et al. Drug Discov Today 2012;17:419-24.
4. National Institute of Mental Health. Definitions of the RDoC domains and constructs. https://www.nimh.nih.gov.
5. Grabb MC, Cross AJ, Potter WZ et al. J Clin Psychopharmacol 2016;36:419-21.
6. Krystal AD, Pizzagalli DA, Smoski M et al. Nat Med 2020;26:760-8.
7. Dedic N, Jones PG, Hopkins SC et al. J Pharmacol Exp Ther 2015;371:1-14.
8. Mishra HK, Ying NM, Luis A et al. Mol Psychiatry 2021;26:3383-94.

DOI:10.1002/wps.21062

Will digital technology address the challenges of drug development in psychiatry?

Pharmacotherapy is likely to remain a mainstream treatment for many mental disorders. A great deal has been learned about psychotropic medications in the past 70 years, and treatment efficacy has improved significantly. However, pharmacotherapy is generally limited to symptomatic relief and cannot provide a cure. In addition, only a certain proportion of patients are able to achieve remission and/or recovery, and the complete disappearance of symptoms remains a distant goal.

The accurate allocation of patients to the most appropriate treatment option based on a deeper understanding of pathophysiology is now needed, along with the development of drugs with novel mechanisms of action. In other words, we need to realize "precision medicine" within psychiatry. To this end, conducting better clinical trials by solving current problems, thereby enabling the faster delivery of new drugs to patients, is important. The extensive review by Correll et al[1] provides very broad and detailed information regarding the above-mentioned issues and carefully explains what is needed to move forward.

As they mention, the lack of sample representativeness in clinical trials, the strong (and increasing yearly) placebo response, the high dropout rate, and the varying reliability of severity assessments are of particular concern. Digital phenotypes derived from personal digital devices[2] seem to have ample potential to address these problems. This potential could be further enhanced by successfully combining new ways of delivering health care using communication technologies such as telemedicine.

Clinical trials often require patients to travel long distances and to make frequent hospital visits, which may reduce the likelihood of trial success. Promoting decentralized clinical trials, i.e., systems that allow patients to participate in a trial without necessarily coming to the hospital, would facilitate patient recruitment and prevent dropouts. The use of digital data to quantify the severity of symptoms in an objective manner could also reduce variations in assessments made at different clinical sites. Frequent assessments are a major burden on patients, but by utilizing ecological momentary assessment via passive monitoring, a method that is becoming increasing feasible[3], therapeutic benefits that were previously difficult to detect might become identifiable. Given the potential of such digital technologies, it seems likely that many currently unmet needs will be addressed. However, the story is not that simple, and this is not a task that can be completed overnight.

A potentially important question in the use of these digital tools is whether they can assess a patient at a level similar to that of a skilled evaluator meeting the patient in person and taking the time to assess his/her psychopathology. There are many different types of digital phenotypes, ranging from those in which the patient actively provides input on his/her condition (called active data) to passive data, such as sensor data, that do not require the patient's active involvement. The latter provide a wide range of information, including data that can be collected from a smartphone such as geographic range of activity, call logs, text input and search logs, as well as data that can be collected from a wearable device, such as acceleration which can be translated into activity, sleep rhythm, heart rate (or pulse rate), and skin conductance. Furthermore, passive data can be obtained through smart speakers, cameras, or some other devices, for example patient language as quantified by natural language processing, speech rate, acoustics of speech, facial expression, posture and body movement.

Even if these data could objectively quantify a patient's behaviour and/or autonomic nervous system activity, they would not elicit the patient's thoughts or moods and could only serve as surrogate markers. Many studies have reported that it was possible to distinguish between patients with mental disorders and healthy volunteers[4], or detect early sign of relapse[5] with a relatively high degree of accuracy from these data, but there is still large room for improvement. Even when a pathological feature can be identified, it is often unclear whether it is a state or a trait marker[6].

Many of these predictive models utilize

machine learning, but it should be noted that, although this technique may fit the specific population from which the data were obtained, the generalizability of findings may not always be high. In addition, determining how to accommodate differences across patients' lifestyles is especially important: the identification of digital phenotypes common to patients across cultures might be difficult.

Nonetheless, the advancement of the above technologies and the accumulation of the relevant knowledge may benefit not only clinical trials but also real-world clinical practice. Gold-standard evaluations may be difficult to perform in time-constrained clinical settings, but "measurement-based psychiatry" could be delivered more easily with those technologies. In fact, commercially available wearable devices can already quantify sleep and activity, and some practitioners may be using such data to treat patients. Specifically, the accumulation of longitudinal data on individual patients would be useful for identifying changes over time. A large cohort study that collects digital data would allow to identify which patients with which digital phenotypes respond to which treatments. As a result, the selection of drugs with the greatest likelihood of being effective for individual patients might become possible.

Concerns about the use of digital tools in clinical practice should also be considered. The question is what kind of long-term changes might occur as face-to-face treatment is replaced by the use of information and communication technology and digital tools. One often discussed issue is the digital divide, i.e., the risk that those who are unable to successfully use digital tools will be left out of health care[7]. Since the COVID-19 pandemic, psychiatric care has been delivered almost entirely remotely in some countries, but it is necessary to investigate whether this has the same therapeutic effect as face-to-face care. A large body of evidence already shows that telemedicine is no less effective than face-to-face care, but it remains unresolved whether this is true even for long-term treatment over multiple years[8]. Furthermore, there is a chance that the focus will shift to improving digital device-derived outcomes rather than actual patient recovery, if treatment effects are assessed using digital phenotypes rather than humans.

As we accumulate digital phenotypic data in the future, it will be important to study how these data are connected to pathophysiology. For example, studies that explore the relationship between brain functional connectivity and digital phenotypes would be useful. If a treatment has been identified that is effective for a specific pattern of functional connectivity, digital phenotyping may be able to identify the patients who are the best candidates for that treatment.

Even if the above-mentioned hurdles are overcome and a regulatory-accepted digital methodology is developed, there is no guarantee that such a methodology would be the best way to quantify mental disorder symptoms over the long term. Sensing technology and analytical methods are constantly evolving, and they can quickly become obsolete. The continued use of once-established standards for many years might nullify the advantages of digital technologies[9].

In conclusion, a great potential seems to have emerged from the use of digital technologies to foster the progress of psychopharmacology. Interdisciplinary research and development with the goal of actually improving the outcomes of people with mental disorders are now needed.

Taishiro Kishimoto
Hills Joint Research Laboratory for Future Preventive Medicine and Wellness, Keio University School of Medicine, Tokyo, Japan

1. Correll CU, Solmi M, Cortese S et al. World Psychiatry 2023;22:48-74.
2. Torous J, Kiang MV, Lorme J et al. JMIR Ment Health 2016;3:e16.
3. Yim SJ, Lui LMW, Lee Y et al. J Affect Disord 2020; 274:602-9.
4. Tazawa Y, Liang KC, Yoshimura M et al. Heliyon 2020;6:e03274.
5. Henson P, D'Mello R, Vaidyam A et al. Transl Psychiatry 2021;11:28.
6. Schiweck C, Lutin E, De Raedt W et al. Front Psychiatry 2021;12:696170.
7. Hilty D, Naslund JA, Ahuja S et al. In: Stein DJ, Fineberg NA, Chamberlain SR (eds). Mental health in a digital world. London: Academic Press, 2022: 11-39.
8. Kinoshita S, Cortright K, Crawford A et al. Psychol Med 2020; doi: https://doi.org/10.1017/S0033291720004584.
9. Torous J, Bucci S, Bell IH et al. World Psychiatry 2021;20:318-35.

DOI:10.1002/wps.21063

Ongoing phase 2/3 trials of psychotropic drugs: is help finally on the way?

In their comprehensive review, Correll et al[1] identify four important problem areas that have slowed the development of better pharmacological treatments for people suffering from serious mental disorders, such as schizophrenia, major depressive disorder and bipolar disorder. These impediments include the limited knowledge of the pathophysiology of these disorders; the lack of biological markers to stratify patient groups and individualize treatment selection; a restricted number of potentially relevant mechanisms of action for novel drug development; and a variety of methodological problems that impair signal detection in randomized controlled trials (RCTs).

The review is divided into two segments, one summarizing current research on promising drugs being studied in phase 2 or 3 trials, and the other reviewing methodological refinements that might improve the validity and efficiency of clinical research. In this piece, I will largely focus on the areas that I know best, though the authors' review of recent developments in the treatment of dementia provides a sobering summary of just how much more work there needs to be done.

For acute treatment of schizophrenia and related disorders, the authors identified 176 trials of a diverse group of compounds, largely targeting non-dopaminergic mechanisms. They found that only about one quarter of these RCTs had reported results and, among these, only about one quarter demonstrated efficacy on the primary dependent

measure. Further, they determined that only a handful of these drugs had progressed to phase 3.

Two of the most interesting drugs that have moved on to phase 3 are KarXT, which is a fixed combination of xanomeline – a muscarinic M1/M4 agonist – and the peripherally acting anticholinergic trospium chloride[2], and ulotaront, the first trace amine-associated receptor 1 (TAAR-1) agonist to show efficacy in a placebo controlled trial[3]. Despite their substantial differences, both drugs are particularly noteworthy because of the absence of extrapyramidal and metabolic side effects. If efficacy and safety are confirmed in the next phase of larger scale studies, these compounds could go a long way towards addressing critical unmet needs, by virtue of having novel mechanisms of action and more favorable tolerability profiles. Unfortunately, the review also documents that another important unmet need in this area of therapeutics, namely treatment of negative symptoms, has not yielded much in the way of truly novel and promising developments.

It was not too long ago that the process of discovery of truly novel drugs for treatment of major depressive disorder seemed like an exercise in futility, as one after another drug with theoretically relevant mechanisms of action failed to delivery significant clinical effects[4]. What a difference a decade can make! The authors identified nearly 180 trials and found that 19 out of 43 RCTs had reported significant effects.

The serendipitous observation that intravenous ketamine – at sub-anesthetic doses – could have rapid and large antidepressant effects stimulated a wave of drug development focused on glutamatergic neurotransmission. The paradigm-changing nature of intravenous ketamine therapy extended beyond its mode of delivery and the rapidity of effects: this is a controlled substance, yet its antidepressant effects, which typically persist for 3-5 days, extend long after the intoxicating effects have dissipated.

It was also noteworthy that the dissociative and euphorogenic effects of intravenous ketamine were not closely linked to the likelihood of symptom improvement, which further suggested that the properties that lead to drug misuse or abuse are not essential to its antidepressant effects[4]. Nevertheless, there was considerable caution about the potential risks of this treatment, and nearly 20 years elapsed between the first observations of antidepressant effects and the approval by the US Food and Drug Administration (FDA) of the first treatment directly resulting from this line of research.

Beyond harvesting the "low hanging fruit", i.e. other modes of administration of ketamine and commercialization of its stereoisomers (S- and R-ketamine), research has also focused on other molecules that modulate glutamatergic neurotransmission, including a proprietary combination of dextromethorphan – the ancient cough suppressant – and bupropion[5]. This medication has recently been approved by the FDA for treatment of major depressive disorder, becoming the first orally administered treatment in this line of therapeutics. A second orally administered medication, esmethadone[6], is now in phase 3. Interestingly, despite its lineage, this last drug is essentially devoid of opioid activity.

Another line of research explored the therapeutic implications of the observation that GABAergic neurons modulate glutamatergic neurotransmission. Demonstration that a short course of intravenous treatment with the neurosteroid brexanolone, an allosteric modulator of GABA-A, could produce rapid antidepressant effects in women with postpartum depression quickly led to identification of a closely related compound, zuranolone, suitable for oral administration. Importantly, though the original discovery plan of these compounds was directed at postpartum depression, it was quickly recognized that this mechanism of action was relevant to treatment of depression in both men and women[7]. Of additional interest is the possibility that these treatments are suitable for intermittent or periodic treatment.

Interestingly, whereas the antidepressant effects of the treatments reviewed above appear to be unrelated to their potentially intoxicating or psychotomimetic effects, the fact that ketamine is a controlled substance may have helped open the door to reexamination of the therapeutic potential of hallucinogens such as psilocybin[8]. In this case, the intensity of the "psychedelic" experience is thought to be essential to the antidepressant effect, as is the belief – on clinical/experiential grounds – that the "trip" should be carefully guided to maximize the clinical benefit. As few safety concerns have emerged to date from phase 2 and early phase 3 studies of psilocybin, it may be that the field will need to wait until post-marketing for more rigorous studies to examine the amount and content of the adjunctive psychotherapeutic support necessary for an optimal result.

In contrast to developments in schizophrenia and major depressive disorder, the authors were unable to identify any drugs currently in development for either acute treatment of mania or prophylaxis of bipolar disorder. Of course, it is almost axiomatic that, once a compound has established efficacy for treatment of acute schizophrenia, interest in its use in mania will follow. Moreover, they identified no compounds in phase 3 for treatment of bipolar depression. That said, the regulatory pathway of lurasidone and, more recently, lumetaperone illustrates that drugs such as KarXT and ulotaront may hold promise for people with bipolar depression, as might drugs such as zuranolone and esmethadone.

The second segment of Correll et al's paper provides an excellent summary of some of the most recent strategies used to improve signal detection in clinical trials. As diagnostic heterogeneity, imprecision of measurement, and various factors that inflate the impact of placebo-expectancy effects on RCT outcomes, will continue to be a way of life for researchers for the foreseeable future, it is wise to incorporate as many of the authors' recommendations as practicable in the next generation of research.

I believe that our best hope for improved signal detection is the establishment of networks of rigorously trained and monitored investigators working together with access to populations of "real-world patients", in a manner analogous to the way that our peers working in cancer treatment have collaborated for the past few decades.

Michael E. Thase
Perelman School of Medicine, University of Pennsylvania, Corporal M.J. Crescenz Veterans Affairs Medical Center, Philadelphia, PA, USA

1. Correll CU, Solmi M, Cortese S et al. World Psychiatry 2023;22:48-74.
2. Brannan SK, Sawchak S, Miller AC et al. N Engl J Med 2021;384:717-26.

3. Koblan KS, Kent J, Hopkins SC et al. N Engl J Med 2020;382:1497-506.
4. Thase ME. CNS Spectr 2017;22:39-48.
5. Tabuteau H, Jones A, Anderson A et al. Am J Psychiatry 2022; doi: 10.1176/appi.ajp.21080800.
6. Fava M, Stahl S, Pani L et al. Am J Psychiatry 2021; 179:122-31.
7. Gunduz-Bruce H, Silber C, Kaul I et al. N Engl J Med 2019;381:903-11.
8. Davis AK, Barrett FS, May DG et al. JAMA Psychiatry 2021;78:481-9.

DOI:10.1002/wps.21064

The future of psychopharmacology: challenges beyond efficacy and tolerability

The paper by Correll et al[1] provides a comprehensive and timely overview of recent developments in psychopharmacology and offers hope for much needed breakthroughs after a period of stagnation in the field. It also considers some of the major challenges slowing further progress, including our limited understanding of the neurobiological underpinnings of psychiatric disorders and the difficulties encountered in designing and conducting trials that are able to adequately assess treatment effects on thoughts, emotions and behaviour.

From the first serendipitous discoveries of compounds with psychotropic effects to modern-day targeted drug development, advances over the years have been considerable, to the extent that we now have agents with at least some beneficial effect for most psychiatric disorders. In particular, the introduction of fluoxetine more than three decades ago heralded an era of drug design aimed at specific neurotransmitter pathways, with an upsurge of interest in psychopharmacology by the pharmaceutical industry, clinicians and the general public. Numerous new agents were introduced, and their therapeutic indications broadened.

These new treatments have not only strengthened the armamentarium of clinicians, but also fundamentally transformed our conceptualization of psychiatric disorders[2]. Consequently, the role of the psychiatrist has changed, and medication management has become a central function of clinical care. As such, an extensive knowledge of psychopharmacology is now a prerequisite for practicing psychiatrists. The danger here, of course, is that the other essential components of clinical care are neglected, and we become regarded as little more than "pill pushers". The challenge, particularly in busy clinical settings, is to balance medication management with a patient-centred approach to care, in order to establish the best possible therapeutic alliance, which in turn enhances treatment engagement and medication adherence[3].

After the initial euphoria accompanying the Prozac era came the realization that our expectations had been unrealistic. The newer generation of psychotropic drugs displayed at best only subtle efficacy advantages over their predecessors, and, while the novel pharmacological profiles effectively addressed adverse effects of the older agents, a new set of tolerability and safety concerns emerged. Over the past two decades, there has been a steady decline in the number of new psychotropic drugs introduced, mainly due to market saturation, escalating costs and the influx of generics[4].

We have witnessed a substantial waning of enthusiasm, and many of the pharmaceutical companies have withdrawn from psychotropic drug development. However, this has also forced those of us in the field to re-think our approach – to target novel mechanisms and to design clinical trials in a way that they are more likely to demonstrate efficacy advantages. Consequently, several promising new agents have progressed to the stage of clinical development, as highlighted in Correll et al's paper. Hopefully, some of these agents will be introduced to clinical practice in the near future, with the potential of not only providing us with more and better options for treating our patients, but also to advance our understanding of the pathophysiology of these disorders.

There are important considerations in the pharmacological treatment of psychiatric disorders that go beyond the efficacy and tolerability of the compounds. In order to be effective, most pharmacological interventions for psychiatric disorders require continuous treatment over a protracted period. As is the case with all chronic treatments, non-adherence is a major consideration[5]. However, with many psychiatric disorders, the problem is further compounded by impairment of insight. This is the case particularly with psychotic disorders and cognitive disorders. In psychosis, insight impairment is characterized by illness unawareness and failure to recognize the need for treatment. These features have enormous implications when considering treatment options and in shared decision-making processes. In such cases, the burden of responsibility for ensuring adherence to treatment should not be left with the patient. This aspect has been recognized by some pharmaceutical companies, which have invested much effort into the development of ways of providing treatment that are more likely than oral medications to provide assured, uninterrupted delivery. In this regard, long-acting injectable formulations have received the most attention.

There are also ethical and philosophical considerations in relation to the ongoing development of new psychopharmacological agents. It could be argued that the costs of developing new and better agents are not justified if they are inaccessible to the majority of individuals who would benefit from their use. This is increasingly the case, and not just in low- and middle-income countries. Even in more developed settings, the exorbitant costs of some newer psychotropic drugs have placed them beyond the reach of many.

On the other hand, the alternative ethical argument is that the best available treatments should be made accessible to all. Indeed, as stipulated in the constitution of the World Health Organization, access to the highest attainable standard of health care is one of the fundamental rights of every human being[6]. Unfortunately, in the

real world, this is not the case, particularly for mental health[7]. Across vast populations, mental health literacy is rudimentary and health care services inaccessible. The enormous treatment gap in these settings is surely an indictment of modern global health care. So, rather than questioning the need for psychotropic drug development, we should be encouraging those who continue to search for new and better agents – but at the same time we should be championing for their greater availability to those in need.

Robin Emsley
Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

1. Correll CU, Solmi M, Cortese S et al. World Psychiatry 2023;22:48-74.
2. Baldessarini RJ. Can J Psychiatry 2014;59:401-5.
3. Thompson L, McCabe R. BMC Psychiatry 2012; 12:87.
4. Braslow JT, Marder SR. Annu Rev Clin Psychol 2019;15:25-50.
5. Cheen MHH, Tan YZ, Oh LF et al. Int J Clin Pract 2019;73:e1335073.
6. Adhanom Ghebreyesus T. Health is a fundamental human right. https://www.who.int.
7. Raviola G, Naslund JA, Smith SL et al. Curr Psychiatry Rep 2019;21:1-13.

DOI:10.1002/wps.21065

Clinical trials of novel psychotropic agents: some caveats

In their paper, Correll et al[1] present proposals for strategies to "de-risk" trials of novel psychotropic agents. However, several of their suggestions may inadvertently increase the risk that clinical trials be uninformative, especially when considering requirements for drug approval. Here we provide our perspective on their advice.

The authors begin their "de-risking" advice with some foundational concepts related to validity, power, and *a priori* hypothesis generation. They go on to discuss the importance of "clinical equipoise" in randomized controlled trials. This emphasis is reasonable. Without clinical equipoise, trials are vulnerable to bias and are more difficult to interpret. For example, the enthusiasm for psychedelic drug development from both the lay press and investigators may contribute to difficulties separating drug effect from expectation bias.

Correll et al subsequently offer suggestions for modifying trial designs in an attempt to avoid failed studies. One recommendation is to consider adaptive trial designs whereby the beginning of the trial informs its later stages. The US Food and Drug Administration (FDA) has published a guidance for industry on the use of adaptive trials[2]. Compared with a traditional clinical trial, patients enrolling at the start of an adaptive clinical trial may not have the same experience as those enrolling later in the trial (e.g., possible doses). This may lead to challenges in interpreting the trial results. Further, although adaptive trials may be designed to maximize the possibility of quickly detecting efficacy with limited enrollment, more subjects may still be needed to characterize safety. A positive adaptive trial may not translate into approval if there are safety signals that must be explored in larger or longer studies. Sponsors considering adaptive studies in phase 3 should discuss their plans with regulatory authorities before implementation.

Phase 2 is an important part of dose exploration. Correll et al suggest using adaptive trials to determine the maximum tolerated dose (MTD) and prevent "expensive and underpowered multi-armed studies". However, as they later acknowledge, there are challenges to using an MTD when a dose range may be required. The MTD as determined early in a study may not translate to the optimum dose, considering benefit-risk, as the study progresses.

A Phase 2 program examining several doses based on phase 1 data (e.g., receptor binding, tolerability) need not be adequately powered to demonstrate safety and efficacy for each arm. It is meant to inform a phase 3 program. Sponsors sometimes design phase 2 studies with characteristics of adequate and well-controlled investigations in hopes that a positive trial may be used to support a marketing application. However, if there are dosing, endpoint, population or safety issues, this approach may ultimately prove more costly.

The paper's discussion of precision medicine versus generalizability is important. We acknowledge that particular mechanisms of action may have benefits particularly applicable to subpopulations, and that enriched trials may improve the chance of detecting an efficacy signal. However, development programs should focus not on artificially narrowed populations, but on a population widely inclusive of those likely to receive benefit.

A reasonable starting point for separating promising subgroup effects from *post-hoc* artifact is biological plausibility. Although a collection of clinical characteristics could be representative of a biological construct, there is a public health interest in determining what that underlying construct is. The authors suggest that positive studies from an enriched population could lead to an approval for use of the drug in a subpopulation, with studies of a broader population deferred to post-approval. However, in the absence of a biologically plausible subgroup definition supported by strong scientific understanding, we do not support this approach. Sponsors should explore scientifically justified potential subgroups in phase 2, refer to the appropriate guidance[3], and discuss plans with regulatory authorities.

Placebo lead-in studies have often not met expectations in psychiatric disorders. Sequential parallel design remains an unproven alternative to traditional placebo lead-in strategies. As with adaptive trials in general, there are significant challenges in interpreting the results of such studies. There is not a standard method for analyzing the results of sequential parallel design studies, and employing such a design in phase 3 entails risk on the part of a sponsor. Sponsors considering sequential parallel design should discuss this with regulatory authorities.

Correll et al state that "FDA…[has] taken the position that to assess the efficacy of a new treatment for many mental disorders is not possible without a placebo-controlled design". This is not accurate[4,5]. The Code of Federal Regulations, Title 21 (section 314.126)[6] describes the characteristics of an

adequate and well-controlled clinical investigation, and specifically mentions other types of controls – such as active treatment concurrent control and no treatment concurrent control – in addition to placebo concurrent control. Placebo-controlled trials are often favored and chosen by sponsors because they typically produce the most readily interpretable results.

Regarding generalizability of clinical trial results, Correll et al note that many "real world" patients would not qualify for pharmaceutical trials because of comorbidities. Sponsors should be prepared to justify their exclusion criteria, focusing on comorbidities that are expected to complicate interpretation of the study or decrease the likelihood of detecting an effect (e.g., active substance use disorders). The paper suggests requiring post-marketing studies to examine drug efficacy in "real world" patients; however, the FDA does not have the statutory authority to require such studies[7].

Correll et al describe scenarios in which rapid recruitment may impact study quality. Baseline symptom inflation and diagnostic imprecision may speed recruitment but will also make demonstrating efficacy more difficult. Although small sites may be a source of heterogeneity, they may simply be recruiting judiciously. Therefore, we recommend caution regarding the suggestion to drop poorly recruiting sites early in the study.

We agree that some new technologies might have the *potential* to improve assessments; however, before incorporating novel assessments (e.g., digital endpoints), we recommend that sponsors submit supportive evidence that the technology is fit-for-purpose. For example, a computerized system for assessing patient speech may seem to be an improvement on established subjective clinician ratings. However, it is the subjective clinical ratings which would have been tied to dysfunction and prognosis. Unless the computerized system also reflects dysfunction and prognosis, it may not be fit-for-purpose. Additionally, sponsors should ensure that including technology does not discourage or prevent certain groups from enrollment or introduce unanticipated biases.

Sponsors should discuss novel statistical approaches with regulatory authorities prior to starting clinical trials. Regarding the suggestion to use an endpoint that reflects symptom course over time (rather than at discrete time points), this may or may not be acceptable for a given trial. Such averaged endpoints may reflect improvement at the start of a trial that is lost as the trial progresses, leading to questions about the durability of effect.

Before attempting something novel in a development program, sponsors should meet with regulatory authorities, which can often refer companies to pertinent published guidances, help think through regulatory requirements, and use experience from other programs to offer recommendations.

Bernard A. Fischer, Tiffany R. Farchione
Division of Psychiatry, Office of Neuroscience, Office of New Drugs, Center for Drug Research and Development, US Food and Drug Administration, Silver Spring, MD, USA

The views expressed in this commentary are those of the authors and do not necessarily represent the position of the US Food and Drug Administration, the US Department of Health and Human Services, or the US Government.

1. Correll CU, Solmi M, Cortese S et al. World Psychiatry 2023;22:48-74.
2. US Food and Drug Administration. Adaptive designs for clinical trials of drugs and biologics. www.fda.gov.
3. US Food and Drug Administration. Enrichment strategies for clinical trials to support approval of human drugs and biological products. www.fda.gov.
4. US Food and Drug Administration. Providing clinical evidence of effectiveness for human drug and biological products. www.fda.gov.
5. US Food and Drug Administration. Demonstrating substantial evidence of effectiveness for human drug and biological products. www.fda.gov.
6. US Food and Drug Administration. CFR - Code of Federal Regulations Title 21. www.fda.gov.
7. US Food and Drug Administration. Postmarketing requirements and commitments: introduction. www.fda.gov.

DOI:10.1002/wps.21067

Tough times never last too long: the future of psychopharmacology

The progress of psychopharmacology has witnessed very different scenarios over the past seven decades. Obviously, the greatest impact occurred with the introduction of the first effective medications, such as chlorpromazine, imipramine, lithium and benzodiazepines[1]. Further refinements based on a better understanding of the pharmacological mechanisms behind those serendipitous findings led to drugs that were friendlier in terms of tolerability. Now, hopefully, we may be entering a new era with more innovative and personalized therapies.

After three decades of "me-too" drugs, the business profits from that drug development model are now exhausted, and practically all those drugs have become generic. This has given an unprecedented push to the search of alternative targets and mechanisms of action. The paradox is that this is happening in the context of recent cuts in the investment of big pharma companies in neuroscience. However, smaller companies and bio-techs have taken over, and there is a bunch of promising novel drugs for the management of schizophrenia, depression, and stress-related disorders, as very well discussed by Correll et al[2].

The situation is somewhat less optimistic for bipolar disorder and addiction, where repurposing is the rule rather than the exception. Some of the promising agents for these indications will only get approved if they are successful for their primary indication, for example schizophrenia[3]. However, it has to be considered that, in many countries, there are no incentives for secondary indications (they require further investment in clinical trials and sometimes they imply price or reimbursement cuts that companies prefer to avoid).

No one knows at present time how many of the new drugs that are at late stages of development will reach the market, but there are good reasons to be optimistic that at least a few will make it and may be available to patients with mental disorders soon. In schizophrenia, the new mechanisms not involving dopamine antagonism or modulation may provide opportunities to non-responders to the traditional treatments, and to tackle

orphan dimensions such as negative symptoms. In depression, practically all novel therapies have in common a fast onset of action, which may save lives by reducing suicide risk and improve the quality of life of patients since treatment start, especially for those in whom the conventional treatments failed. New drugs, combined with some particular forms of adjunctive psychotherapy, may make a difference for those suffering from post-traumatic stress disorder.

Further aspects that may foster optimism are the progress associated with the classification of psychopharmacologic agents[4], and the focus on transdiagnostic targets[5], such as emotional dysregulation and cognitive impairment. Finally, advances in the implementation of precision psychiatry[6] may provide further opportunities to explore biomarker-based targets rather than traditional clinical endpoints.

Nevertheless, some hurdles are still there. An obvious one is the increasing difficulties in signal detection with placebo-controlled trials[7] and the limited alternatives to placebo-controlled designs[8], as well as problems related to the representativeness of the patients enrolled in those trials and the generalizability of the findings[9]. Regulatory agencies are not consistent across the world in their requirements for marketing approval of medications, and this carries increased costs and inequalities. The stigma associated with psychiatric conditions is still a major cause of shortage of investments in research as compared to other areas of medicine, despite the huge prevalence of these disorders.

Precision psychiatry will hopefully evolve over the present decade, but will likely pose novel challenges. Health care access is still an issue in many parts of the world, and this is particularly true for mental illness. The benefits of precision psychiatry and novel treatments, with their associated increased costs, may not be available for all, and cause further inequities. Given the high prevalence of psychiatric disorders, governments will likely face huge budget and reimbursement challenges as diagnostic and therapeutic progress makes the care of the mentally ill increasingly expensive.

I am not particularly confident that there will be a perfect correlation between biomarkers and deep clinical phenotyping in psychiatry, although there is plenty of room for improvement in performing thorough psychopathological assessments in large samples of patients and including that information in the current clinically poor datasets of big consortia of genetics (e.g., Psychiatric Genetics Consortium) and neuroimaging (e.g., ENIGMA). But even if so-called "molecular psychopathology" ends up being too unspecific, there is hope that future biomarkers may be better correlated with functioning, making their use fruitful as relevant treatment targets. The rise of digital tools may be instrumental in this regard, yielding objective behavioral data for the assessment and monitoring of personalized outcomes. This would be relevant not only for clinical trial design, but also for clinical practice.

The future of psychopharmacology depends on this, but also on establishing synergies with other treatment modalities, such as neuromodulation and advanced psychotherapies. Hence investments, either from public or charity budgets, and ideally from both, are urgently needed in psychiatry and related disciplines. Large population datasets, covering the whole life span, need to be deeply studied with all the available relevant tools and technology, as defined by consensus of worldwide experts. This is the time to make a real step further, filling the gaps described by Correll et al[2], and pursuing better health and justice for the mentally ill.

Efforts in searching better diagnosis and treatment of psychiatric disorders should go hand in hand with better health care access, early intervention initiatives, prevention, and promotion of mental health in the general population. The future of psychopharmacology is unequivocally linked to the future of psychiatry as a discipline. The stigma associated to mental disorders and to pharmacological tools for the disorders of the brain is perhaps the greatest barrier to overcoming these tough times, which should not last too long.

Eduard Vieta
Institute of Neuroscience, Hospital Clinic, University of Barcelona, IDIBAPS, CIBERSAM, Barcelona, Catalonia, Spain

1. Vieta E. Rev Psiquiatr Salud Ment 2020;13:1-4.
2. Correll CU, Solmi M, Cortese S et al. World Psychiatry 2023;22:48-74.
3. Gimenez-Palomo A, Vieta E. Eur Neuropsychopharmacol 2022;62:4-6.
4. Zohar J, Levy DM. Eur Neuropsychopharmacol 2022;57:36-8.
5. Solmi M, Bodini L, Cocozza S et al. Eur Neuropsychopharmacol 2020;41:16-27.
6. Vieta E. Rev Psiquiatr Salud Ment 2015;8:117-8.
7. Huneke NTM. Eur Neuropsychopharmacol 2022; 62:7-9.
8. Similon MVM, Paasche C, Krol F et al. Eur Neuropsychopharmacol 2022;60:91-9.
9. Taipale H, Schneider-Thoma J, Pinzón-Espinosa J et al. JAMA Psychiatry 2022;79:210-8.

DOI:10.1002/wps.21066

Impact of mental disorders on clinical outcomes of physical diseases: an umbrella review assessing population attributable fraction and generalized impact fraction

Elena Dragioti[1,2], Joaquim Radua[3-5], Marco Solmi[3,6-9], Corentin J. Gosling[8,10,11], Dominic Oliver[3,12], Filippo Lascialfari[13], Muhammad Ahmed[3], Samuele Cortese[8,14-16], Andrés Estradé[3], Gonzalo Arrondo[8,17], Mary Gouva[2], Michele Fornaro[18], Agapi Batiridou[2], Konstantina Dimou[2], Dimitrios Tsartsalis[19], Andre F. Carvalho[20], Jae Il Shin[21,22], Michael Berk[20], Silvia Stringhini[23-25], Christoph U. Correll[9,26-28], Paolo Fusar-Poli[3,13,29]

[1]Pain and Rehabilitation Centre and Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden; [2]Research Laboratory Psychology of Patients, Families and Health Professionals, School of Health Sciences, University of Ioannina, Ioannina, Greece; [3]Early Psychosis: Interventions and Clinical-detection (EPIC) Lab, Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK; [4]Imaging of Mood- and Anxiety-Related Disorders Group, Institut d'Investigacions Biomèdiques August Pi i Sunyer, CIBERSAM, University of Barcelona, Barcelona, Spain; [5]Department of Clinical Neuroscience, Centre for Psychiatric Research and Education, Karolinska Institutet, Stockholm, Sweden; [6]Department of Psychiatry, University of Ottawa, Ottawa, ON, Canada; [7]Department of Mental Health, Ottawa Hospital, Ottawa, ON, Canada; [8]Centre for Innovation in Mental Health, School of Psychology, Faculty of Environmental and Life Sciences, University of Southampton, Southampton, UK; [9]Department of Child and Adolescent Psychiatry, Charité Universitätsmedizin, Berlin, Germany; [10]DysCo Lab, Paris Nanterre University, Nanterre, France; [11]Laboratoire de Psychopathologie et Processus de Santé, Université Paris Cité, Boulogne-Billancourt, France; [12]Department of Psychiatry, University of Oxford, Oxford, UK; [13]Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy; [14]Clinical and Experimental Sciences (CNS and Psychiatry), Faculty of Medicine, University of Southampton, and Solent NHS Trust, Southampton, UK; [15]Division of Psychiatry and Applied Psychology, School of Medicine, University of Nottingham, Nottingham, UK; [16]Hassenfeld Children's Hospital at NYU Langone, New York, NY, USA; [17]Mind-Brain Group, Institute for Culture and Society, University of Navarra, Pamplona, Spain; [18]Section of Psychiatry, Department of Neuroscience, Reproductive Science and Odontostomatology, Federico II University of Naples, Naples, Italy; [19]Department of Emergency Medicine, Hippokration Hospital, Athens, Greece; [20]Institute for Mental and Physical Health and Clinical Translation (IMPACT), School of Medicine and Barwon Health, Deakin University, Geelong, VIC, Australia; [21]Department of Pediatrics, Yonsei University College of Medicine, Seoul, South Korea; [22]Department of Pediatrics, Severance Children's Hospital, Seoul, South Korea; [23]Division of Primary Care, Geneva University Hospitals, Geneva, Switzerland; [24]University Centre for General Medicine and Public Health, University of Lausanne, Lausanne, Switzerland; [25]Department of Health and Community Medicine, University of Geneva, Geneva, Switzerland; [26]Department of Psychiatry, Zucker Hillside Hospital, Northwell Health, Glen Oaks, NY, USA; [27]Department of Psychiatry and Molecular Medicine, Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY, USA; [28]Center for Psychiatric Neuroscience, Feinstein Institute for Medical Research, Manhasset, NY, USA; [29]OASIS Service, South London and Maudsley NHS Foundation Trust, London, UK

Empirical evidence indicates a significant bidirectional association between mental disorders and physical diseases, but the prospective impact of mental disorders on clinical outcomes of physical diseases has not been comprehensively outlined. In this PRISMA- and COSMOS-E-compliant umbrella review, we searched PubMed, PsycINFO, Embase, and Joanna Briggs Institute Database of Systematic Reviews and Implementation Reports, up to March 15, 2022, to identify systematic reviews with meta-analysis that examined the prospective association between any mental disorder and clinical outcomes of physical diseases. Primary outcomes were disease-specific mortality and all-cause mortality. Secondary outcomes were disease-specific incidence, functioning and/or disability, symptom severity, quality of life, recurrence or progression, major cardiac events, and treatment-related outcomes. Additional inclusion criteria were further applied to primary studies. Random effect models were employed, along with I^2 statistic, 95% prediction intervals, small-study effects test, excess significance bias test, and risk of bias (ROBIS) assessment. Associations were classified into five credibility classes of evidence (I to IV and non-significant) according to established criteria, complemented by sensitivity and subgroup analyses to examine the robustness of the main analysis. Statistical analysis was performed using a new package for conducting umbrella reviews (https://metaumbrella.org). Population attributable fraction (PAF) and generalized impact fraction (GIF) were then calculated for class I-III associations. Forty-seven systematic reviews with meta-analysis, encompassing 251 non-overlapping primary studies and reporting 74 associations, were included (68% were at low risk of bias at the ROBIS assessment). Altogether, 43 primary outcomes (disease-specific mortality: n=17; all-cause mortality: n=26) and 31 secondary outcomes were investigated. Although 72% of associations were statistically significant (p<0.05), only two showed convincing (class I) evidence: that between depressive disorders and all-cause mortality in patients with heart failure (hazard ratio, HR=1.44, 95% CI: 1.26-1.65), and that between schizophrenia and cardiovascular mortality in patients with cardiovascular diseases (risk ratio, RR=1.54, 95% CI: 1.36-1.75). Six associations showed highly suggestive (class II) evidence: those between depressive disorders and all-cause mortality in patients with diabetes mellitus (HR=2.84, 95% CI: 2.00-4.03) and with kidney failure (HR=1.41, 95% CI: 1.31-1.51); that between depressive disorders and major cardiac events in patients with myocardial infarction (odds ratio, OR=1.52, 95% CI: 1.36-1.70); that between depressive disorders and dementia in patients with diabetes mellitus (HR=2.11, 95% CI: 1.77-2.52); that between alcohol use disorder and decompensated liver cirrhosis in patients with hepatitis C (RR=3.15, 95% CI: 2.87-3.46); and that between schizophrenia and cancer mortality in patients with cancer (standardized mean ratio, SMR=1.74, 95% CI: 1.41-2.15). Sensitivity/subgroup analyses confirmed these results. The largest PAFs were 30.56% (95% CI: 27.67-33.49) for alcohol use disorder and decompensated liver cirrhosis in patients with hepatitis C, 26.81% (95% CI: 16.61-37.67) for depressive disorders and all-cause mortality in patients with diabetes mellitus, 13.68% (95% CI: 9.87-17.58) for depressive disorders and major cardiac events in patients with myocardial infarction, 11.99% (95% CI: 8.29-15.84) for schizophrenia and cardiovascular mortality in patients with cardiovascular diseases, and 11.59% (95% CI: 9.09-14.14) for depressive disorders and all-cause mortality in patients with kidney failure. The GIFs confirmed the preventive capacity of these associations. This umbrella review demonstrates that mental disorders increase the risk of a poor clinical outcome in several physical diseases. Prevention targeting mental disorders – particularly alcohol use disorders, depressive disorders, and schizophrenia – can reduce the incidence of adverse clinical outcomes in people with physical diseases. These findings can inform clinical practice and trans-speciality preventive approaches cutting across psychiatric and somatic medicine.

Key words: Mental disorders, physical diseases, outcomes, disease-specific mortality, all-cause mortality, trans-speciality prevention

(World Psychiatry 2023;22:86–104)

Both physical diseases and mental disorders contribute significantly to the increasing burden on health care systems worldwide[1,2]. Cardiovascular diseases, cancer, chronic respiratory diseases, and diabetes are accountable for more than 50% of global deaths[1], while mental disorders are the third leading cause of disease burden, with depressive disorders accounting for 37% of all

years of life lost to disability, followed by anxiety disorders (23%) and schizophrenia (12%)[2].

The Cartesian dichotomy of mental disorder-physical disease is challenged by empirical evidence from primary studies[3], meta-analyses[3-7], and umbrella reviews[8,9] showing significant prospective associations between the two realms. For instance, individuals with schizophrenia, compared to the general population, have a higher incidence of metabolic and cardiovascular diseases and of cancer[10-13]; those with mood disorders are at higher risk of developing cancer and diabetes mellitus[7,14]; and those with borderline personality disorder have a higher risk to develop a gastrointestinal disease, arthritis and chronic pain. Moreover, mental disorders have been found to increase the burden of physical diseases[10,15,16].

Neurobiologically, the core mechanisms that are likely to drive the neuroprogression of mental disorders – such as inflammation, oxidative stress, apoptosis, and mitochondrial dysfunction – overlap with the mechanisms driving somatoprogression[17]. Moreover, mental disorders interfere with adherence to healthy behaviors and treatment[18]. Consequently, the occurrence of mental disorders often worsens the prognosis of physical diseases. For example, depressive and anxiety disorders are associated with a higher mortality risk in people with cancer[19,20], cardiovascular diseases[21,22], chronic obstructive pulmonary disease[23], and diabetes mellitus[24,25]. The recent COVID-19 pandemic has also indicated that mental disorders are associated with higher disease severity and mortality[26-28].

Despite this accumulating evidence, studies concerning the impact of mental disorders on clinical outcomes of physical diseases are often restricted to small sets of associations, sometimes with conflicting results, and therefore hold limited clinical relevance[9]. Relevant confounders, such as differences in diagnostic methods, the timing of the diagnosis of mental disorders[9] and the effect of psychiatric medications[12], have not been systematically controlled for. Furthermore, the observed associations have generally not been appraised using established classification criteria to grade the credibility of the evidence and control for several types of biases.

Another limitation is that the reported associations are not directly informative for clinical practice. For example, it is unclear to what extent preventive approaches for mental disorders could reduce the incidence of clinical outcomes of physical diseases. To address this question, it is essential to quantify the proportional reduction of disease that would occur if a given risk factor is eliminated (population attributable fraction, PAF)[29], or partially reduced (generalized impact fraction, GIF)[30-32], in a specific population. To our knowledge, no study has estimated the meta-analytic PAF or GIF of the most robust associations between mental disorders and clinical outcomes in patients with physical diseases.

This is the first umbrella review comprehensively summarizing the evidence concerning the prospective impact of mental disorders on clinical outcomes of physical diseases using established classification criteria of evidence that address multiple biases[33-35], controlling for relevant confounders, and estimating the related meta-analytic PAF and GIF. Providing a solid and rigorous synthesis of this evidence is crucial to promote sound etiopathological research and to implement effective preventive strategies cutting across psychiatry and somatic medicine[36].

METHODS

We followed the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) 2020 statement[37] and the Conducting Systematic Reviews and Meta-Analyses of Observational Studies of Etiology (COSMOS-E) guidelines[38]. The study protocol is available at the Center for Open Science (https://osf.io/dt4fu).

Search strategy and selection criteria

We systematically searched PubMed, PsycINFO, Embase, and Joanna Briggs Institute Database of Systematic Reviews and Implementation Reports from inception to March 15, 2022, to identify systematic reviews with meta-analysis that examined the prospective association between any mental disorder and clinical outcomes of physical diseases. Primary outcomes were disease-specific and all-cause mortality. Secondary outcomes were disease-specific incidence, functioning and/or disability, symptom severity, quality of life, recurrence and progression, major cardiac events, and treatment-related outcomes.

Categories of mental disorders were stratified according to the corresponding ICD-10 diagnostic blocks, in line with previous studies[39,40], and defined by standard diagnostic criteria or requirements (i.e., any version of the ICD or the DSM), or established diagnostic research criteria (e.g., Research Diagnostic Criteria[41]), or validated assessment instruments with cut-offs that map onto discrete ICD/DSM diagnoses (e.g., Patient Health Questionnaire[42]).

We focused on categories of physical diseases associated with the highest burden according to the 2019 Global Burden of Disease Study[1] and other recent studies[11]: cardiovascular diseases (e.g., coronary heart disease), chronic respiratory diseases (e.g., chronic obstructive pulmonary disease), neurological diseases (e.g., multiple sclerosis), nutritional and metabolic diseases (e.g., obesity), endocrine system diseases (e.g., diabetes mellitus), kidney diseases, neoplasms, digestive diseases (e.g., liver cirrhosis), infectious diseases (e.g., human immunodeficiency virus, HIV infection), and musculoskeletal diseases (e.g., low back pain).

As a search strategy, we combined key terms and Medical Subject Headings (MeSH) terms related to these categories of mental disorders and physical diseases with terms related to the clinical outcomes of interest and to systematic reviews or meta-analyses (full details are described in supplementary information). The reference lists of the records identified during the screening process were also searched. Four independent investigators screened the records based on title and abstract reading. After excluding those that were not relevant, the full texts of the remaining records were further assessed for inclusion. Any discrepancy was solved through discussions with a fifth senior investigator.

We included: a) systematic reviews with meta-analysis of observational studies with a prospective design, with meta-analytic summary estimates derived from at least two primary studies; b) primarily investigating the association between mental disorders and clinical outcomes of physical diseases (defined as above); c) published in English.

We excluded: a) systematic reviews without meta-analysis; b) systematic reviews with meta-analysis of individual participant data or network meta-analysis; c) systematic reviews with meta-analysis of randomized controlled trials, interventions, study designs other than prospective (cross-sectional and retrospective case-control studies are subject to recall bias and reverse causality); d) meta-analyses of data not identified via systematic reviews; e) meta-analyses mixing mental disorders and physical diseases without providing distinguishable association measures; e) systematic reviews or meta-analyses using unclear diagnostic criteria not operationalized as above; f) fully overlapping datasets.

When two systematic reviews or meta-analyses presented overlapping data on the same association, only the one with the largest dataset in terms of number of primary studies was retained for the specific association (the two meta-analyses could be non-overlapping for other associations). In the case of similar datasets, we selected the meta-analysis with the highest study quality. When two meta-analyses presented minimally overlapping or not overlapping datasets, nevertheless still addressing the same association, both meta-analyses were included.

Additional inclusion/exclusion criteria were applied to each of the primary studies included in the systematic reviews. Primary study-level inclusion criteria were: a) prospective cohort or longitudinal study (if a meta-analysis included multiple study designs such as randomized controlled trials and prospective studies, we only retained prospective studies); b) examining longitudinally the impact of a mental disorder on clinical outcomes of a physical disease (defined as above); c) distinguishing study participants with a mental disorder (exposed) or not (unexposed) who develop (cases) or not (controls) at least one clinical outcome of a physical disease.

Primary study-level exclusion criteria were: a) studies investigating psychiatric symptoms only but not mental disorders; b) studies reporting on clinical outcomes only for mixed categories of mental or physical diseases (e.g., anxiety and depressive disorders, or diabetes and stroke), without distinguishable estimates per pair of disorders; c) studies using unclear diagnostic criteria not operationalized as above (e.g., continuous psychometric scales without established cut-offs to estimate categorical diagnoses); d) studies reporting on outcomes other than those of interest.

Risk of bias

Four independent investigators assessed the risk of bias in the included systematic reviews by using the Risk of Bias in Systematic Reviews (ROBIS) tool[43], which has shown good reliability and construct validity in systematic reviews[44]. Any discrepancy was solved through discussions with a fifth investigator.

The ROBIS tool is applied in three phases: 1) assess relevance (optional), 2) identify concerns with the review process, and 3) judge risk of bias in the review[43]. In this study, we employed phases 2 and 3. Phase 2 is divided into four domains. Domain 1 assesses concerns regarding the specification of study eligibility criteria; domain 2 evaluates any concerns regarding methods used to identify/select studies; domain 3 covers concerns regarding methods used to collect data and appraise studies; and domain 4 focuses on concerns regarding the synthesis of results. Phase 3 assesses the overall ROBIS risk of bias in the interpretation of review findings[43,45].

Data extraction

Data extraction was performed independently by three investigators and verified by a fourth investigator.

For each eligible systematic review, we extracted the standard identifier (PubMed identifier, PMID, or digital object identifier, DOI), the first author, the year and journal of publication, the number of prospective primary studies, and the specific populations evaluated. We also extracted the study-specific association measures (odds ratio, OR; risk ratio, RR; hazard ratio, HR; and standardized mortality ratio, SMR), with their 95% confidence intervals (CIs), or the indirect information needed to estimate the association measure.

For each primary study, we extracted the specific population, the number of cases (number of outcome events in participants with a mental disorder), the number of non-cases (number of outcome events in participants without mental disorders), the sample size, the method used to diagnose physical diseases, and the confounders to be tested in subgroup analyses – i.e., the method used to diagnose mental disorders, the timing of mental disorder diagnosis (before or after the diagnosis of a physical disorder), the type of estimates (fully/partially adjusted or unadjusted), the age and sex of participants, and the exposure to psychiatric medications.

For primary studies, we extracted in decreasing order of preference the fully adjusted estimates (e.g., controlling for all available covariates), the partially adjusted estimates (e.g., controlling only for age and sex or some of the covariates reported in the study) and the unadjusted estimates. Whenever studies used multiple control groups, we only considered data from participants without a mental disorder (non-exposed).

We also recorded the quality score of the primary studies and the scale used (when reported) to assess quality; otherwise, we rated the study with the Newcastle-Ottawa scale (NOS)[46].

Statistical analysis

The main effect size of interest was the prospective association between mental disorders and clinical outcomes of physical diseases, indexed by the meta-analytic OR, RR, HR or SMR measures and eventually converted into equivalent odds ratios (eORs)[33] for comparative purposes. The direction of the effect sizes was harmonized[47]: an eOR greater than 1 indexed an increased likelihood of the outcome, while an eOR less than 1 indexed a decreased

likelihood of the outcome.

Whenever studies provided effect sizes for independent subgroups (e.g., they presented effect sizes for males and females separately), we pooled them using the Borenstein method[48]. When multiple outcomes (e.g., all-cause mortality and cardiovascular mortality) were assessed in the same primary study, we estimated a pooled effect size[10], assuming a correlation of 0.8 between outcomes[49,50].

Random effects models with the restricted maximum likelihood (REML) variance estimator were employed[50]. The I^2 statistic was computed to evaluate inconsistency (I^2>50% indicated high inconsistency)[51], together with the 95% prediction intervals to estimate the plausible range in which the effect sizes of future studies are expected to fall[52]. The presence of small-study effects was tested with Egger's regression asymmetry test ($p \leq 0.05$[53]).

The presence of excess significance bias was calculated by using the new Test for Excess Statistical Significance (TESS) and the Proportion of Statistical Significance Test (PSST)[54]. Both TESS and PSST have desirable statistical properties: adequate control of Type I errors and high statistical power, which takes inconsistency into account[54]. The presence of excess significance bias was assumed if either TESS or PSST was greater than the Z-score of 1.645[54].

Associations were classified into five levels of evidence according to established classification criteria[9,33-35,55]: convincing (class I: N>1,000 cases, $p<10^{-6}$, no evidence of small-study effects or excess significance bias, 95% prediction interval not including the null, and no large inconsistency); highly suggestive (class II: N>1,000 cases, $p<10^{-6}$, largest study with a statistically significant effect, and class I criteria not met); suggestive (class III: N>1,000 cases, $p<10^{-3}$, and class I and II criteria not met); weak (class IV: all other associations with $p \leq 0.05$); and non-significant (NS: all associations with p>0.05).

A sensitivity analysis was performed by removing the criterion of N>1,000 cases to examine the robustness of the main analysis when smaller numbers of cases were included[56]. Subgroup analyses were also performed for associations supported by class I/II evidence to test confounders identified at the primary study level. We stratified the analyses by: a) diagnostic method (standard diagnostic criteria vs. research criteria vs. validated assessment instruments with cut-offs that map onto discrete categories); b) timing of mental diagnosis (diagnosis of mental disorder confirmed before or after the diagnosis of physical disease); c) follow-up duration (>5 vs. ≤5 years); d) type of estimates (adjusted vs. unadjusted); e) age of participants (<50 vs. ≥50 years old); f) exposure to psychiatric medications (yes/no); and g) sex (majority of males vs. majority of females).

The PAF analysis was conducted for each class I-III association, following a method previously established[57]. Prevalence data (± 95% CIs) of mental disorders in physical diseases were extracted from the primary studies as the total number of those exposed and those in the total population of interest (e.g., the population of patients with cardiovascular diseases). The calculation of the PAF was based on Levin's formula[58], which requires the RR estimate and the prevalence of the risk factor[59]. We converted all ORs to RRs using a standard formula[60]. 95% CIs for the PAFs were derived using a method previously validated[40]. For each association, we created 50,000 random RRs according to the RR 95% CI and 50,000 random prevalences according to the prevalence 95% CI. We then combined the random RRs and prevalences to derive 50,000 PAF estimations, from which we derived the PAF 95% CI.

While the PAF assumes a perfect intervention that fully eradicates the risk factor (i.e., 100% reduction of its prevalence)[61], such a complete removal is usually unrealistic. We thus performed additional analyses by computing the GIF for factors with the largest PAFs (since the GIF is ≤PAF, the GIF analysis would be futile for smaller PAFs). The GIF estimates the proportional reduction in disease incidence given a graded reduction in the prevalence of a risk factor[61].

All analyses were performed in R software, version 4.1.2, using a new evidence synthesis package developed to conduct umbrella reviews: the metaumbrella package[50,62], also available as a browser-based graphical app (https://metaumbrella.org).

RESULTS

Database search results

The search identified 21,612 potentially relevant records, and 18,610 titles/abstracts were screened after duplicate removal (see Figure 1). Altogether, 551 full-text papers were checked for eligibility, and 47 systematic reviews with meta-analysis were eventually included in the umbrella review[13,19,20,22-24,26,63-102].

The systematic reviews were published between 2004 and 2022, including a total of 251 non-overlapping primary (prospective) studies. They reported on 43 primary outcomes (disease-specific mortality: n=17; all-cause mortality: n=26) and 31 secondary outcomes (disease-specific incidence: n=6; disease-specific functioning and/or disability: n=1; disease-specific symptom severity: n=7; disease-specific recurrence or progression: n=8; major cardiac events: n=7; and treatment-related outcomes: n=2). No disease-specific quality of life outcome was reported.

The total number of participants included in each systematic review ranged from 159[75] to 11,309,529[13] (median: 3,717, interquartile range, IQR: 1,154-22,786). The participants' age ranged from 17[72,85] to 99 years[97], and all but one systematic review[20] included both males and females. The number of primary (prospective) studies included in each systematic review ranged from 2[73,75,78,95,99] to 27[76] (median: 5, IQR: 3-8); their follow-up duration ranged from three[79] to 29 years[86]. About 79% of the primary studies in each systematic review were of high quality.

Most (n=38, 81%) systematic reviews examined associations between mood or anxiety disorders and clinical outcomes of physical diseases: 30 (63.8%) studied the associations of mood disorders[19,24,63-65,67,68,71-74,81,82,84-94,96-100,102], and five (10.8%) the associations of anxiety disorders[22,66,70,77,95], mostly with outcomes of cardiovascular, neoplastic, endocrine, infectious, neurological or respiratory diseases. Three studies (6.4%) investigated the associations of both anxiety and mood disorders with outcomes of neoplastic, neurological and respiratory diseases[20,23,78].

The other diagnostic blocks were less investigated. Four sys-

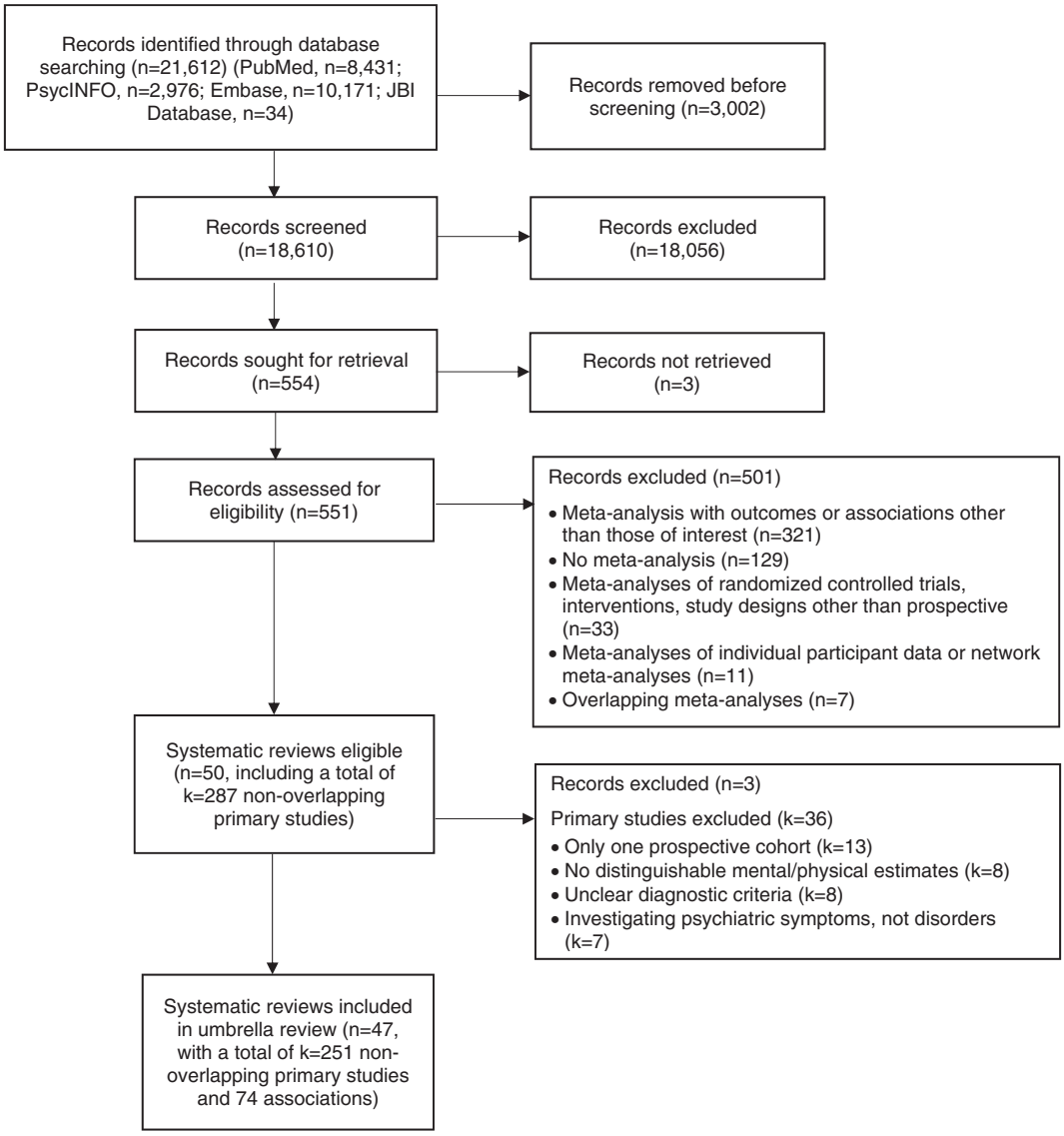

Figure 1 PRISMA flow chart, JBI - Joanna Briggs Institute

tematic reviews (8.5%) studied organic, including symptomatic, mental disorders in relation to outcomes of cardiovascular, infectious or neurological diseases[26,69,76,79]. Two (4.2%) studied schizophrenia with regard to outcomes of neoplastic diseases[83,101]; one (2.1%) studied both mood disorders and schizophrenia in relation to outcomes of cardiovascular diseases[13]; one (2.1%) studied alcohol use disorders in regard to outcomes of liver diseases[75]; and one (2.1%) separately studied anxiety disorders, depressive disorders and Alzheimer's disease in relation to outcomes of a neurological disease[75].

More than half (n=30, 63.8%) of the systematic reviews ascertained mental disorders using a combination of standard diagnostic criteria or requirements (DSM/ICD), research criteria and validated assessment measures with established cut-offs that map onto ICD/DSM diagnoses. Eleven (23.5%) ascertained mental disorders using exclusively the third of the above-mentioned approaches[20,22,63,74,75,87,89,93,95,96,102]. Only six (12.7%) used standard diagnostic criteria or requirements (any version of DSM or ICD) exclusively[26,76,80,83,99,101] (for details, see supplementary information).

There were no systematic reviews with meta-analysis examining the impact of mental disorders from the other ICD-10 diagnostic blocks on clinical outcomes of physical diseases.

Risk of bias

An overall summary of the ROBIS assessment of the systematic reviews is provided in the supplementary information. A total of 26 (55.3%) reviews were at low risk of bias across all phase 2 domains. In Phase 2, 35 (74.5%) systematic reviews had a low risk of bias in domain 1, 34 (72.3%) in domain 2, 26 (55.3%) in domain 3, and 31 (66%) in domain 4. A total of 32 (68.1%) systematic re-

views were rated as at low risk of bias in phase 3, which indexes the overall ROBIS risk of bias[43,45].

Summary of associations

A total of 74 associations were analyzed. Fifty-three (71.6%) presented a statistically significant effect ($p<0.05$), but only 15 of those (28.3%) reached $p<10^{-6}$. The number of cases was greater than 1,000 for 30 associations (40.5%). Twenty-eight associations (37.8%) presented large inconsistency ($I^2>50\%$), while for 12 (16.2%) the 95% prediction interval did not include the null hypothesis. Additionally, the evidence for small-study effects was noted for nine associations (12.1%), and excess significance bias was noted for 19 (25.6%) associations.

The summary of the associations for classes I-IV is shown in Figures 2 and 3. Only two associations (2.7%) showed a convincing level of evidence (class I), and six (8.1%) showed highly suggestive evidence (class II). Of the remaining associations, three (4.1%) showed suggestive evidence (class III), 42 (56.7%) weak evidence (class IV), and 21 (28.4%) had no evidence. In the following sections, we primarily describe the associations with the highest classes (I-III) of evidence.

Associations of neurotic, stress-related and somatoform disorders with clinical outcomes of physical diseases

None of the 13 associations in this diagnostic block was supported by convincing or highly suggestive evidence (class I and II) for either primary or secondary outcomes. Only the association between anxiety disorders and cardiovascular mortality in patients with cardiovascular diseases (RR=1.46, 95% CI: 1.17-1.82) presented a suggestive evidence level (class III). There was weak evidence (class IV) for four associations concerning secondary outcomes. No evidence was found for the remaining eight associations concerning primary and secondary outcomes (see Figures 2 and 3, Table 1 and supplementary information).

After removing the N>1,000 cases criterion in sensitivity analysis, the two associations between anxiety disorders and major cardiac events were upgraded from weak (class IV) to suggestive evidence (class III). The level of evidence of the other associations remained unchanged (see Table 1 and supplementary information).

Associations of mood disorders with clinical outcomes of physical diseases

Among the 49 associations in this diagnostic block, only that between depressive disorders and all-cause mortality among patients with heart failure (HR=1.44, 95% CI: 1.26-1.65) presented a convincing level of association (class I) (see Figure 2 and Table 2).

Highly suggestive evidence (class II) was found for associations between depressive disorders and all-cause mortality in patients with kidney failure (HR=1.41, 95% CI: 1.31-1.51) and in those with diabetes mellitus (HR=2.84, 95% CI: 2.00-4.03); for the association between depressive disorders and major cardiac events in patients with myocardial infarction (OR=1.52, 95% CI: 1.36-1.70); and for the association between depressive disorders and dementia in patients with diabetes mellitus (HR=2.11, 95% CI: 1.77-2.52) (see Figure 2, Table 2 and supplementary information).

There was suggestive evidence (class III) for two associations: that between bipolar disorder and cardiovascular mortality in patients with cardiovascular diseases (RR=1.65, 95% CI: 1.32-2.06),

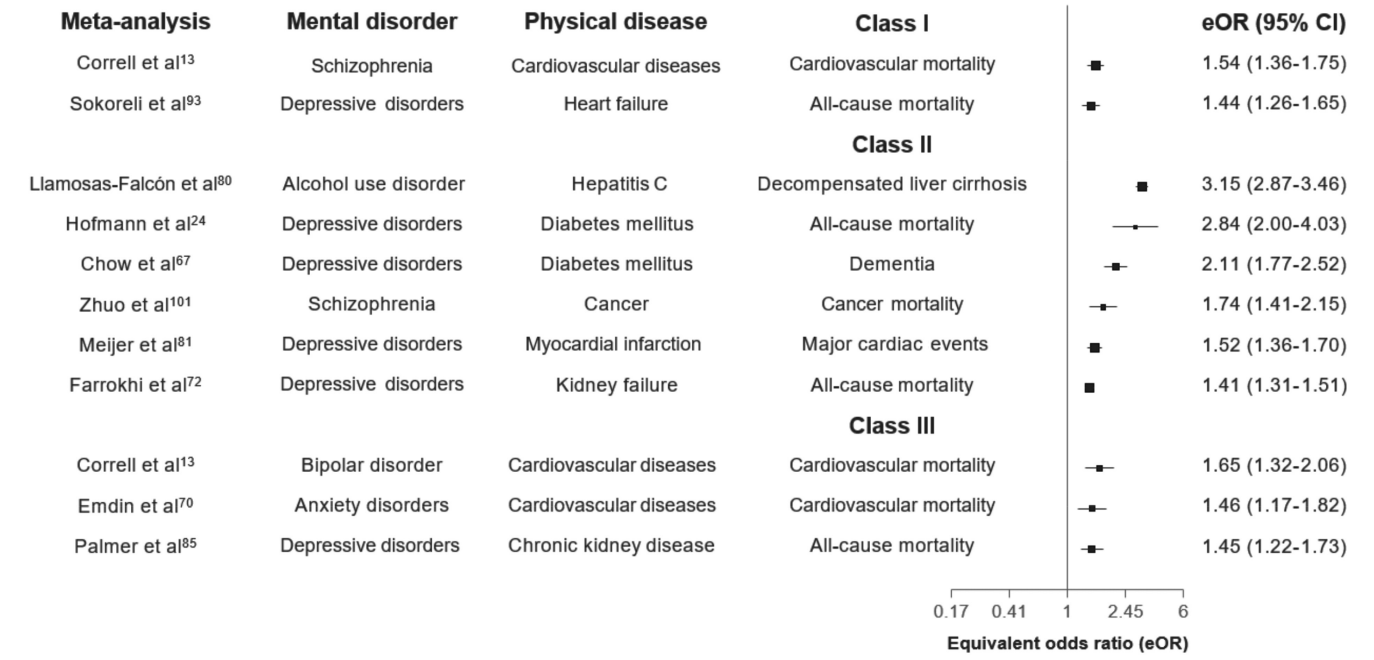

Figure 2 Forest plot of prospective associations between mental disorders and clinical outcomes of physical diseases, stratified by class I, II and III of evidence

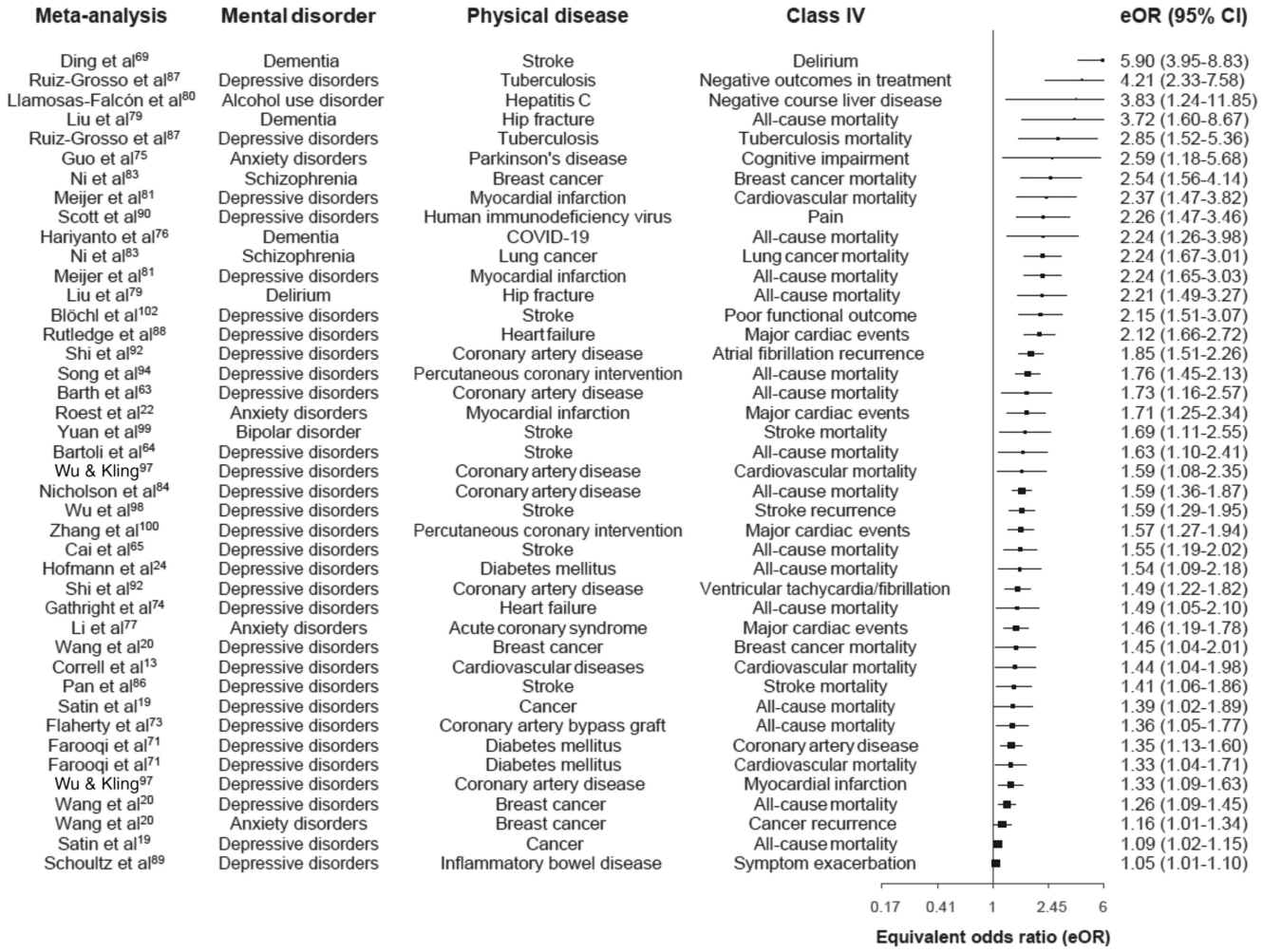

Figure 3 Forest plot of prospective associations between mental disorders and clinical outcomes of physical diseases, stratified by class IV of evidence

and that between depressive disorders and all-cause mortality in patients with chronic kidney disease (RR=1.45, 95% CI: 1.22-1.73). There was either weak (class IV) or no evidence of association for all other primary and secondary outcomes (see Figure 3, Table 2 and supplementary information).

After removing the N>1,000 cases criterion in sensitivity analysis, there was no change in the level of class I, II and III evidence (see Table 2).

Three associations between depressive disorders and primary outcomes were upgraded from weak (class IV) to highly suggestive evidence (class II): those with all-cause mortality in patients with myocardial infarction, percutaneous coronary intervention, and coronary artery disease (see Table 2). The same upgrade was observed for the associations between depressive disorders and two secondary outcomes: major cardiac events in patients with heart failure, and atrial fibrillation recurrence in patients with coronary artery disease (see supplementary information).

One association between depressive disorders and a primary outcome was upgraded from weak (class IV) to suggestive evidence (class III): that with cardiovascular mortality in patients with myocardial infarction (see Table 2). The same upgrade was observed for seven associations between depressive disorders and secondary outcomes: poor functional outcome and stroke recurrence in patients with stroke; major cardiac events in patients with percutaneous coronary intervention; ventricular tachycardia/fibrillation in patients with coronary artery disease; coronary artery disease in patients with diabetes mellitus; negative treatment outcomes in patients with tuberculosis; and pain in patients with HIV infection (see supplementary information).

Associations of mental and behavioural disorders due to psychoactive substance use with clinical outcomes of physical diseases

No association in this diagnostic block was supported by convincing evidence (class I), and there were no data on primary outcomes. The association between alcohol use disorder and de-

Table 1 Level of evidence for the association of neurotic, stress-related and somatoform disorders with primary outcomes of physical diseases

Study	Mental disorder	Physical disease	Outcome	k	Effect size (95% CI)	N cases	p random effects	I^2 %	PI (95% CI)	SSE/ESB	LS	eOR	CE	CES
Neurotic, stress-related and somatoform disorders in patients with cardiovascular diseases														
Emdin et al[70]	Anxiety disorders	Cardiovascular diseases	Cardiovascular mortality	3	RR: 1.46 (1.17-1.82)	3,475	7.2e-04	0.00	0.35-6.04	No/No	No	1.46	III	III
Celano et al[66]	Anxiety disorders	Coronary artery disease	All-cause mortality	8	OR: 1.25 (0.96-1.64)	904	>0.05	43.85	0.7-2.26	No/Yes	No	1.25	NS	NS
Li et al[77]	Anxiety disorders	Acute coronary syndrome	All-cause mortality	5	RR: 1.03 (0.70-1.51)	961	>0.05	44.05	0.35-3.05	No/No	No	1.03	NS	NS
Neurotic, stress-related and somatoform disorders in patients with other physical diseases														
Atlantis et al[23]	Anxiety disorders	Chronic obstructive pulmonary disease	All-cause mortality	3	RR: 1.11 (0.90-1.36)	32	>0.05	0.00	0.29-4.17	No/No	No	1.11	NS	NS
Wang et al[20]	Anxiety disorders	Breast cancer	All-cause mortality	3	HR: 1.07 (0.92-1.23)	1,049	>0.05	0.00	0.42-2.69	No/No	No	1.07	NS	NS

CE – class of evidence, CES – class of evidence after sensitivity analysis (removing the N>1,000 cases criterion), CI – confidence interval, eOR – equivalent odds ratio, ESB – excess significance bias, HR – hazard ratio, LS – largest study with significant effect, NS – not significant, OR – odds ratio, RR – risk ratio, PI – prediction interval, SSE – small study effect

Table 2 Level of evidence for the association of mood disorders with primary outcomes of physical diseases

Study	Mental disorder	Physical disease	Outcome	k	Effect size (95% CI)	N cases	p random effects	I^2%	PI (95% CI)	SSE/ESB	LS	eOR	CE	CES
Mood disorders in patients with cardiovascular diseases														
Sokoreli et al[93]	Depressive disorders	Heart failure	All-cause mortality	4	HR: 1.44 (1.26-1.65)	1,377	1.4e-07	0.00	1.07-1.94	No/No	Yes	1.44	I	I
Gathright et al[74]	Depressive disorders	Heart failure	All-cause mortality	9	HR: 1.49 (1.05-2.10)	1,283	2.5e-02	79.85	0.46-4.79	Yes/Yes	No	1.49	IV	IV
Correll et al[13]	Bipolar disorder	Cardiovascular diseases	Cardiovascular mortality	6	RR: 1.65 (1.32-2.06)	8,923	9.0e-06	80.43	0.86-3.14	No/No	Yes	1.65	III	III
Meijer et al[81]	Depressive disorders	Myocardial infarction	Cardiovascular mortality	5	OR: 2.37 (1.47-3.82)	107	3.8e-04	13.58	0.78-7.22	No/No	No	2.37	IV	III
Meijer et al[81]	Depressive disorders	Myocardial infarction	All-cause mortality	15	OR: 2.24 (1.65-3.03)	725	2.0e-07	48.11	0.92-5.44	No/No	Yes	2.24	IV	II
Song et al[94]	Depressive disorders	Percutaneous coronary intervention	All-cause mortality	6	RR: 1.76 (1.45-2.13)	265	1.1e-08	0.00	1.28-2.41	No/Yes	Yes	1.76	IV	II
Barth et al[63]	Depressive disorders	Coronary artery disease	All-cause mortality	6	HR: 1.73 (1.16-2.57)	1,097	7.1e-03	72.4	0.49-6.12	No/Yes	Yes	1.73	IV	IV
Nicholson et al[84]	Depressive disorders	Coronary artery disease	All-cause mortality	10	RR: 1.59 (1.36-1.87)	412	1.3e-08	9.42	1.32-1.93	No/Yes	Yes	1.59	IV	II
Yuan et al[99]	Bipolar disorder	Stroke	Stroke mortality	2	HR: 1.69 (1.11-2.55)	1,816	3.2e-02	96.52	NA	NA/NA	Yes	1.69	IV	IV
Bartoli et al[64]	Depressive disorders	Stroke	All-cause mortality	5	RR: 1.63 (1.10-2.41)	237	1.5e-02	58.87	0.49-5.39	No/No	No	1.63	IV	IV
Cai et al[65]	Depressive disorders	Stroke	All-cause mortality	8	HR: 1.55 (1.19-2.02)	24,022	1.0e-03	74.47	0.69-3.5	Yes/Yes	Yes	1.55	IV	IV
Wu & Kling[97]	Depressive disorders	Coronary artery disease	Cardiovascular mortality	5	HR: 1.59 (1.08-2.35)	1,654	1.9e-02	82.00	0.41-6.23	Yes/Yes	Yes	1.59	IV	IV
Correll et al[13]	Depressive disorders	Cardiovascular diseases	Cardiovascular mortality	5	OR: 1.44 (1.04-1.98)	8,319	2.6e-02	86.29	0.46-4.44	No/No	No	1.44	IV	IV
Pan et al[86]	Depressive disorders	Stroke	Stroke mortality	4	HR: 1.41 (1.06-1.86)	5,007	1.7e-02	36.91	0.76-2.59	No/No	No	1.41	IV	IV
Flaherty et al[73]	Depressive disorders	Coronary artery bypass graft	All-cause mortality	2	HR: 1.36 (1.05-1.77)	239	2.1e-02	0.00	NA	NA/NA	Yes	1.36	IV	IV
Mood disorders in patients with chronic respiratory diseases														
Atlantis et al[23]	Depressive disorders	Chronic obstructive pulmonary disease	All-cause mortality	6	RR: 2.04 (0.87-4.77)	215	>0.05	73.8	0.12-34.24	No/Yes	Yes	2.04	NS	NS
Courtwright et al[68]	Depressive disorders	Lung transplant	Posttransplant mortality	2	HR: 1.01 (0.99-1.04)	218	>0.05	0.00	NA	NA/NA	No	1.01	NS	NS

Table 2 Level of evidence for the association of mood disorders with primary outcomes of physical diseases *(continued)*

Study	Mental disorder	Physical disease	Outcome	k	Effect size (95% CI)	N cases	p random effects	I^2%	PI (95% CI)	SSE/ESB	LS	eOR	CE	CES
Mood disorders in patients with endocrine system diseases														
Farrokhi et al[72]	Depressive disorders	Kidney failure	All-cause mortality	6	HR: 1.41 (1.31-1.51)	1,834	1.0e-22	12.85	1.28-1.55	Yes/Yes	Yes	1.41	II	II
Hofmann et al[24]	Depressive disorders	Diabetes mellitus	All-cause mortality	7	HR: 2.84 (2.00-4.03)	2,108	4.7e-09	88.81	0.88-9.15	No/No	Yes	1.93	II	II
Hofmann et al[24]	Depressive disorders	Diabetes mellitus	All-cause mortality	6	HR: 1.54 (1.09-2.18)	3,725	1.4e-02	85.18	0.48-4.99	No/Yes	No	1.54	IV	IV
Palmer et al[85]	Depressive disorders	Chronic kidney disease	All-cause mortality	13	HR: 1.45 (1.22-1.73)	2,066	2.0e-05	40.69	0.95-2.22	Yes/Yes	Yes	1.45	III	III
Farooqi et al[71]	Depressive disorders	Diabetes mellitus	Cardiovascular mortality	3	HR: 1.33 (1.04-1.71)	468	2.3e-02	14.51	0.27-6.66	No/No	No	1.33	IV	IV
van Dooren et al[96]	Depressive disorders	Diabetes mellitus	Cardiovascular mortality	2	HR: 1.60 (0.69-3.72)	169	>0.05	77.41	NA	NA/NA	No	1.60	NS	NS
Mood disorders in patients with cancer														
Satin et al[19]	Depressive disorders	Cancer	All-cause mortality	3	RR: 1.39 (1.02-1.89)	55	3.5e-02	0.00	0.19-10.08	No/No	No	1.39	IV	IV
Satin et al[19]	Depressive disorders	Cancer	All-cause mortality	8	HR: 1.09 (1.02-1.15)	1,490	5.2e-03	60.07	0.95-1.24	Yes/No	No	1.09	IV	IV
Wang et al[20]	Depressive disorders	Breast cancer	Breast cancer mortality	2	HR: 1.45 (1.04-2.01)	313	2.7e-02	0.00	NA	NA/NA	No	1.45	IV	IV
Wang et al[20]	Depressive disorders	Breast cancer	All-cause mortality	6	HR: 1.26 (1.09-1.45)	2,021	1.3e-03	0.00	1.03-1.53	No/No	No	1.26	IV	IV
Shi et al[91]	Depressive disorders	High-grade brain tumor	All-cause mortality	3	HR: 1.31 (0.86-1.99)	836	>0.05	0.00	0.09-19.69	No/No	No	1.31	NS	NS
Shi et al[91]	Depressive disorders	Glioma	Glioma mortality	5	RR: 0.74 (0.54-1.02)	627	>0.05	48.61	0.27-2.07	No/No	No	0.74	NS	NS
Mood disorders in patients with other physical diseases														
Ruiz-Grosso et al[87]	Depressive disorders	Tuberculosis	Tuberculosis mortality	2	OR: 2.85 (1.52-5.36)	53	1.1e-03	0.00	NA	NA/NA	Yes	2.85	IV	IV

CE – class of evidence, CES – class of evidence after sensitivity analysis (removing the N>1,000 cases criterion), CI – confidence interval, eOR – equivalent odds ratio, ESB – excess significance bias, HR – hazard ratio, LS – largest study with significant effect, NA – not assessable, NS – not significant, OR – odds ratio, RR – risk ratio, PI – prediction interval, SSE – small study effect

compensated liver cirrhosis in patients with hepatitis C (RR=3.15, 95% CI: 2.87-3.46) presented highly suggestive evidence (class II). After removing the N>1,000 cases criterion in sensitivity analysis, there was no change in the level of evidence (see supplementary information).

Associations of schizophrenia with clinical outcomes of physical diseases

In this diagnostic block, one association presented convincing evidence (class I): that between schizophrenia and cardiovascular mortality in patients with cardiovascular diseases (RR=1.54, 95% CI: 1.36-1.75). One further association was supported by highly suggestive evidence (class II): that between schizophrenia and cancer mortality in patients with cancer (SMR=1.74, 95% CI: 1.41-2.15) (see Figure 2 and Table 3). Two associations presented weak evidence (class IV): those between schizophrenia and cancer mortality in patients with breast and lung cancer (see Figure 3 and Table 3).

After removing the N>1,000 cases criterion in sensitivity analysis, the association between schizophrenia and cancer mortality was upgraded from weak (class IV) to highly suggestive (class II) in patients with lung cancer, and from weak (class IV) to suggestive (class III) in patients with breast cancer. The level of evidence of the other two associations remained unchanged (see Table 3).

Associations of organic, including symptomatic, mental disorders with clinical outcomes of physical diseases

No association in this diagnostic block was supported by convincing, highly suggestive, or suggestive evidence (classes I, II and III). There was weak evidence (class IV) of the association between both dementia and delirium with all-cause mortality in patients with hip fracture; of the association between dementia and all-cause mortality in patients with COVID-19 infection; and of the association between dementia and delirium in patients with stroke (see Table 3 and supplementary information).

After removing the N>1,000 cases criterion in sensitivity analysis, the association between dementia and delirium in patients with stroke was upgraded from weak (class IV) to convincing evidence (class I), while the association between delirium and all-cause mortality in patients with hip fracture was upgraded from weak (class IV) to suggestive (class III) evidence (see Table 3 and supplementary information).

Subgroup analyses

Not all planned subgroup analyses were possible, due to the lack of data (see supplementary information).

When restricting the analyses to standard diagnostic criteria (any version of DSM or ICD), the class II association between depressive disorders and all-cause mortality in patients with diabetes mellitus was downgraded to weak (class IV) evidence. When restricting the analyses to studies formulating a diagnosis of mental disorder before the diagnosis of physical disease (of course, clinical outcomes always followed the diagnosis of a mental disorder), the level of evidence of class I and II associations remained unchanged.

When restricting the analyses to follow-up duration >5 years, the class I association between schizophrenia and cardiovascular mortality in patients with cardiovascular diseases, and the class II associations between depressive disorders and all-cause mortality in patients with kidney failure and diabetes mellitus were downgraded to suggestive or weak evidence (class III and IV). When restricting the analyses to adjusted estimates, only the class I association between schizophrenia and cardiovascular mortality in patients with cardiovascular diseases was downgraded to weak (class IV) evidence.

When restricting the analyses to age of participants <50 years, the class I association between schizophrenia and cardiovascular mortality in cardiovascular diseases was downgraded to weak (class IV) evidence. When restricting the analyses to samples exposed to psychiatric treatments, all class I and II associations were downgraded to either suggestive (class III) or weak (class IV) evidence.

When restricting the analyses to studies including in their samples a majority of males, the class I association between schizophrenia and cardiovascular mortality in patients with cardiovascular diseases, and between depressive disorders and all-cause mortality in patients with heart failure, were downgraded to highly suggestive (class II) or weak (class IV) evidence. The class II associations between depressive disorders and all-cause mortality in patients with kidney failure and diabetes mellitus were downgraded to suggestive or weak evidence (class III or IV).

It is important to note that all the subgroup analyses were conducted in a very small number of primary studies (see supplementary information) and are, therefore, highly underpowered.

Population attributable fraction (PAF) and generalized impact fraction (GIF)

The largest PAF was that for the association of alcohol use disorder with decompensated liver cirrhosis in patients with hepatitis C (30.56%, 95% CI: 27.67-33.49) (see Table 4). GIF analysis showed that alcohol use disorder should be reduced by 33% to prevent 10% of decompensated liver cirrhosis in hepatitis C (see also supplementary information).

The PAFs for the association of depressive disorders with all-cause mortality in patients with diabetes mellitus and kidney failure were respectively 26.81% (95% CI: 16.61-37.67) and 11.59% (95% CI: 9.09-14.14). The PAF for the association of depressive disorders with cardiac events in patients with myocardial infarction was 13.68% (95% CI: 9.87-17.58) (see Table 4). GIF analyses showed that depressive disorders should be reduced by 37% and by 86% to prevent 10% of all-cause mortality in patients with diabetes mellitus and kidney failure, respectively, and be reduced by

Table 3 Level of evidence for the association of schizophrenia and organic, including symptomatic, mental disorders with primary outcomes of physical diseases

Study	Mental disorder	Physical disease	Outcome	k	Effect size (95% CI)	N cases	p random effects	I^2%	PI (95% CI)	SSE/ESB	LS	eOR	CE	CES
Schizophrenia in patients with cardiovascular diseases and cancer														
Correll et al[13]	Schizophrenia	Cardiovascular diseases	Cardiovascular mortality	7	RR: 1.54 (1.36-1.75)	9,097	2.2e-11	27.82	1.19-2.00	No/No	No	1.54	I	I
Zhuo et al[101]	Schizophrenia	Cancer	Cancer mortality	3	SMR: 1.74 (1.41-2.15)	6,145	2.9e-07	66.53	0.17-17.56	No/No	Yes	1.72	II	II
Ni et al[83]	Schizophrenia	Breast cancer	Breast cancer mortality	2	RR: 2.54 (1.56-4.14)	175	1.7e-04	0.00	NA	NA/NA	Yes	2.54	IV	III
Ni et al[83]	Schizophrenia	Lung cancer	Lung cancer mortality	2	RR: 2.24 (1.67-3.01)	192	9.0e-08	0.00	NA	NA/NA	Yes	2.24	IV	II
Organic, including symptomatic, mental disorders in patients with infectious and musculoskeletal system diseases														
Liu et al[79]	Dementia	Hip fracture	All-cause mortality	2	HR: 3.72 (1.6-8.67)	384	2.3e-03	72.52	NA	NA/NA	Yes	3.72	IV	IV
Liu et al[79]	Delirium	Hip fracture	All-cause mortality	6	HR: 2.21 (1.49-3.27)	638	7.5e-05	64.54	0.65-7.51	No/No	Yes	2.21	IV	III
Hariyanto et al[76]	Dementia	COVID-19	All-cause mortality	2	RR: 2.24 (1.26-3.98)	4,417	5.8e-03	89.51	NA	NA/NA	Yes	2.24	IV	IV
Liu et al[26]	Dementia	COVID-19	All-cause mortality	2	OR: 3.27 (0.34-31.43)	148	>0.05	47.02	NA	NA/NA	Yes	3.27	NS	NS

CE – class of evidence, CES – class of evidence after sensitivity analysis (removing the N>1,000 cases criterion), CI – confidence interval, eOR – equivalent odds ratio, ESB – excess significance bias, HR – hazard ratio, LS – largest study with significant effect, NA – not assessable, NS – not significant, OR – odds ratio, RR – risk ratio, PI – prediction interval, SMR – standardized mortality ratio, SSE – small study effect

Table 4 Meta-analytical population attributable fraction (PAF) for the associations supported by the largest evidence (classes I, II and III)

Mental disorder	Physical disease	Outcome	Risk ratio (95% CI)	Prevalence of mental disorder in physical disease (95% CI)	PAF (95% CI)
Depressive disorders	Heart failure	All-cause mortality	1.44 (1.26-1.65)	17.72% (16.89-18.56)	7.25% (4.38-10.34)
Schizophrenia	Cardiovascular diseases	Cardiovascular mortality	1.54 (1.36-1.75)	25.17% (25.08-25.30)	11.99% (8.29-15.84)
Depressive disorders	Diabetes mellitus	Dementia	2.11 (1.77-2.52)	6.66% (6.60-6.71)	6.89% (4.87-9.19)
Depressive disorders	Kidney failure	All-cause mortality	1.41 (1.31-1.51)	32.11% (31.30-32.93)	11.59% (9.09-14.14)
Depressive disorders	Diabetes mellitus	All-cause mortality	2.84 (2.00-4.03)	19.91% (19.07-20.79)	26.81% (16.61-37.67)
Alcohol use disorder	Hepatitis C	Decompensated liver cirrhosis	3.15 (2.87-3.46)	20.50% (20.30-20.70)	30.56% (27.67-33.49)
Depressive disorders	Myocardial infarction	Major cardiac events	1.52 (1.36-1.70)	30.58% (29.62-31.56)	13.68% (9.87-17.58)
Schizophrenia	Cancer	Cancer mortality	1.74 (1.41-2.14)	11.05% (10.75-11.36)	7.53% (4.31-11.21)
Bipolar disorder	Cardiovascular diseases	Cardiovascular mortality	1.65 (1.32-2.06)	3.41% (3.81-4.10)	2.17% (1.16-3.76)
Anxiety disorders	Cardiovascular diseases	Cardiovascular mortality	1.46 (1.17-1.82)	5.50% (5.41-5.64)	2.47% (0.93-4.33)
Depressive disorders	Chronic kidney disease	All-cause mortality	1.45 (1.22-1.73)	10.50% (10.01-10.96)	4.53% (2.24-7.12)

73% to prevent 10% of major cardiac events in patients with myocardial infarction (see Figure 4 and supplementary information).

The PAF of the association of schizophrenia with cardiovascular mortality in patients with cardiovascular diseases was 11.99% (95% CI: 8.29-15.84) (see Table 4). GIF analysis showed that schizophrenia prevalence should be reduced by 83% to prevent 10% of cardiovascular mortality in patients with cardiovascular diseases (see supplementary information).

The PAFs for other class I-III associations are reported in Table 4. They were 7.53% (95% CI: 4.31-11.21) for the association between schizophrenia and cancer mortality in patients with cancer; 7.25% (95% CI: 4.38-10.34) for the association between depressive disorders and all-cause mortality in patients with heart failure; 4.53% (95% CI: 2.24-7.12) for the association between depressive disorders and all-cause mortality in patients with chronic kidney disease; 2.47% (95% CI: 0.93-4.33) for the association between anxiety disorders and cardiovascular mortality in patients with cardiovascular diseases; and 2.17% (95% CI: 1.16-3.76) for the association between bipolar disorder and cardiovascular mortality in patients with cardiovascular diseases.

DISCUSSION

In this umbrella review, we evaluated 47 systematic reviews with meta-analysis, including 251 non-overlapping primary studies, testing 74 prospective associations between mental disorders and 43 primary and 31 secondary clinical outcomes of physical diseases. This is the first attempt to comprehensively evaluate the impact of the entire spectrum of mental disorders on the clinical outcomes of physical diseases, using established grading criteria that control for several biases. This is also the first study to employ *metaumbrella*, a comprehensive suite of statistical packages developed for conducting umbrella reviews[50,62]. We also estimated for the first time the meta-umbrella preventive capacity (meta-analytic PAFs) of the associations supported by class I-III evidence to establish reliable, evidence-based and actionable targets that can be prioritized in clinical practice.

An additional strength of this work is the in-depth screening of primary studies included in each systematic review to selectively include only data reflecting prospective associations. This choice mitigates the reverse causality bias and ensures the temporality of the examined associations, where exposures (mental disorders) always preceded the event investigated (clinical outcomes of physical diseases). Furthermore, we also screened primary studies to include only those using robust diagnostic or research criteria, or validated instruments with specific cut-offs mapped to discrete categories of mental disorders. This approach overcomes the significant noise derived from studies that mistake continuous symptoms or self-reported subjective "experiences" for categorical mental disorders, which characterizes the existing transdiagnostic literature[103,104]. Our refined evidence synthesis method resulted in more than two-thirds (68%) of the included systematic reviews having a low risk of bias and nearly 80% of the selected

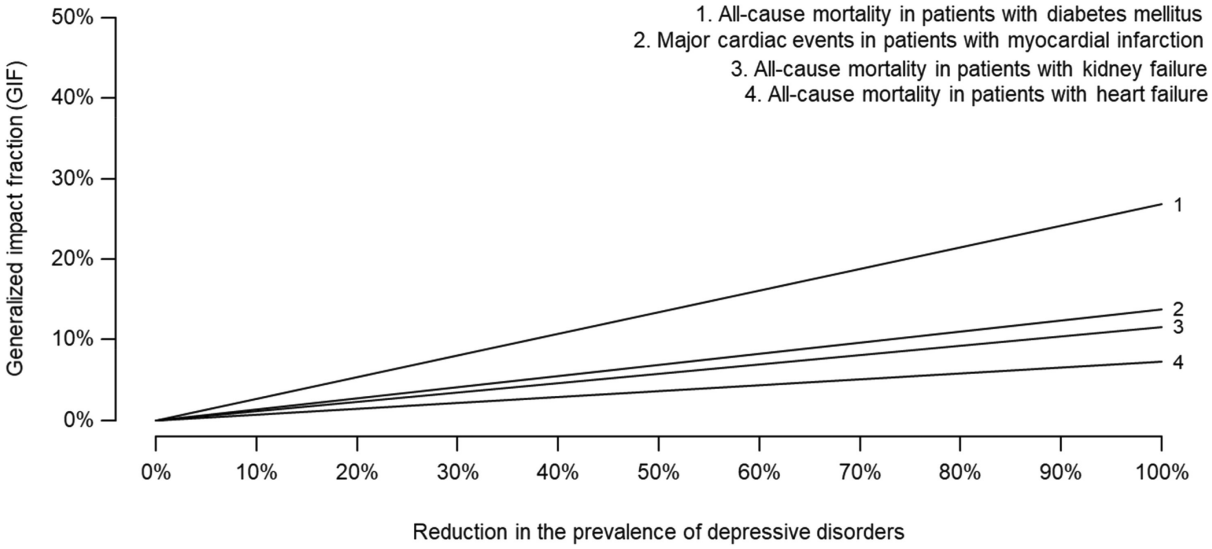

Figure 4 Meta-analytic generalized impact fraction (GIF) of depressive disorders for all-cause mortality and major cardiac events in several physical diseases

primary studies scoring high on quality assessments.

Mood disorders (especially depressive disorders) emerged as credible risk factors for adverse clinical outcomes in cardiovascular diseases, as most associations in this class were supported by the largest evidence (classes I, II or III). The most robust association (class I) was that between depressive disorders and all-cause mortality among patients with heart failure, which remained at the same level of evidence after conducting subgroup analyses accounting for confounders. Other highly suggestive/suggestive associations were those between depressive disorders and the risk of major cardiac events in patients with myocardial infarction (class II), and between bipolar disorder and the risk of cardiovascular mortality in patients with cardiovascular disease (class III).

Overall, the association between depressive disorders and cardiovascular diseases is a consolidated area of research across psychiatry and somatic medicine[7,105-108], although the underlying mechanisms are not fully understood[107]. The pathophysiology of these conditions may share common mechanisms, including behavioral, biological and medication-related ones[108-112], forming an interdependent network[113].

Behavioral mechanisms may include unhealthy habits (smoking, excessive alcohol consumption, physical inactivity, unhealthy diet, medication non-adherence) that accelerate pathophysiological processes, such as atherosclerosis, leading to poor health outcomes and increased mortality[109,110,112-114].

Biological mechanisms may include alterations in the autonomic nervous system, plasminogen activator inhibitor-1 and fibrinogen levels, endothelial function, and neurohormonal factors, as well as diminished heart rate variability, and genetic alterations of the serotonin transporter[109,110,112,113,114]. Molecular inflammatory mechanisms involving interleukins (IL-6 and IL-1β) and C-reactive protein, as well as an oxidative stress imbalance, may also point to common pathways between mood and cardiovascular condictions[109,115-118].

Mechanisms associated with treatment (for example, antidepressants use) may include cardiotoxicity[109,110,113,114], or the alteration of platelet activation[111] leading to an increased incidence of major cardiac events and sudden death[109-111,113,114]. However, the latter is unlikely a strong mechanism, especially when using selective serotonin reuptake inhibitors, which reduce platelet aggregation[119,120].

We also found highly suggestive (class II) evidence that depressive disorders increase all-cause mortality risk in patients with diabetes mellitus and kidney failure. The increased mortality in diabetes mellitus is due to insulin resistance and metabolic factors (e.g., abdominal obesity and dyslipidemia). These factors are aggravated by depressive disorders, which are independently associated with insulin resistance[121] and metabolic syndrome (elevated adipose tissue and dyslipidemia[122,123]). The increased mortality in depressed patients with kidney failure may be due to suboptimal compliance with complex medication regimens[123-125].

Highly suggestive (class II) evidence was similarly found for the association between depressive disorders and an increased risk of dementia in patients with diabetes mellitus[67]. Both depressive disorders and diabetes mellitus have been shown to increase the incidence of dementia individually and synergistically[126], with the metabolic-brain axis as a key mediator connecting these conditions[126]. Depressive disorders are associated with micro/macro vascular alterations[127,128], insulin resistance[121] and neuroinflammation[129]; these factors may increase the risk of dementia in this patient population[130,131]. Stress and psychosocial determinants of health may also be key mediators in how these systems interact[126].

These are clinically highly relevant findings, as depression prevention and/or treatment has great potential to improve overall health and outcomes in common physical diseases that are as-

sociated with severe biopsychosocial and societal burden (e.g., dementia is a rising problem in ageing societies[132]) and premature mortality. Our PAF analysis directly informs the prioritization of these approaches and associated resources on the basis of evidence-based potential preventive gains. For example, this study provides the first robust meta-umbrella evidence showing that preventing depressive disorders could reduce up to one-third of mortality rates across various physical conditions.

Screening for depression in patients with cardiovascular diseases is recommended by the US Preventive Services Task Force and the American Heart Association[133,134]. Furthermore, independent meta-analyses showed that psychotherapy/psychoeducation can have a preventive effect by reducing the severity of symptoms before the onset of depressive disorders[135-137]. Randomized controlled trials demonstrated that collaborative care, which includes patient preferences, cognitive intervention and/or lifestyle advice, drug treatment management, and relapse prevention[138], or physical exercise[139,140], can specifically reduce depression in patients with cardiovascular diseases or diabetes, including low- and middle-income countries[141,142]. These interventions could, at the same time, have an impact on depressive disorders and improve self-management of physical diseases in patients with mental and physical multimorbidity[143]. Our GIF analysis confirms these benefits; the reduction of mortality rates remains clinically relevant even if preventive interventions are only partially effective. Taken together, these findings call for a new generation of translational research validating preventive approaches for depressive disorders in physical conditions.

The association between schizophrenia and increased cardiovascular and cancer mortality in patients with these physical diseases was also supported by convincing or highly suggestive evidence (class I and II, respectively). The higher mortality risk in schizophrenia compared to the general population is substantial and particularly marked during the early stages of the disorder[144]. The increased risk of cardiovascular and cancer mortality may be due to suboptimal cardiovascular[145,146] and cancer screening[147] in patients with schizophrenia, coupled with high cigarette smoking[145], frequent metabolic syndrome (obesity, hypertension, diabetes, hyperlipidemia)[148-152], physical inactivity, drug and alcohol use, and poor adherence to medication[153-155].

Although antipsychotics can lead to adverse cardiometabolic effects that are a risk factor for cardiovascular mortality[156], a recent meta-analysis showed that all-cause mortality risk at the population level is substantially reduced with antipsychotic use versus no antipsychotic use (RR=0.71)[144]. The reason for this paradoxical relationship can be found in a nationwide database within-subject analysis, where ongoing antipsychotic treatment was associated with higher adherence to statins, antihypertensive and antidiabetic medications[157]. Thus, greater psychiatric stability via antipsychotic treatment improves not only healthy lifestyle behaviors but also adherence to medications for secondary physical illness prevention[144].

Furthermore, our PAF analysis suggests that preventing psychosis in young people at clinical high risk can produce physical health benefits in terms of reduced cardiovascular and cancer mortality, in addition to improved mental health outcomes[158-165] (indicated prevention).

Highly suggestive evidence (class II) was also found for the association of alcohol use disorder with decompensated liver cirrhosis in patients infected with hepatitis C virus. Indeed, alcohol use disorder leads to alterations in cytokine production, lipopolysaccharide-TLR4 signalling, and reactive oxygen species[166], factors that increase hepatotoxicity[167,168]. Patients with alcohol use disorder are also frequently medically ineligible for hepatitis C treatment[169].

Our PAF analysis demonstrates that about one-third of decompensated liver cirrhosis in patients with hepatitis C could be averted by preventing alcohol use disorder (the largest PAF in our study). Thus, alcohol use disorder should be identified and managed as much as possible to improve psychiatric as well as physical health outcomes. Screening for unhealthy alcohol use in primary care settings in adults, including pregnant women, and providing brief behavioral counselling interventions is an evidence-based approach to reducing unhealthy alcohol use, as recommended by the US Preventive Services Task Force[170].

There are some limitations to this study. First, while we avoided the limitations of retrospective or case-control study designs by selecting only prospective systematic reviews with meta-analysis and prospective primary studies, the observed associations do not represent pathophysiological causality. For example, although we preferably focused on adjusted estimates, we could not specifically address the role of single confounders, such as genetic effects, body mass index or metabolic risk factors, which may at least partially account for the observed associations. Second, there were few relevant systematic reviews with meta-analysis in child and adolescent populations, and for mental disorders other than depressive disorders. For example, we did not find any relevant meta-analysis that considered patients with anorexia nervosa or personality disorders. Third, the results of the subgroup analyses should be viewed with caution, due to the granularity of the reported data and the very limited statistical power. Finally, our PAF findings are specific to the populations affected with physical diseases and cannot be applied to the general population.

Acknowledging these caveats, our study has several implications. We demonstrated at a meta-umbrella review level that mental disorders significantly impair the health and life expectancy of individuals with physical diseases, and quantified for the first time the associated preventive capacity. Our findings may be particularly relevant for informing the prioritization of preventive approaches for physical diseases via improved detection and management of mental disorders, with currently the best evidence and actionable targets for alcohol use disorders, depression and schizophrenia.

These approaches are likely to be particularly relevant for young people, given the early age at onset of most mental disorders[40,171]. Prevention for youth is currently driven by initiatives siloed in physical diseases, such as cancer and obesity[143,165]. However, preventing the onset of mental disorders can become a tantalizing strategy for reducing at the same time the risk of developing phy-

sical diseases[143]. Indeed, the cost and risk associated with preventive approaches (e.g., ethical concerns[172]) can be offset by concurrently reducing the burden of both psychiatric disorders and physical diseases[165,173]. Integrating early detection and prevention of mental health and physical conditions may be particularly cost-effective in resource-constrained settings[142].

This strategy would require innovative integrated or, at least, co-located clinical services for emerging mental and physical conditions, overcoming the limited preventive capacity of current health care services[165]. Indeed, youth-friendly mental and physical health care services are being developed and tested worldwide[174-177], and promise to achieve the much-needed cross-disciplinary fertilization of expertise which is essential to reduce the Cartesian dichotomy between mental and physical knowledge, education and research.

In conclusion, this umbrella review demonstrates that mental disorders increase the risk of several poor clinical outcomes in patients with physical diseases. Prevention targeting mental disorders – particularly alcohol use disorders, depressive disorders, and schizophrenia – can reduce the incidence of adverse clinical outcomes in physical diseases. These findings can inform clinical practice and trans-speciality preventive approaches cutting across psychiatric and somatic medicine.

ACKNOWLEDGEMENTS

G. Arrondo is supported by the European Social Fund and the Spanish Research Agency (RYC2020-030744-I/AEI/10.13039/501100011033). M. Berk is supported by National Health and Medical Research Council (NHMRC) Senior Principal Research Fellowship and Leadership 3 Investigator grants (1156072 and 2017131). Supplementary information on this study is available at https://github.com/eldragio/DRAGIOTI_UR_2022/raw/main/Supplementary%20information.pdf.

REFERENCES

1. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020;396:1204-22.
2. GBD 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Psychiatry 2022;9:137-50.
3. Cortese S, Sun S, Zhang J et al. Association between attention deficit hyperactivity disorder and asthma: a systematic review and meta-analysis and a Swedish population-based study. Lancet Psychiatry 2018;5:717-26.
4. Lim LF, Solmi M, Cortese S. Association between anxiety and hypertension in adults: a systematic review and meta-analysis. Neurosci Biobehav Rev 2021; 131:96-119.
5. Cortese S, Moreira-Maia CR, St Fleur D et al. Association between ADHD and obesity: a systematic review and meta-analysis. Am J Psychiatry 2016;173:34-43.
6. Kim JY, Choi MJ, Ha S et al. Association between autism spectrum disorder and inflammatory bowel disease: a systematic review and meta-analysis. Autism Res 2022;15:340-52.
7. Vancampfort D, Correll CU, Galling B et al. Diabetes mellitus in people with schizophrenia, bipolar disorder and major depressive disorder: a systematic review and large scale meta-analysis. World Psychiatry 2016;15:166-74.
8. Arrondo G, Solmi M, Dragioti E et al. Associations between mental and physical conditions in children and adolescents: an umbrella review. Neurosci Biobehav Rev 2022;137:104662.
9. Machado MO, Veronese N, Sanches M et al. The association of depression and all-cause and cause-specific mortality: an umbrella review of systematic reviews and meta-analyses. BMC Med 2018;16:112.
10. Scott KM, Lim C, Al-Hamzawi A et al. Association of mental disorders with subsequent chronic physical conditions: World Mental Health Surveys from 17 countries. JAMA Psychiatry 2016;73:150-8.
11. Correll CU, Detraux J, De Lepeleire J et al. Effects of antipsychotics, antidepressants and mood stabilizers on risk for physical diseases in people with schizophrenia, depression and bipolar disorder. World Psychiatry 2015;14: 119-36.
12. Leucht S, Burkard T, Henderson J et al. Physical illness and schizophrenia: a review of the literature. Acta Psychiatr Scand 2007;116:317-33.
13. Correll CU, Solmi M, Veronese N et al. Prevalence, incidence and mortality from cardiovascular disease in patients with pooled and specific severe mental illness: a large-scale meta-analysis of 3,211,768 patients and 113,383,368 controls. World Psychiatry 2017;16:163-80.
14. Mezuk B, Eaton WW, Albrecht S et al. Depression and type 2 diabetes over the lifespan: a meta-analysis. Diabetes Care 2008;31:2383-90.
15. Rugulies R. Depression as a predictor for coronary heart disease. a review and meta-analysis. Am J Prev Med 2002;23:51-61.
16. Momen NC, Plana-Ripoll O, Agerbo E et al. Association between mental disorders and subsequent medical conditions. N Engl J Med 2020;382:1721-31.
17. Morris G, Puri BK, Walker AJ et al. Shared pathways for neuroprogression and somatoprogression in neuropsychiatric disorders. Neurosci Biobehav Rev 2019; 107:862-82.
18. Lando J, Williams SM, Williams B et al. A logic model for the integration of mental health into chronic disease prevention and health promotion. Prev Chronic Dis 2006;3:A61.
19. Satin JR, Linden W, Phillips MJ. Depression as a predictor of disease progression and mortality in cancer patients: a meta-analysis. Cancer 2009;115:5349-61.
20. Wang X, Wang N, Zhong L et al. Prognostic value of depression and anxiety on breast cancer recurrence and mortality: a systematic review and meta-analysis of 282,203 patients. Mol Psychiatry 2020;25:3186-97.
21. Meijer A, Conradi HJ, Bos EH et al. Adjusted prognostic association of depression following myocardial infarction with mortality and cardiovascular events: individual patient data meta-analysis. Br J Psychiatry 2013;203:90-102.
22. Roest AM, Martens EJ, Denollet J et al. Prognostic association of anxiety post myocardial infarction with mortality and new cardiac events: a meta-analysis. Psychosom Med 2010;72:563-9.
23. Atlantis E, Fahey P, Cochrane B et al. Bidirectional associations between clinically relevant depression or anxiety and COPD: a systematic review and meta-analysis. Chest 2013;144:766-77.
24. Hofmann M, Kohler B, Leichsenring F et al. Depression as a risk factor for mortality in individuals with diabetes: a meta-analysis of prospective studies. PLoS One 2013;8:e79809.
25. Iversen MM, Nefs G, Tell GS et al. Anxiety and depressive symptoms as predictors of all-cause mortality among people with insulin-naive type 2 diabetes: 17-year follow-up of the second Nord-Trondelag Health Survey (HUNT2), Norway. PLoS One 2016;11:e0160861.
26. Liu L, Ni SY, Yan W et al. Mental and neurological disorders and risk of COVID-19 susceptibility, illness severity and mortality: a systematic review, meta-analysis and call for action. EClinicalMedicine 2021;40:101111.
27. Rogers JP, Chesney E, Oliver D et al. Psychiatric and neuropsychiatric presentations associated with severe coronavirus infections: a systematic review and meta-analysis with comparison to the COVID-19 pandemic. Lancet Psychiatry 2020;7:611-27.
28. Vai B, Mazza MG, Marisa CD et al. Joint European policy on the COVID-19 risks for people with mental disorders: an umbrella review and evidence- and consensus-based recommendations for mental and public health. Eur Psychiatry 2022;65:e47.
29. Bray F, Soerjomataram I. Population attributable fractions continue to unmask the power of prevention. Br J Cancer 2018;118:1031-2.
30. Shield KD, Parkin DM, Whiteman DC et al. Population attributable and preventable fractions: cancer risk factor surveillance, and cancer policy projection. Curr Epidemiol Rep 2016;3:201-11.
31. Whiteman DC, Webb PM, Green AC et al. Cancers in Australia in 2010 attributable to modifiable factors: summary and conclusions. Aust N Z J Public Health 2015;39:477-84.
32. Rezende LF, Eluf-Neto J. Population attributable fraction: planning of diseases prevention actions in Brazil. Rev Saude Publica 2016;50:30.
33. Radua J, Ramella-Cravaro V, Ioannidis JPA et al. What causes psychosis? An umbrella review of risk and protective factors. World Psychiatry 2018;17:49-66.
34. Dragioti E, Solmi M, Favaro A et al. Association of antidepressant use with adverse health outcomes: a systematic umbrella review. JAMA Psychiatry 2019; 76:1241-55.

35. Ioannidis JP. Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. CMAJ 2009;181:488-93.
36. Sartorius N. Comorbidity of mental and physical diseases: a main challenge for medicine of the 21st century. Shanghai Arch Psychiatry 2013;25:68-9.
37. Page MJ, McKenzie JE, Bossuyt PM et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;10:89.
38. Dekkers OM, Vandenbroucke JP, Cevallos M et al. COSMOS-E: Guidance on conducting systematic reviews and meta-analyses of observational studies of etiology. PLoS Med 2019;16:e1002742.
39. Arango C, Dragioti E, Solmi M et al. Risk and protective factors for mental disorders beyond genetics: an evidence-based atlas. World Psychiatry 2021;20: 417-36.
40. Dragioti E, Radua J, Solmi M et al. Global population attributable fraction of potentially modifiable risk factors for mental disorders: a meta-umbrella systematic review. Mol Psychiatry 2022; doi: 10.1038/s41380-022-01586-8.
41. Spitzer RL, Endicott J, Robins E. Research Diagnostic Criteria: rationale and reliability. Arch Gen Psychiatry 1978;35:773-82.
42. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16:606-13.
43. Whiting P, Savovic J, Higgins JP et al. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. J Clin Epidemiol 2016;69:225-34.
44. Buhn S, Mathes T, Prengel P et al. The risk of bias in systematic reviews tool showed fair reliability and good construct validity. J Clin Epidemiol 2017;91: 121-8.
45. Li L, Yang L, Luo B et al. Acupuncture for post-stroke cognitive impairment: an overview of systematic reviews. Int J Gen Med 2022;15:7249-64.
46. Wells GA, Shea B, O'Connell D et al. The Newcastle-Ottawa scale (NOS) for assessing the quality of non-randomised studies in meta-analyses. http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm.
47. Fusar-Poli P, Radua J. Ten simple rules for conducting umbrella reviews. Evid Based Ment Health 2018;21:95-100.
48. Borenstein M, Hedges LV, Higgins JPT et al. Introduction to meta-analysis. Hoboken: Wiley, 2009.
49. Fisher ZF, Tipton E. robumeta: an R-package for robust variance estimation in meta-analysis. arXiv:1503.02220.
50. Gosling CJ, Cartigny A, Mellier BC et al. Efficacy of psychosocial interventions for autism spectrum disorder: an umbrella review. Mol Psychiatry 2022; doi: 10.1038/s41380-022-01670-z.
51. Higgins JP, Thompson SG, Deeks JJ et al. Measuring inconsistency in meta-analyses. BMJ 2003;327:557-60.
52. Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. BMJ 2011;342:d549.
53. Sterne JA, Sutton AJ, Ioannidis JP et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ 2011;343:d4002.
54. Stanley TD, Doucouliagos H, Ioannidis JPA et al. Detecting publication selection bias through excess statistical significance. Res Synth Methods 2021; 12:776-95.
55. Belbasis L, Bellou V, Evangelou E et al. Environmental risk factors and multiple sclerosis: an umbrella review of systematic reviews and meta-analyses. Lancet Neurol 2015;14:263-73.
56. Markozannes G, Aretouli E, Rintou E et al. An umbrella review of the literature on the effectiveness of psychological interventions for pain reduction. BMC Psychol 2017;5:31.
57. Mukadam N, Sommerlad A, Huntley J et al. Population attributable fractions for risk factors for dementia in low-income and middle-income countries: an analysis using cross-sectional survey data. Lancet Glob Health 2019;7:e596-603.
58. Levin ML. The occurrence of lung cancer in man. Acta Unio Int Contra Cancrum 1953;9:531-41.
59. Lin C-K, Chen S-T. Estimation and application of population attributable fraction in ecological studies. Environ Health 2019;18:52.
60. Zhang J, Yu KF. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. JAMA 1998;280:1690-1.
61. Mansournia MA, Altman DG. Population attributable fraction. BMJ 2018; 360:k757.
62. Gosling CJ, Solanes A, Fusar-Poli P et al. metaumbrella: the first comprehensive suite to perform data analysis in umbrella reviews with stratification of the evidence. Evid Based Ment Health (in press).
63. Barth J, Schumacher M, Herrmann-Lingen C. Depression as a risk factor for mortality in patients with coronary heart disease: a meta-analysis. Psychosom Med 2004;66:802-13.
64. Bartoli F, Di Brita C, Crocamo C et al. Early post-stroke depression and mortality: meta-analysis and meta-regression. Front Psychiatry 2018;9:530.
65. Cai W, Mueller C, Li YJ et al. Post stroke depression and risk of stroke recurrence and mortality: a systematic review and meta-analysis. Ageing Res Rev 2019;50:102-9.
66. Celano CM, Millstein RA, Bedoya CA et al. Association between anxiety and mortality in patients with coronary artery disease: a meta-analysis. Am Heart J 2015;170:1105-15.
67. Chow YY, Verdonschot M, McEvoy CT et al. Associations between depression and cognition, mild cognitive impairment and dementia in persons with diabetes mellitus: a systematic review and meta-analysis. Diabetes Res Clin Pract 2022;185:109227.
68. Courtwright AM, Salomon S, Lehmann LS et al. The effect of pretransplant depression and anxiety on survival following lung transplant: a meta-analysis. Psychosomatics 2016;57:238-45.
69. Ding G, Hua S, Chen J et al. Does cognitive decline/dementia increase delirium risk after stroke? Psychogeriatrics 2021;21:605-11.
70. Emdin CA, Odutayo A, Wong CX et al. Meta-analysis of anxiety as a risk factor for cardiovascular disease. Am J Cardiol 2016;118:511-9.
71. Farooqi A, Khunti K, Abner S et al. Comorbid depression and risk of cardiac events and cardiac mortality in people with diabetes: a systematic review and meta-analysis. Diabetes Res Clin Pract 2019;156:107816.
72. Farrokhi F, Abedi N, Beyene J et al. Association between depression and mortality in patients receiving long-term dialysis: a systematic review and meta-analysis. Am J Kidney Dis 2014;63:623-35.
73. Flaherty LB, Wood T, Cheng A et al. Pre-existing psychological depression confers increased risk of adverse cardiovascular outcomes following cardiac surgery: a systematic review and meta-analysis. J Thorac Cardiovasc Surg 2017;154:1578-86.e1.
74. Gathright EC, Goldstein CM, Josephson RA et al. Depression increases the risk of mortality in patients with heart failure: a meta-analysis. J Psychosom Res 2017;94:82-9.
75. Guo Y, Liu FT, Hou XH et al. Predictors of cognitive impairment in Parkinson's disease: a systematic review and meta-analysis of prospective cohort studies. J Neurol 2021;268:2713-22.
76. Hariyanto TI, Putri C, Arisa J et al. Dementia and outcomes from coronavirus disease 2019 (COVID-19) pneumonia: a systematic review and meta-analysis. Arch Gerontol Geriatr 2021;93:104299.
77. Li J, Ji F, Song J et al. Anxiety and clinical outcomes of patients with acute coronary syndrome: a meta-analysis. BMJ Open 2020;10:e034135.
78. Li JQ, Tan L, Wang HF et al. Risk factors for predicting progression from mild cognitive impairment to Alzheimer's disease: a systematic review and meta-analysis of cohort studies. J Neurol Neurosurg Psychiatry 2016;87:476-84.
79. Liu Y, Wang Z, Xiao W. Risk factors for mortality in elderly patients with hip fractures: a meta-analysis of 18 studies. Aging Clin Exp Res 2018;30:323-30.
80. Llamosas-Falcon L, Shield KD, Gelovany M et al. Alcohol use disorders and the risk of progression of liver disease in people with hepatitis C virus infection – a systematic review. Subst Abuse Treat Prev Policy 2020;15:45.
81. Meijer A, Conradi HJ, Bos EH et al. Prognostic association of depression following myocardial infarction with mortality and cardiovascular events: a meta-analysis of 25 years of research. Gen Hosp Psychiatry 2011;33:203-16.
82. Mourao RJ, Mansur G, Malloy-Diniz LF et al. Depressive symptoms increase the risk of progression to dementia in subjects with mild cognitive impairment: systematic review and meta-analysis. Int J Geriatr Psychiatry 2016;31: 905-11.
83. Ni L, Wu J, Long Y et al. Mortality of site-specific cancer in patients with schizophrenia: a systematic review and meta-analysis. BMC Psychiatry 2019;19:323.
84. Nicholson A, Kuper H, Hemingway H. Depression as an aetiologic and prognostic factor in coronary heart disease: a meta-analysis of 6362 events among 146 538 participants in 54 observational studies. Eur Heart J 2006;27:2763-74.
85. Palmer SC, Vecchio M, Craig JC et al. Association between depression and death in people with CKD: a meta-analysis of cohort studies. Am J Kidney Dis 2013;62:493-505.
86. Pan A, Sun Q, Okereke OI et al. Depression and risk of stroke morbidity and mortality: a meta-analysis and systematic review. JAMA 2011;306:1241-9.
87. Ruiz-Grosso P, Cachay R, de la Flor A et al. Association between tuberculosis and depression on negative outcomes of tuberculosis treatment: a systematic review and meta-analysis. PLoS One 2020;15:e0227472.
88. Rutledge T, Reis VA, Linke SE et al. Depression in heart failure a meta-analytic review of prevalence, intervention effects, and associations with clinical outcomes. J Am Coll Cardiol 2006;48:1527-37.
89. Schoultz M, Beattie M, Gorely T et al. Assessment of causal link between psychological factors and symptom exacerbation in inflammatory bowel disease:

a systematic review utilising Bradford Hill criteria and meta-analysis of prospective cohort studies. Syst Rev 2020;9:169.
90. Scott W, Arkuter C, Kioskli K et al. Psychosocial factors associated with persistent pain in people with HIV: a systematic review with meta-analysis. Pain 2018;159:2461-76.
91. Shi C, Lamba N, Zheng LJ et al. Depression and survival of glioma patients: a systematic review and meta-analysis. Clin Neurol Neurosurg 2018;172:8-19.
92. Shi S, Liu T, Liang J et al. Depression and risk of sudden cardiac death and arrhythmias: a meta-analysis. Psychosom Med 2017;79:153-61.
93. Sokoreli I, de Vries JJG, Pauws SC et al. Depression and anxiety as predictors of mortality among heart failure patients: systematic review and meta-analysis. Heart Fail Rev 2016;21:49-63.
94. Song X, Song J, Shao M et al. Depression predicts the risk of adverse events after percutaneous coronary intervention: a meta-analysis. J Affect Disord 2020; 266:158-64.
95. Taggart Wasson L, Shaffer JA, Edmondson D et al. Posttraumatic stress disorder and nonadherence to medications prescribed for chronic medical conditions: a meta-analysis. J Psychiatr Res 2018;102:102-9.
96. van Dooren FE, Nefs G, Schram MT et al. Depression and risk of mortality in people with diabetes mellitus: a systematic review and meta-analysis. PLoS One 2013;8:e57058.
97. Wu Q, Kling JM. Depression and the risk of myocardial infarction and coronary death: a meta-analysis of prospective cohort studies. Medicine 2016;95: e2815.
98. Wu QE, Zhou AM, Han YP et al. Poststroke depression and risk of recurrent stroke: a meta-analysis of prospective studies. Medicine 2019;98:e17235.
99. Yuan M, Xiao ZL, Zhou HY et al. Bipolar disorder and the risk for stroke incidence and mortality: a meta-analysis. Neurol Sci 2022;43:467-76.
100. Zhang WY, Nan N, Song XT et al. Impact of depression on clinical outcomes following percutaneous coronary intervention: a systematic review and meta-analysis. BMJ Open 2019;9:e026445.
101. Zhuo C, Tao R, Jiang R et al. Cancer mortality in patients with schizophrenia: systematic review and meta-analysis. Br J Psychiatry 2017;211:7-13.
102. Blochl M, Meissner S, Nestler S. Does depression after stroke negatively influence physical disability? A systematic review and meta-analysis of longitudinal studies. J Affect Disord 2019;247:45-56.
103. Fusar-Poli P. TRANSD recommendations: improving transdiagnostic research in psychiatry. World Psychiatry 2019;18:361-2.
104. Fusar-Poli P, Solmi M, Brondino N et al. Transdiagnostic psychiatry: a systematic review. World Psychiatry 2019;18:192-207.
105. Whooley MA, Wong JM. Depression and cardiovascular disorders. Annu Rev Clin Psychol 2013;9:327-54.
106. Gan Y, Gong Y, Tong X et al. Depression and the risk of coronary heart disease: a meta-analysis of prospective cohort studies. BMC Psychiatry 2014;14:371.
107. De Hert M, Detraux J, Vancampfort D. The intriguing relationship between coronary heart disease and mental disorders. Dialogues Clin Neurosci 2018; 20:31-40.
108. Gold SM, Köhler-Forsberg O, Moss-Morris R et al. Comorbid depression in medical diseases. Nat Rev Dis Primers 2020;6:69.
109. Shao M, Lin X, Jiang D et al. Depression and cardiovascular disease: shared molecular mechanisms and clinical implications. Psychiatry Res 2020;285: 112802.
110. Penninx BW. Depression and cardiovascular disease: epidemiological evidence on their linking mechanisms. Neurosci Biobehav Rev 2017;74:277-86.
111. Grippo AJ, Johnson AK. Biological mechanisms in the relationship between depression and heart disease. Neurosci Biobehav Rev 2002;26:941-962.
112. Joynt KE, Whellan DJ, O'Connor CM. Depression and cardiovascular disease: mechanisms of interaction. Biol Psychiatry 2003;54:248-61.
113. Stapelberg NJ, Neumann DL, Shum DH et al. A topographical map of the causal network of mechanisms underlying the relationship between major depressive disorder and coronary heart disease. Aust N Z J Psychiatry 2011;45: 351-69.
114. Hare DL, Toukhsati SR, Johansson P et al. Depression and cardiovascular disease: a clinical review. Eur Heart J 2014;35:1365-72.
115. Khandaker GM, Zuber V, Rees JMB et al. Shared mechanisms between coronary heart disease and depression: findings from a large UK general population-based cohort. Mol Psychiatry 2020;25:1477-86.
116. Howren MB, Lamkin DM, Suls J. Associations of depression with C-reactive protein, IL-1, and IL-6: a meta-analysis. Psychosom Med 2009;71:171-86.
117. Jiménez-Fernández S, Gurpegui M, Díaz-Atienza F et al. Oxidative stress and antioxidant parameters in patients with major depressive disorder compared to healthy controls before and after antidepressant treatment: results from a meta-analysis. J Clin Psychiatry 2015;76:1658-67.
118. Jiménez-Fernández S, Gurpegui M, Garrote-Rojas D et al. Oxidative stress parameters and antioxidants in adults with unipolar or bipolar depression versus healthy controls: systematic review and meta-analysis. J Affect Disord 2022;314:211-21.
119. Taylor CB, Youngblood ME, Catellier D et al. Effects of antidepressant medication on morbidity and mortality in depressed patients after myocardial infarction. Arch Gen Psychiatry 2005;62:792-8.
120. O'Connor CM, Jiang W, Kuchibhatla M et al. Antidepressant use, depression, and survival in patients with heart failure. Arch Intern Med 2008;168:2232-7.
121. Fernandes BS, Salagre E, Enduru N et al. Insulin resistance in depression: a large meta-analysis of metabolic parameters and variation. Neurosci Biobehav Rev 2022;139:104758.
122. Penninx BWJH, Milaneschi Y, Lamers F et al. Understanding the somatic consequences of depression: biological mechanisms and the role of depression symptom profile. BMC Med 2013;11:129.
123. Vancampfort D, Correll CU, Wampers M et al. Metabolic syndrome and metabolic abnormalities in patients with major depressive disorder: a meta-analysis of prevalences and moderating variables. Psychol Med 2014;44:2017-28.
124. Vian J, Pereira C, Chavarria V et al. The renin-angiotensin system: a possible new target for depression. BMC Med 2017;15:144.
125. Kobori H, Nangaku M, Navar LG et al. The intrarenal renin-angiotensin system: from physiology to the pathobiology of hypertension and kidney disease. Pharmacol Rev 2007;59:251-87.
126. Rosenblat JD, Mansur RB, Cha DS et al. Depression, diabetes and dementia. In: Sartorius N, Holt RIG, Maj M (eds). Comorbidity of mental and physical disorders. Basel: Karger, 2015:42-53.
127. Sullivan MD, Katon WJ, Lovato LC et al. Association of depression with accelerated cognitive decline among patients with type 2 diabetes in the ACCORD-MIND trial. JAMA Psychiatry 2013;70:1041-7.
128. Katon W, Pedersen HS, Ribe AR et al. Effect of depression and diabetes mellitus on the risk for dementia: a national population-based cohort study. JAMA Psychiatry 2015;72:612-9.
129. Mousten IV, Sørensen NV, Christensen RHB et al. Cerebrospinal fluid biomarkers in patients with unipolar depression compared with healthy control individuals: a systematic review and meta-analysis. JAMA Psychiatry 2022;79: 571-81.
130. Brzezińska A, Bourke J, Rivera-Hernández R et al. Depression in dementia or dementia in depression? Systematic review of studies and hypotheses. Curr Alzheimer Res 2020;17:16-28.
131. Jamieson A, Goodwill AM, Termine M et al. Depression related cerebral pathology and its relationship with cognitive functioning: a systematic review. J Affect Disord 2019;250:410-8.
132. Brayne C, Miller B. Dementia and aging populations – A global priority for contextualized research and health policy. PLoS Med 2017;14:e1002275.
133. Siu AL, US Preventive Services Task Force (USPSTF), Bibbins-Domingo K et al. Screening for depression in adults: US Preventive Services Task Force Recommendation Statement. JAMA 2016;315:380-7.
134. Fan RR, Rudnick SB, Minami HR et al. Depression screening in patients with vascular disease. Vascular 2022; doi: 10.1177/17085381221084817.
135. Breedvelt JJF, Kandola A, Kousoulis AA et al. What are the effects of preventative interventions on major depressive disorder (MDD) in young adults? A systematic review and meta-analysis of randomized controlled trials. J Affect Disord 2018;239:18-29.
136. Cuijpers P, van Straten A, Smit F et al. Preventing the onset of depressive disorders: a meta-analytic review of psychological interventions. Am J Psychiatry 2008;165:1272-80.
137. Salazar de Pablo G, Solmi M, Vaquerizo-Serrano J et al. Primary prevention of depression: an umbrella review of controlled interventions. J Affect Disord 2021;294:957-70.
138. Coventry P, Lovell K, Dickens C et al. Integrated primary care for patients with mental and physical multimorbidity: cluster randomised controlled trial of collaborative care for patients with depression comorbid with diabetes or cardiovascular disease. BMJ 2015;350:h638.
139. Terada T, Cotie LM, Tulloch H et al. Sustained effects of different exercise modalities on physical and mental health in patients with coronary artery disease: a randomized clinical trial. Can J Cardiol 2022;38:1235-43.
140. Reed JL, Terada T, Cotie LM et al. The effects of high-intensity interval training, Nordic walking and moderate-to-vigorous intensity continuous training on functional capacity, depression and quality of life in patients with coronary artery disease enrolled in cardiac rehabilitation: a randomized controlled trial (CRX study). Prog Cardiovasc Dis 2022;70:73-83.
141. Ali MK, Chwastiak L, Poongothai S et al. Effect of a collaborative care model on depressive symptoms and glycated hemoglobin, blood pressure, and se-

rum cholesterol among patients with depression and diabetes in India: the INDEPENDENT randomized clinical trial. JAMA 2020;324:651-62.
142. Srinivasan K, Heylen E, Johnson Pradeep R et al. Collaborative care compared to enhanced standard treatment of depression with co-morbid medical conditions among patients from rural South India: a cluster randomized controlled trial (HOPE Study). BMC Psychiatry 2022;22:394.
143. Goldstein BI, Carnethon MR, Matthews KA et al. Major depressive disorder and bipolar disorder predispose youth to accelerated atherosclerosis and early cardiovascular disease: a scientific statement from the American Heart Association. Circulation 2015;132:965-86.
144. Correll CU, Solmi M, Croatto G et al. Mortality in people with schizophrenia: a systematic review and meta-analysis of relative risk and aggravating or attenuating factors. World Psychiatry 2022;21:248-71.
145. Lambert AM, Parretti HM, Pearce E et al. Temporal trends in associations between severe mental illness and risk of cardiovascular disease: a systematic review and meta-analysis. PLoS Med 2022;19:e1003960.
146. Solmi M, Fiedorowicz J, Poddighe L et al. Disparities in screening and treatment of cardiovascular diseases in patients with mental disorders across the world: systematic review and meta-analysis of 47 observational studies. Am J Psychiatry 2021;178:793-803.
147. Solmi M, Firth J, Miola A et al. Disparities in cancer screening in people with mental illness across the world versus the general population: prevalence and comparative meta-analysis including 4 717 839 people. Lancet Psychiatry 2020;7:52-63.
148. Keenan TE, Yu A, Cooper LA et al. Racial patterns of cardiovascular disease risk factors in serious mental illness and the overall U.S. population. Schizophr Res 2013;150:211-6.
149. Ratliff JC, Palmese LB, Reutenauer EL et al. Obese schizophrenia spectrum patients have significantly higher 10-year general cardiovascular risk and vascular ages than obese individuals without severe mental illness. Psychosomatics 2013;54:67-73.
150. Dickerson F, Origoni A, Schroeder J et al. Mortality in schizophrenia and bipolar disorder: clinical and serological predictors. Schizophr Res 2016;170:177-83.
151. Nordentoft M, Plana-Ripoll O, Laursen TM. Cancer and schizophrenia. Curr Opin Psychiatry 2021;34:260-5.
152. Vancampfort D, Stubbs B, Mitchell AJ et al. Risk of metabolic syndrome and its components in people with schizophrenia and related psychotic disorders, bipolar disorder and major depressive disorder: a systematic review and meta-analysis. World Psychiatry 2015;14:339-47.
153. Goff DC, Sullivan LM, McEvoy JP et al. A comparison of ten-year cardiac risk estimates in schizophrenia patients from the CATIE study and matched controls. Schizophr Res 2005;80:45-53.
154. Crump C, Winkleby MA, Sundquist K et al. Comorbidities and mortality in persons with schizophrenia: a Swedish national cohort study. Am J Psychiatry 2013;170:324-33.
155. Kane JM, Kishimoto T, Correll CU. Non-adherence to medication in patients with psychotic disorders: epidemiology, contributing factors and management strategies. World Psychiatry 2013;12:216-26.
156. Pillinger T, McCutcheon RA, Vano L et al. Comparative effects of 18 antipsychotics on metabolic function in patients with schizophrenia, predictors of metabolic dysregulation, and association with psychopathology: a systematic review and network meta-analysis. Lancet Psychiatry 2020;7:64-77.
157. Solmi M, Tiihonen J, Lähteenvuo M et al. Antipsychotics use is associated with greater adherence to cardiometabolic medications in patients with schizophrenia: results from a nationwide, within-subject design study. Schizophr Bull 2022;48:166-75.
158. Oliver D, Arribas M, Radua J et al. Prognostic accuracy and clinical utility of psychometric instruments for individuals at clinical high-risk of psychosis: a systematic review and meta-analysis. Mol Psychiatry 2022; doi: 10.1038/s41380-022-01611-w.
159. Estradé A, Salazar de Pablo G, Zanotti A et al. Public health primary prevention implemented by clinical high-risk services for psychosis. Transl Psychiatry 2022;12:43.
160. Salazar de Pablo G, Radua J, Pereira J et al. Probability of transition to psychosis in individuals at clinical high risk: an updated meta-analysis. JAMA Psychiatry 2021;78:970-8.
161. Catalan A, Salazar de Pablo G, Vaquerizo Serrano J et al. Annual research review: Prevention of psychosis in adolescents – systematic review and meta-analysis of advances in detection, prognosis and intervention. J Child Psychol Psychiatry 2021;62:657-73.
162. Fusar-Poli P, Spencer T, De Micheli A et al. Outreach and support in South-London (OASIS) 2001-2020: twenty years of early detection, prognosis and preventive care for young people at risk of psychosis. Eur Neuropsychopharmacol 2020;39:111-22.
163. Fusar-Poli P, Salazar de Pablo G, Correll CU et al. Prevention of psychosis: advances in detection, prognosis, and intervention. JAMA Psychiatry 2020;77:755-65.
164. Kotlicka-Antczak M, Podgórski M, Oliver D et al. Worldwide implementation of clinical services for the prevention of psychosis: the IEPA early intervention in mental health survey. Early Interv Psychiatry 2020;14:741-50.
165. Fusar-Poli P, Correll CU, Arango C et al. Preventive psychiatry: a blueprint for improving the mental health of young people. World Psychiatry 2021;20:200-21.
166. Nieto JC, Sánchez E, Román E et al. Cytokine production in patients with cirrhosis and TLR4 polymorphisms. World J Gastroenterol 2014;20:17516-24.
167. Fuster D, Sanvisens A, Bolao F et al. Alcohol use disorder and its impact on chronic hepatitis C virus and human immunodeficiency virus infections. World J Hepatol 2016;8:1295-308.
168. Hutchinson SJ, Bird SM, Goldberg DJ. Influence of alcohol on the progression of hepatitis C virus infection: a meta-analysis. Clin Gastroenterol Hepatol 2005;3:1150-9.
169. Abassa KK, Wu XY, Xiao XP et al. Effect of alcohol on clinical complications of hepatitis virus-induced liver cirrhosis: a consecutive ten-year study. BMC Gastroenterol 2022;22:130.
170. Curry SJ, Krist AH, Owens DK et al. Screening and behavioral counseling interventions to reduce unhealthy alcohol use in adolescents and adults: US Preventive Services Task Force Recommendation Statement. JAMA 2018;320:1899-909.
171. Solmi M, Radua J, Olivola M et al. Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. Mol Psychiatry 2022;27:281-95.
172. Fusar-Poli P, Manchia M, Koutsouleris N et al. Ethical considerations for precision psychiatry: a roadmap for research and clinical practice. Eur Neuropsychopharmacol 2022;63:17-34.
173. O'Neil A, Jacka FN, Quirk SE et al. A shared framework for the common mental disorders and non-communicable disease: key considerations for disease prevention and control. BMC Psychiatry 2015;15:15.
174. McGorry PD, Mei C, Chanen A et al. Designing and scaling up integrated youth mental health care. World Psychiatry 2022;21:61-76.
175. Perera S, Hetrick S, Cotton S et al. Awareness of headspace youth mental health service centres across Australian communities between 2008 and 2015. J Ment Health 2020;29:410-7.
176. Malla A, Boksa P, Joober R. Meeting the challenges of the new frontier of youth mental health care. World Psychiatry 2022;21:78-9.
177. Kieling C, Salum GA, Pan PM et al. Youth mental health services: the right time for a global reach. World Psychiatry 2022;21:86-7.

DOI:10.1002/wps.21068

Cognitive behavior therapy vs. control conditions, other psychotherapies, pharmacotherapies and combined treatment for depression: a comprehensive meta-analysis including 409 trials with 52,702 patients

Pim Cuijpers[1-3], Clara Miguel[1], Mathias Harrer[4,5], Constantin Yves Plessen[1,6], Marketa Ciharova[1], David Ebert[4], Eirini Karyotaki[1]

[1]Department of Clinical, Neuro and Developmental Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; [2]Babeş-Bolyai University, International Institute for Psychotherapy, Cluj-Napoca, Romania; [3]WHO Collaborating Centre for Research and Dissemination of Psychological Interventions, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; [4]Psychology & Digital Mental Health Care, Department of Health Sciences, Technical University Munich, Munich, Germany; [5]Department of Clinical Psychology & Psychotherapy, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany; [6]Department of Psychosomatic Medicine, Charité Universitätsmedizin Berlin, Berlin, Germany

Cognitive behavior therapy (CBT) is by far the most examined type of psychological treatment for depression and is recommended in most treatment guidelines. However, no recent meta-analysis has integrated the results of randomized trials examining its effects, and its efficacy in comparison with other psychotherapies, pharmacotherapies and combined treatment for depression remains uncertain. We searched PubMed, PsycINFO, Embase and the Cochrane Library to identify studies on CBT, and separated included trials into several subsets to conduct random-effects meta-analyses. We included 409 trials (518 comparisons) with 52,702 patients, thus conducting the largest meta-analysis ever of a specific type of psychotherapy for a mental disorder. The quality of the trials was found to have increased significantly over time (with increasing numbers of trials with low risk of bias, less waitlist control groups, and larger sample sizes). CBT had moderate to large effects compared to control conditions such as care as usual and waitlist (g=0.79; 95% CI: 0.70-0.89), which remained similar in sensitivity analyses and were still significant at 6-12 month follow-up. There was no reduction of the effect size of CBT according to the publication year (<2001 vs. 2001-2010 vs. >2011). CBT was significantly more effective than other psychotherapies, but the difference was small (g=0.06; 95% CI: 0-0.12) and became non-significant in most sensitivity analyses. The effects of CBT did not differ significantly from those of pharmacotherapies at the short term, but were significantly larger at 6-12 month follow-up (g=0.34; 95% CI: 0.09-0.58), although the number of trials was small, and the difference was not significant in all sensitivity analyses. Combined treatment was more effective than pharmacotherapies alone at the short (g=0.51; 95% CI: 0.19-0.84) and long term (g=0.32; 95% CI: 0.09-0.55), but it was not more effective than CBT alone at either time point. CBT was also effective as unguided self-help intervention (g=0.45; 95% CI: 0.31-0.60), in institutional settings (g=0.65; 95% CI: 0.21-1.08), and in children and adolescents (g=0.41; 95% CI: 0.25-0.57). We can conclude that the efficacy of CBT in depression is documented across different formats, ages, target groups, and settings. However, the superiority of CBT over other psychotherapies for depression does not emerge clearly from this meta-analysis. CBT appears to be as effective as pharmacotherapies at the short term, but more effective at the longer term.

Key words: Depression, cognitive behavior therapy, psychotherapies, Internet-based interventions, meta-analysis, antidepressants, combined treatment

(World Psychiatry 2023;22:105–115)

Depression is a highly prevalent mental disorder, with about 280 million people worldwide suffering from it[1]. The disorder results in considerable loss of quality of life in patients and their families[2], and is associated with increased physical morbidity and premature mortality[3], a considerable disease burden at the population level[1], and enormous economic costs[4]. Several evidence-based interventions are available for the treatment of depression, including pharmacotherapies[5] and psychotherapies[6].

Cognitive behavior therapy (CBT) is by far the most examined type of psychological treatment for depression and is recommended in most treatment guidelines. Several hundreds of randomized controlled trials have tested the effects of CBT[6,7]. Previous meta-analyses have found that CBT is significantly more effective in the treatment of depression than various control conditions[6-8], whereas its effectiveness in comparison with other psychotherapies, pharmacotherapies and combined treatment at the short and longer term, as well as its impact on specific populations of patients and in different formats, remain uncertain[9].

The last comprehensive meta-analysis of CBT for depression was published in 2013[8], while the number of trials has increased exponentially over the years, and many new trials have been published since then. Furthermore, that meta-analysis did not include trials in children/adolescents and inpatients, as well as comparisons with pharmacotherapies and combined treatments, with other psychotherapies, and with unguided digital interventions. More recent meta-analyses have focused on psychological interventions in general, including CBT[6,7], but they have not examined specific characteristics of the participants, the treatment and the study as predictors of outcome.

We decided, therefore, to conduct a new, comprehensive meta-analysis of randomized trials examining the short- and long-term effects of CBT in depression across all treatment formats (i.e., individual, group, unguided and guided self-help), all ages (including children and adolescents), delivered in any setting (including outpatients and inpatients), and compared against control conditions (e.g., waitlist, care as usual) as well as other active treatments (i.e., other psychotherapies, antidepressant medications, and combined treatment).

METHODS

Identification and selection of trials

This study is part of a larger meta-analytic project on psychological treatments for depression[10]. The protocol for the current meta-analysis has been published in the Open Science Framework (http://osf.io/a6p3w).

The trials included in this study were identified through a database which is continuously updated, currently including studies from 1966 to January 1, 2022. For this database, we searched PubMed, PsycINFO, Embase and the Cochrane Library, by combining index and free terms indicative of depression and psychotherapies, with filters for randomized controlled trials. The full search strings can be found in the supplementary information. Furthermore, we checked references of earlier meta-analyses on psychological treatments for depression.

Two independent researchers screened all records, and all papers that could meet inclusion criteria according to one of them were retrieved as full text. The two independent researchers also decided to include or exclude a study in the database, and disagreements were resolved through discussion.

For the current study, we selected randomized controlled trials in which CBT for people with depression was compared with control conditions (care as usual, waitlist, others), other psychotherapies, pharmacotherapies, or combined treatment.

A broad definition of CBT was used: a treatment in which the therapist focuses on the impact of present dysfunctional thoughts on a patient's current behavior and future functioning, and which is aimed at evaluating, challenging and modifying a patient's dysfunctional beliefs (cognitive restructuring). Cognitive restructuring could be combined with other mood management skills, such as behavioral activation, problem-solving, social skills training, or mindfulness. This definition was derived from an extensive study in which different types of psychotherapies were examined by multiple researchers, resulting in a consensus on the definition of each therapy[11].

Depression could be defined as meeting the criteria for a depressive disorder according to a diagnostic interview or as a score above the cut-off on a self-report depression measure. We included trials in which CBT was administered in any format (individual, group, telephone, guided or unguided self-help). We also included trials of outpatients as well as inpatients, and in any age group.

We separated the included studies into several subsets, so that the comparisons from these studies could be pooled in a meta-analysis. In the largest subset, CBT was compared with control conditions. In this subset, we included CBT that was applied individually, in groups, as guided self-help, or in a mixed format, because previous research has shown that these formats have comparable effects[12]. Studies of unguided self-help CBT were included in a separate subset. We also created a separate subset for CBT in inpatients, because these patients differ from outpatients, and the control conditions vary considerably from outpatient settings[13]. A separate subset was also built for studies comparing CBT with pharmacotherapies, CBT with combined treatment, and pharmacotherapies with combined treatment. We created a separate subset for depression in children and adolescents, because therapies usually are less effective in this group.

Quality assessment and data extraction

We assessed the validity of included studies using four criteria of the Risk of Bias (RoB) assessment tool, version 1, developed by the Cochrane Collaboration[14,15]. The RoB tool assesses possible sources of bias in randomized trials, including the adequate generation of allocation sequence; the concealment of allocation to conditions; the prevention of knowledge of the allocated intervention (masking of assessors); and dealing with incomplete outcome data (this was assessed as positive when intention-to-treat analyses were conducted, meaning that all randomized patients were included in the analyses). Two independent researchers evaluated the validity of the included studies, and disagreements were solved through discussion.

We also coded participant characteristics (diagnostic method, recruitment method, target group, mean age, proportion of women, inpatient or outpatient); characteristics of CBT (treatment format, number of sessions), as well as general characteristics of the studies (type of comparison group, publication year, country where the study was conducted). In the studies in which CBT was compared with other therapies, we also categorized the other therapies according to the definitions provided elsewhere[7]. In studies with pharmacotherapies, we also categorized the type of antidepressant: selective serotonin reuptake inhibitor (SSRI), tricyclic antidepressant (TCA), other.

Outcome measures

For each comparison between a psychological treatment and a control condition, the effect size indicating the difference between the two groups at post-test was calculated (Hedges' g)[16]. Effect sizes were calculated by subtracting (at post-test) the average score of the psychotherapy group from the average score of the control group and dividing the result by the pooled standard deviation. Because some studies were expected to have relatively small sample sizes, we corrected the effect size for small sample bias.

When the means and standard deviations were not reported in a study, we used change scores. If these were not reported, we converted binary outcomes to Hedges' g. If these were also not reported, we used other statistics (e.g., p value, t value) to calculate the effect size.

Meta-analyses

To make a historical overview of trials on CBT over time, we conducted bivariable linear regression analyses examining if the characteristics of the trials have changed over time. We limited these analyses to the subset comparing CBT with control condi-

tions, because this was the largest and most homogeneous subset.

The meta-analyses were conducted using the metapsyTools package in R (version 4.1.1) and Rstudio (version 1.1.463 for Mac)[17]. The metapsyTools package was specifically developed for the meta-analytic project of which this study is part. This package imports the functionality of the meta[18], metafor[19], and dmetar[20] packages.

We calculated the pooled effect sizes in several different ways, as implemented in the metapsyTools package, so that we could explore if different pooling methods resulted in different outcomes. In our main model, all effect size data available for a comparison in a specific study were aggregated within that comparison first. These aggregated effects were then pooled across studies and comparisons. An intra-study correlation coefficient of ρ=0.5 was assumed to aggregate effects within comparisons.

We conducted several other analyses to examine whether these main outcomes were robust. First, we estimated the pooled effect using a three-level correlated and hierarchical effects (CHE) model[21]. We assumed an intra-study correlation of ρ=0.5 for this model. Second, we pooled effects while excluding outliers, using the "non-overlapping confidence intervals" approach, in which a study is defined as an outlier when the 95% confidence interval (CI) of the effect size does not overlap with the 95% CI of the pooled effect size[20]. Third, we pooled effects while excluding influential cases, defined by the diagnostics proposed by Viechtbauer and Cheung[22]. Fourth, we calculated the effect when the smallest or largest effect in each study was considered. Fifth, we estimated the pooled effect using only studies with a low risk of bias. We also used three different methods to assess and adjust for potential publication bias[20,23]: Duval and Tweedie's trim and fill procedure[24], Rücker's "limit meta-analysis method"[25], and the selection model[26,27].

A random-effects model was assumed for all analyses. Between-study heterogeneity variance (components) was estimated using restricted maximum likelihood. For models not fitted using robust variance estimation, we applied the Knapp-Hartung method to obtain robust CIs and significance tests of the overall effect[28].

As a test of homogeneity of effect sizes, we calculated the I^2-statistic and its 95% CI, which is an indicator of heterogeneity in percentages. A value of 0% indicates no observed heterogeneity, and larger values indicate increasing heterogeneity, with 25% as low, 50% as moderate, and 75% as high heterogeneity[29]. For the three-level model, we calculated a multilevel extension of I^2, which describes the amount of total variability attributable to heterogeneity within studies (level 2) and heterogeneity between studies (level 3)[20,30]. Because I^2 cannot be interpreted as an absolute measure of the between-study heterogeneity, we also added the prediction interval (PI) to the main analyses, which indicates the range in which the true effect size of 95% of all populations will fall[31,32].

We also estimated the number-needed-to-treat (NNT) for depression using the formulae provided by Furukawa[33] (assuming the control group's event rate at a conservative 17%)[34].

For the main comparison (CBT versus control conditions), we also extracted the rate of response (i.e., a 50% reduction of depressive symptoms compared to baseline). If the response rate was not reported, we estimated it using a method based on the baseline means, the post-test means, the post-test standard deviations and the number of subjects[35]. For studies using the Hamilton Rating Scale for Depression (HAM-D), we also calculated the rate of remission, defined as a score of ≤7 on the 17-item version of that scale[36]. We also calculated the relative risk (RR) for response and remission of CBT compared with the control groups, as well as the NNT (as 1 divided by the risk difference).

In each subset, we conducted a series of subgroup analyses, examining the effects of the interventions according to major characteristics of the participants, interventions and studies. We avoided subgroups of less than five studies, merging them with other subgroups. Because the subset comparing CBT with control conditions was very large, we also conducted a multivariable meta-regression analysis in which all characteristics were included.

RESULTS

Selection and inclusion of studies

After examining a total of 30,889 records (21,563 after removal of duplicates), we retrieved 3,584 full-text papers for further consideration. A total of 409 trials met the inclusion criteria for this meta-analysis (see Figure 1). Selected characteristics of included studies and comparisons are presented in the supplementary information.

Characteristics of included studies

The 409 studies (518 comparisons between CBT and a control condition) included 52,702 patients (27,000 in CBT and 25,702 in control groups). Aggregated characteristics of the studies and comparisons are provided in Table 1.

Most studies recruited participants through the community (n=181, 44.3%) or clinical referrals (n=106, 25.9%). In most studies, the target group was represented by adults in general (n=160, 39.1%); 70 studies aimed at patients with general medical disorders (17.1%), 41 studies at perinatal depression (10.0%), and 27 studies at children or adolescents (9.0%).

In the majority of studies (n=226, 55.3%), depression was defined as meeting the criteria for a depressive disorder according to a diagnostic interview, while in 162 studies (39.3%) it was defined as a score above the cut-off on a self-report depression measure. The mean age of participants in the studies was 40.1±14.98 years; the average proportion of women was 69%. Most studies were conducted in the US (n=141, 34.5%) or in the UK or other European countries (n=141, 34.5%). Most studies (n=249, 60.8%) were published since 2011.

Among the 518 comparisons, the majority tested an individual CBT format (n=206, 39.8%), while 141 examined a group format (27.2%), 84 a guided self-help format (16.2%), and 39 an unguided self-help format (7.5%). In 211 comparisons (40.7%), CBT was administered in more than 12 sessions.

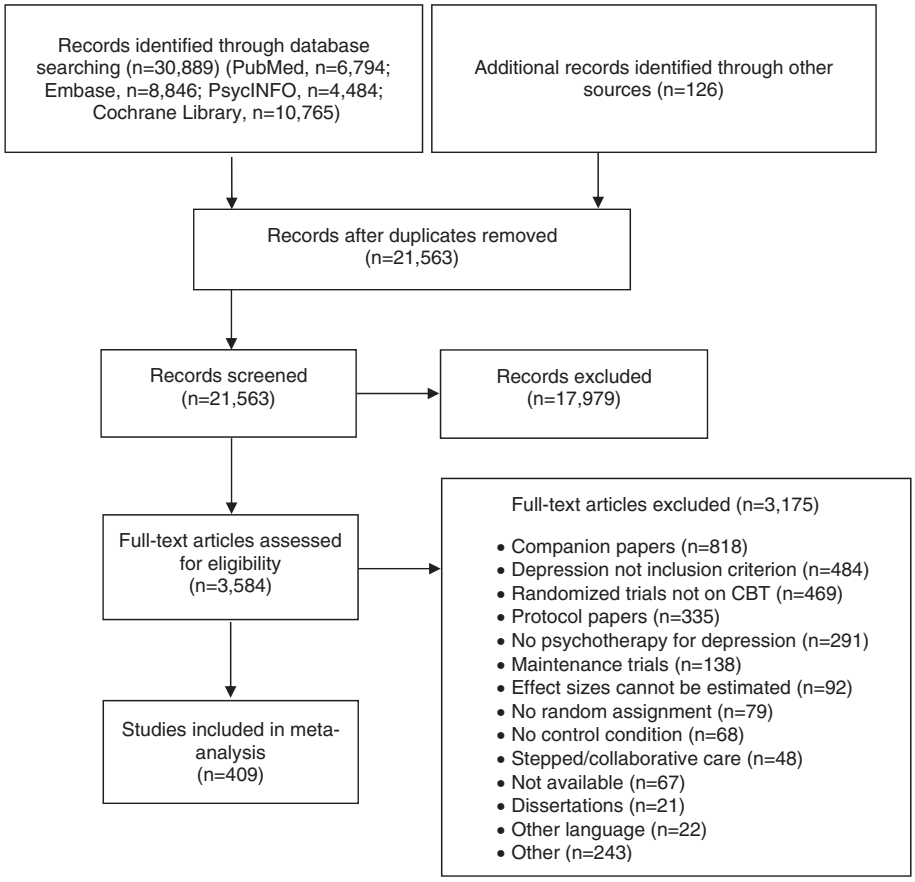

Figure 1 PRISMA flow chart, CBT – cognitive behavior therapy

Of the 409 studies, 224 (54.8%) reported an adequate generation of allocation sequence, 201 (49.1%) an adequate concealment of allocation to conditions, and 101 (24.7%) an adequate prevention of knowledge of the allocated intervention (masking of assessors); 262 (64.1%) conducted intention-to-treat analyses. Risk of bias was low across all four domains in 131 studies (32.0%), for two or three domains in 173 studies (42.2%), and for no or one domain in 105 studies (25.7%).

Historical overview

The historical overview was limited to the subset comparing CBT with control conditions (241 studies with 271 comparisons, including 12,907 patients in CBT arms and 12,199 in control conditions). The cumulative number of studies over time is shown in Figure 2.

The bivariable linear regression analyses found that the number of trials examining depressed patients with general medical disorders and women with perinatal depression increased significantly over time (p=0.007 and p=0.012, respectively). The use of waitlist as the control condition decreased significantly over time (p=0.001), while the number of studies with low risk of bias increased significantly (p<0.001), as well as the number of trials in non-Western countries (p=0.005). The number of participants in each comparison also increased significantly (p<0.001), while the number of sessions of CBT decreased significantly over time (p=0.03). All the other characteristics of CBT trials did not change over time (see also supplementary information).

CBT versus control conditions

The main effect size indicating the overall difference between CBT and control conditions after treatment was g=0.79 (95% CI: 0.70-0.89), corresponding to an NNT of 3.8 (see Table 2). Heterogeneity was very high (I^2=85; 95% CI: 83-86), and the prediction interval ranged from –0.45 to 2.04.

The sensitivity analyses supported the main findings (see Table 2 and supplementary information). Heterogeneity was considerably lower after excluding outliers (I^2=26; 95% CI: 11-39), but the number of outliers that had to be removed was large (n=77). The effect size was smaller for studies with low risk of bias (g=0.60; 95% CI: 0.49-0.71) and after adjusting for publication bias (g=0.47, 95% CI: 0.35-0.59 using the trim and fill procedure).

The subgroup analyses indicated that the effect size in studies with low risk of bias was significantly lower than in other studies (p<0.001), and that the effect size differed across countries (higher in non-Western countries; p=0.003) and treatment formats (higher

Table 1 Aggregated characteristics of included studies and comparisons

Included studies (n=409)		
Recruitment, n (%)	Community	181 (44.3)
	Clinical	106 (25.9)
	Other	122 (29.6)
Target group, n (%)	Children	12 (2.9)
	Adolescents	25 (6.1)
	Adults	160 (39.1)
	Elderly	26 (6.4)
	General medical	70 (17.1)
	Perinatal	41 (10.0)
	Other	75 (18.3)
Age, years (mean±SD)		40.1±15.0
Gender (% female)		69.0
Diagnosis, n (%)	Meeting criteria for depressive disorder	226 (55.3)
	Score above cut-off on self-report depression measure	162 (39.3)
	Other	21 (5.1)
Country, n (%)	US	141 (34.5)
	UK	44 (10.8)
	Other European countries	97 (23.7)
	Australia	33 (8.1)
	Canada	25 (6.1)
	East Asia	30 (7.3)
	Other	39 (9.5)
Year of publication, n (%)	<1980	4 (1.0)
	1981-1990	32 (7.8)
	1991-2000	41 (10.0)
	2001-2010	83 (20.3)
	2011-2020	219 (53.5)
	2021	30 (7.3)
Overall risk of bias (RoB), n (%)	0 (high)	20 (4.9)
	1	85 (20.8)
	2	73 (17.8)
	3	100 (24.4)
	4 (low)	131 (32.0)
RoB: Adequate sequence generation, n (%)		224 (54.8)
RoB: Adequate allocation concealment, n (%)		201 (49.1)
RoB: Adequate masking of assessors, n (%)		101 (24.7)
RoB: Intention-to-treat analyses, n (%)		262 (64.1)
Included comparisons (n=518)		
Format, n (%)	Individual	206 (39.8)
	Group	141 (27.2)
	Guided self-help	84 (16.2)
	Unguided self-help	39 (7.5)
	Other/mixed	48 (9.3)
Number of sessions, n (%)	<8	120 (23.2)
	8-12	141 (27.2)
	>12	211 (40.7)
	Not reported/relevant	46 (8.9)

for group formats; p=0.02). There was no reduction of the effect size of CBT according to the publication year (<2001 vs. 2001-2010 vs. >2011) (p=0.43). We entered all variables in a multivariable meta-regression analysis and found that, after adjustment for all variables, only the use of a waitlist control condition (p=0.02) and whether the trial was conducted in an "other" country (not the US, Europe, East Asia, Canada or Australia; p=0.001) had a significant impact on the effect size (see supplementary information).

CBT was still effective at 6 to 9 month follow-up (g=0.74, 95% CI: 0.36-1.11) and at 10 to 12 month follow-up (g=0.49, 95% CI: 0.01-0.98), and this was confirmed in most sensitivity analyses (see Table 2 and supplementary information). Heterogeneity was high in most analyses. At 13 to 24 month follow-up, the main effect size was no longer significant (g=0.22, 95% CI: –0.12 to 0.56), although this may be related to the small number of studies (n=8).

The response rate was 0.42 (95% CI: 0.39-0.45) in CBT and 0.19 (95% CI: 0.18-0.21) in the control conditions, which resulted in a RR of 2.13 (95% CI: 1.96-2.32) and a NNT of 4.7 (95% CI: 4.0-5.5) in favor of CBT (see Table 3). Most sensitivity analyses indicated similar outcomes, except that there was significant publication bias, and the RR was lower in studies with low risk of bias. The response rates differed significantly across control conditions, with the lowest rate for waitlist (see Table 3 and supplementary information).

The remission rate was 0.36 (95% CI: 0.31-0.42) for CBT and 0.15 (0.12-0.18) for control conditions, which resulted in a RR of 2.45 (95% CI: 2.06-2.92), and a NNT of 3.6 (95% CI: 2.7-5.0). This rate remained very similar in the sensitivity analyses, although it was somewhat lower (but still significant) after adjustment for publication bias. These findings should be considered with caution, because the difference between reported and estimated remission rates was significant (p=0.02) (see Table 3 and supplementary information).

CBT versus other psychotherapies

CBT was compared with other psychotherapies in 87 studies (82 comparisons; 6,480 participants, including 3,148 in CBT and 3,332 in the other therapies). The main analyses indicated a very small, but significant effect of CBT over other therapies (g=0.06; 95% CI: 0-0.12; NNT=63), with low heterogeneity (I^2=31; 95% CI: 10-47) (see Table 4).

When limiting the studies to those with low risk of bias, or excluding outliers, or after adjustment for publication bias, the difference between CBT and other psychotherapies was no longer significant. In the subgroup analyses in which we examined the

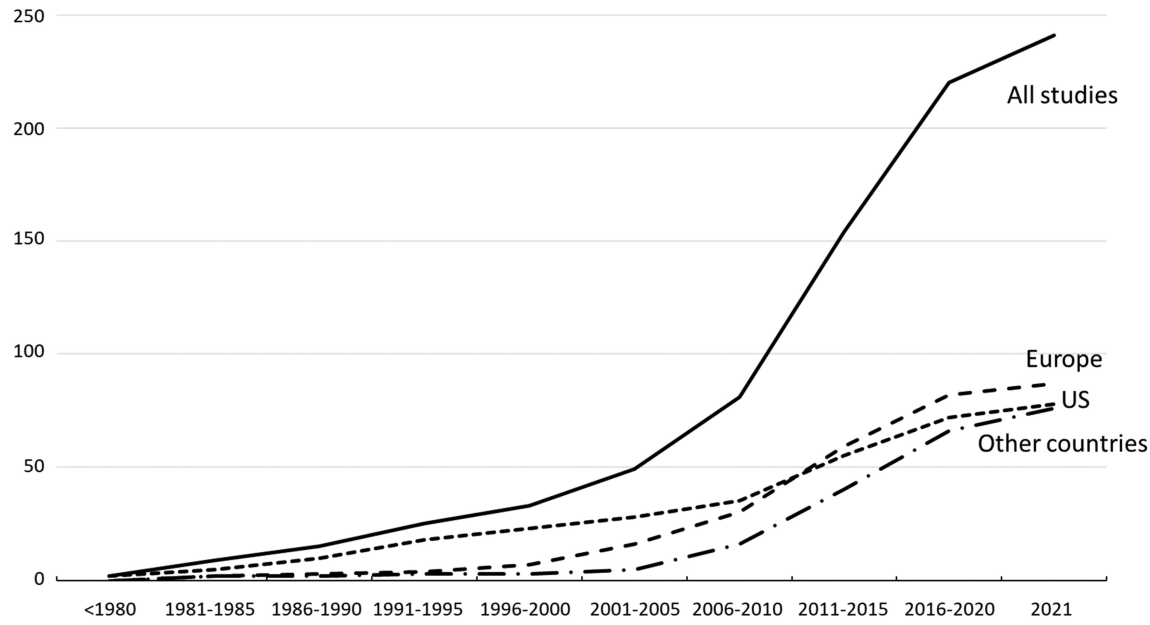

Figure 2 Randomized trials comparing cognitive behavior therapy (CBT) with control conditions: cumulation over time

different psychotherapies that were compared with CBT, we found no indication that one of these therapies was more or less effective than CBT (see Table 4 and supplementary information).

The number of studies reporting longer-term outcomes was small, and no significant differences between CBT and other psychotherapies were found at 6-9 months, 9-12 months, or 13-24 months (see Table 4 and supplementary information).

CBT versus pharmacotherapies and combined treatment

CBT was compared with pharmacotherapies in 38 studies (38 comparisons; 2,979 participants, including 1,459 in CBT groups and 1,520 in pharmacotherapy groups). No significant difference was found between CBT and pharmacotherapies (g=0.08; 95% CI: –0.07 to 0.24). The same was observed in sensitivity analyses, although one of the analyses examining publication bias indicated a small, but significant effect in favor of pharmacotherapies. None of the subgroup analyses pointed at a significant difference between subgroups of studies (see Table 4 and supplementary information).

At 6 to 12 month follow-up, CBT was more effective than pharmacotherapies (g=0.34; 95% CI: 0.09-0.58; NNT=10.2). This was confirmed in most sensitivity analyses, although the number of studies with low risk of bias was small and the effect size no longer significant. In two of the three analyses adjusting for publication bias, this finding was also not significant anymore (see Table 4 and supplementary information).

Combined treatment was compared with pharmacotherapy alone in 18 studies (18 comparisons; 1,658 participants, including 827 in the combined and 831 in the pharmacotherapy conditions). Combined treatment was more effective than pharmacotherapy (g=0.51; 95% CI: 0.19-0.84) and that was confirmed in most sensitivity analyses, although the number of trials with low risk of bias was small. After adjustment for publication bias, the effects were no longer significant. No significant differences were found in subgroup analyses (see Table 4 and supplementary information).

Combined treatment was not significantly more effective than CBT alone (g=0.19; 95% CI: –0.11 to 0.50) in the 15 relevant studies (14 comparisons; 644 participants, including 325 in the combined and 319 in the CBT only conditions). Only one of three analyses in which we adjusted for publication bias resulted in a significant effect size in favor of combined treatment. Because of the limited number of trials, we could only conduct a limited number of subgroup analyses, and none of them resulted in significant differences between subgroups (see Table 4 and supplementary information).

At 6 to 12 month follow-up, combined treatment was more effective than pharmacotherapy alone (g=0.32, 95% CI: 0.09-0.55), but this finding was not confirmed in all sensitivity analyses. Combined treatment was not more effective than CBT alone (g=0.11; 95% CI: –0.38 to 0.60) (see Table 4 and supplementary information).

Other comparisons

Unguided self-help CBT (Internet-based or not) had a small to moderate effect on depression (g=0.45; 95% CI: 0.31-0.60), based on 36 studies (39 comparisons; 11,720 participants, including 6,206 in the CBT and 5,514 in the control conditions). The effects of unguided CBT were significant in all sensitivity analyses, although they were somewhat smaller in two of three analyses adjusting for publication bias. Subgroup analyses indicated that waitlist-controlled trials resulted in larger effect sizes (p=0.03), and studies in Europe resulted in smaller effects (p=0.01). We also found that studies conducted after 2011 had significantly larger effects than

Table 2 Cognitive behavior therapy (CBT) vs. control conditions: main analyses

	n	g (95% CI)	I^2 (95% CI)	PI	NNT
Post-test					
All comparisons	271	0.79 (0.70-0.89)	85 (83-86)	–0.45 to 2.04	3.8
Outliers removed	194	0.70 (0.65-0.74)	26 (11-39)	0.49 to 0.90	4.4
Only low risk of bias	90	0.60 (0.49-0.71)	77 (72-81)	–0.22 to 1.42	5.2
Three-level model	460	0.81 (0.72-0.90)	90 (-)	–0.56 to 2.17	3.7
Publication bias correction	349	0.47 (0.35-0.59)	90 (89-91)	–1.52 to 2.46	7.0
6-9 month follow-up					
All comparisons	78	0.74 (0.36-1.11)	91 (89-92)	–1.90 to 3.37	4.1
Outliers removed	65	0.42 (0.33-0.50)	63 (51-72)	–0.10 to 0.93	8.0
Only low risk of bias	29	0.91 (0.46-1.36)	94 (92-95)	–1.46 to 3.28	3.2
Three-level model	119	0.74 (0.40-1.08)	98 (-)	–2.17 to 3.65	4.1
Publication bias correction	93	0.30 (–0.23 to 0.83)	94 (93-95)	–4.31 to 4.91	11.4
10-12 month follow-up					
All comparisons	22	0.49 (0.01-0.98)	91 (88-93)	–1.68 to 2.67	6.5
Outliers removed	20	0.22 (0.10-0.35)	74 (59-83)	–0.25 to 0.70	16.0
Only low risk of bias	4	0.28 (–0.25 to 0.82)	87 (68-94)	–1.29 to 1.86	12.3
Three-level model	30	0.50 (0.03-0.96)	97 (-)	–1.65 to 2.64	6.5
Publication bias correction	22	0.49 (0.01-0.98)	91 (88-93)	–1.68 to 2.67	6.5
13-24 month follow-up					
All comparisons	8	0.22 (–0.12 to 0.56)	86 (75-93)	–0.77 to 1.21	16.2
Outliers removed	7	0.09 (–0.10 to 0.27)	11 (0-74)	–0.24 to 0.42	42.9
Only low risk of bias	3	–0.01 (–0.17 to 0.16)	0 (0-90)	–1.20 to 1.18	416.3
Three-level model	13	0.22 (–0.14 to 0.59)	80 (-)	–0.68 to 1.13	16.0
Publication bias correction	11	0.44 (0.09-0.80)	89 (83-93)	–0.71 to 1.60	7.4

PI – prediction interval, NNT – number needed to treat. The reported publication bias correction is that using the trim and fill procedure.

earlier studies (p=0.01), suggesting that the effects may have improved over time (see Table 5 and supplementary information).

We could compare CBT in institutional settings to control conditions in 10 studies (11 comparisons; 448 participants, including 275 in CBT and 173 in the control conditions). Five studies (six comparisons) were conducted in psychiatric inpatient settings, four in nursing homes, and one in another institutional setting. None of the trials was rated as at low risk of bias. We found a moderate to large effect (g=0.65; 95% CI: 0.21-1.08) with high heterogeneity, which remained significant in most sensitivity analyses, but was no longer significant in two of the three analyses adjusting for publication bias (see Table 5 and supplementary information). Because of the small number of trials and the low quality, we did not conduct subgroup analyses.

In children and adolescents, CBT was compared to control conditions in 37 studies (39 comparisons; 3,667 participants, including 1,859 in CBT and 1,808 in control groups). We found a moderate effect (g=0.41; 95% CI: 0.25-0.57; NNT=8.1), with high heterogeneity (I^2=78; 95% CI: 70-84). The effect size remained similar across most sensitivity analyses. The number of studies with low risk of bias was low and the effect size was no longer significant in this subset. One of the effect sizes adjusted for publication bias was also not significant (see Table 5 and supplementary information). In the subgroup analyses, we found that waitlist control groups resulted in significantly larger effect sizes than other control conditions (p=0.01), and studies with low risk of bias resulted in significantly lower effect sizes than other studies (p=0.04).

DISCUSSION

This is the largest meta-analysis ever of a specific type of psychotherapy for a mental disorder, including 409 RCTs (518 comparisons) with 52,702 patients. CBT was found to be effective in depression when compared to control conditions such as usual care and waitlist, with a moderate to large effect size (g=0.79). This effect was robust in several sensitivity analyses, although it was somewhat smaller for studies with low risk of bias (g=0.60) and

Table 3 Cognitive behavior therapy (CBT) vs. control conditions: response and remission rates, relative risk (RR) and number-needed-to-treat (NNT)

	n	Rate (95% CI)	I^2 (95% CI)	RR (95% CI)	I^2 (95% CI)	NNT (95% CI)
Response						
All CBT conditions	238	0.42 (0.39-0.45)	82 (79-84)	2.13 (1.96-2.32)	47 (38-54)	4.7 (4.0-5.5)
Reported	10	0.42 (0.28-0.59)	91 (85-94)	2.32 (1.43-3.77)	46 (0-74)	4.0 (1.9-12.2)
Estimated	228	0.42 (0.39-0.45)	81 (79-83)	2.13 (1.95-2.32)	47 (38-54)	4.7 (4.0-5.5)
Outliers excluded	162	0.42 (0.40-0.43)	31 (16-43)	2.25 (2.07-2.44)	10 (0-25)	4.2 (3.7-4.9)
Publication bias correction	259	0.39 (0.36-0.42)	84 (82-85)	1.66 (1.48-1.85)	59 (54-64)	8.0 (6.2-11.0)
Low risk of bias	78	0.39 (0.35-0.44)	86 (83-88)	1.84 (1.64-2.07)	40 (21-54)	6.3 (4.9-8.2)
All control groups	238	0.19 (0.18-0.21)	67 (63-72)			
Reported	10	0.17 (0.10-0.25)	73 (48-86)			
Estimated	228	0.19 (0.18-0.21)	67 (62-71)			
Outliers excluded	192	0.19 (0.18-0.20)	14 (0-29)			
Publication bias correction	310	0.24 (0.22-0.26)	72 (68-75)			
Low risk of bias	78	0.21 (0.18-0.24)	73 (66-78)			
Type: Waitlist*	110	0.17 (0.15-0.19)	50 (38-60)			
Type: Care as usual	104	0.21 (0.18-0.24)	75 (70-79)			
Type: Other control	24	0.23 (0.19-0.26)	60 (37-74)			
Remission						
All CBT conditions	69	0.36 (0.31-0.42)	80 (75-84)	2.45 (2.06-2.92)	26 (0-45)	3.6 (2.7-5.0)
Reported**	10	0.49 (0.38-0.60)	73 (48-86)	2.36 (1.71-3.25)	18 (0-59)	3.9 (2.3-7.4)
Estimated	59	0.34 (0.29-0.40)	80 (75-85)	2.47 (2.01-3.03)	26 (0-47)	3.6 (2.6-5.2)
Outliers excluded	49	0.36 (0.33-0.39)	48 (27-63)	2.47 (2.08-2.93)	10 (0-34)	3.6 (2.7-4.9)
Publication bias correction	80	0.43 (0.37-0.50)	83 (80-86)	1.83 (1.44-2.31)	41 (24-54)	6.3 (4.0-12.0)
Low risk of bias	14	0.33 (0.22-0.47)	87 (80-92)	2.17 (1.57-2.99)	30 (0-63)	4.5 (2.6-9.2)
All control groups	69	0.15 (0.12-0.18)	70 (61-76)			
Reported	10	0.19 (0.13-0.29)	66 (34-83)			
Estimated	59	0.14 (0.10-0.18)	71 (62-77)			
Outliers excluded	56	0.14 (0.12-0.16)	29 (1-49)			
Publication bias correction	98	0.24 (0.19-0.31)	75 (70-80)			
Low risk of bias	14	0.18 (0.13-0.24)	67 (41-81)			
Type: Waitlist	34	0.12 (0.08-0.17)	59 (40-72)			
Type: Care as usual	27	0.15 (0.10-0.21)	76 (65-83)			
Type: Other control	8	0.21 (0.15-0.29)	71 (40-86)			

*difference among types of control conditions, p=0.006, **difference between reported and estimated remission rates, p=0.02

after adjustment for publication bias (g=0.47). CBT was still significantly effective at 6-9 month (g=0.74) and 10-12 month (g=0.49) follow-up, and this was confirmed in most sensitivity analyses.

A total of 42% of patients receiving CBT responded to treatment, while the response rate was only 19% in control groups, with a NNT of 4.7 in favor of CBT. The remission rate was 36% in patients receiving CBT, compared to 15% in control conditions, with a NNT of 3.6.

Comparative trials suggest that CBT is significantly more effective than other psychotherapies, but the difference is small (g=0.06) and does not remain significant in most sensitivity analyses. The effects of CBT are comparable to those of pharmacotherapies at the short term, but CBT is significantly more effective at 6 to 12 months (g=0.34). Combined treatment is significantly more effective than pharmacotherapy alone, at the short (g=0.51) and the longer term (g=0.32), but combined treatment is not more effective than CBT alone at either time point.

Table 4 Cognitive behavior therapy (CBT) vs. other active treatments

	n	g (95% CI)	I^2 (95% CI)	NNT
CBT vs. other psychotherapies				
All studies	87	0.06 (0-0.12)	31 (10-47)	63
Outliers removed	81	0.04 (–0.01 to 0.09)	1 (0-27)	93.9
Only low risk of bias	24	0.02 (–0.05 to 0.09)	0 (0-45)	200.4
Publication bias correction	92	0.04 (–0.03 to 0.11)	44 (28-56)	93.4
Long-term effect (at 6-9 months)	18	–0.03 (–0.14 to 0.07)	0 (0-50)	117.2
Long-term effect (at 9-12 months)	14	–0.09 (–0.19 to 0.01)	12 (0-50)	47.7
Compared to supportive therapy	22	0.12 (–0.07 to 0.31)	54 (26-72)	31.2
Compared to interpersonal therapy	9	0.00 (–0.12 to 0.12)	0 (0-65)	18.0
Compared to psychodynamic therapy	7	0.21 (–0.10 to 0.52)	47 (0-78)	17.1
Compared to behavioral activation	10	0.02 (–0.17 to 0.20)	28 (0-66)	196.6
Compared to 3rd wave therapies	2	–0.05 (–1.21 to 1.11)	0 (-)	81.0
Compared to problem-solving therapy	2	0.12 (–0.21 to 0.44)	0 (-)	31.2
Compared to other psychotherapies	35	0.05 (–0.04 to 0.14)	23 (0-49)	77.2
CBT vs. pharmacotherapies				
All studies	38	0.08 (–0.07 to 0.24)	66 (52-76)	46.1
Outliers removed	32	–0.03 (–0.13 to 0.07)	34 (0-57)	135.0
Only low risk of bias	8	–0.06 (–0.38 to 0.27)	66 (29-84)	70.6
Publication bias correction	44	–0.05 (–0.25 to 0.15)	76 (68-82)	81.7
Long-term effect (at 6-12 months)	12	0.34 (0.09-0.58)	53 (10-76)	10.2
Combined treatment vs. pharmacotherapy alone				
All studies	18	0.51 (0.19-0.84)	71 (53-82)	6.3
Outliers removed	16	0.41 (0.23-0.60)	49 (8-71)	8.1
Only low risk of bias	5	0.27 (–0.42 to 0.96)	77 (43-90)	13.1
Publication bias correction	21	0.34 (–0.08 to 0.76)	79 (68-86)	10.1
Long-term effect (at 6-12 months)	6	0.32 (0.09-0.55)	29 (0-71)	10.6
Combined treatment vs. CBT alone				
All studies	15	0.19 (–0.11 to 0.50)	68 (45-81)	22.4
Outliers removed	13	0.19 (–0.01 to 0.39)	18 (0-56)	22.8
Only low risk of bias	2	–0.24 (–12.73 to 12.25)	94 (82-98)	14.7
Publication bias correction	18	0.37 (0.03-0.72)	77 (63-85)	12.8
Long-term effect (at 6-12 months)	5	0.11 (–0.38 to 0.60)	25 (0-70)	34.8

NNT – number needed to treat. The reported publication bias correction is that using the trim and fill procedure.

Most trials examine CBT in an individual, group or guided self-help format, and we previously showed that there are no significant differences between these formats[12]. In the current meta-analysis, we could also include a set of trials of unguided self-help CBT, and found that this was also effective, with a small to moderate effect size (g=0.45). CBT was also found to be effective in inpatient settings (g=0.65), as well as in children and adolescents (g=0.41).

Research on CBT has evolved over time. The quality of studies has improved, which can be seen from the increasing number of trials with low risk of bias, the decrease in the use of waitlist control groups, and the increase in sample sizes of included studies. The number of treatment sessions has significantly decreased over the years. In a meta-regression analysis, we could not confirm that the effect size of CBT has decreased over time, as was suggested in an earlier study[37].

The findings of this study should be considered in the light of some limitations. First, heterogeneity was high in many analyses, and subgroup and meta-regression analyses could not identify all

Table 5 Other comparisons between cognitive behavior therapy (CBT) and control conditions

	n	g (95% CI)	I^2 (95% CI)	NNT
Unguided self-help CBT				
All comparisons	39	0.45 (0.31-0.60)	78 (71-84)	7.2
Outliers removed	34	0.43 (0.34-0.52)	51 (28-67)	7.7
Only low risk of bias	18	0.40 (0.27-0.52)	59 (32-76)	8.4
Publication bias correction	53	0.25 (0.07-0.43)	84 (80-88)	14.2
CBT in institutional settings				
All comparisons	11	0.65 (0.21-1.08)	70 (45-84)	4.8
Outliers removed	10	0.49 (0.15-0.83)	52 (2-77)	6.6
Publication bias correction	13	0.41 (–0.14 to 0.96)	81 (68-88)	8.2
CBT in children and adolescents				
All comparisons	39	0.41 (0.25-0.57)	78 (70-84)	8.1
Outliers removed	32	0.33 (0.23-0.43)	24 (0-51)	10.3
Only low risk of bias	8	0.17 (–0.10 to 0.45)	78 (57-89)	21
Publication bias correction	55	0.10 (–0.09 to 0.30)	86 (82-899	36.8

NNT – number needed to treat. The reported publication bias correction is that using the trim and fill procedure.

sources of this heterogeneity, suggesting that there are differences between trials that cannot be explained by the extracted characteristics. Second, risk of bias was high in many of the included trials, and the effect sizes of the trials with low risk of bias were significantly lower in some of the analyses. Fortunately, the number of studies was so large that we could examine outcomes in subsets of trials with low risk of bias. Finally, we found indications of publication bias in many analyses, although several findings remained robust after correcting for this bias.

We can conclude that CBT is effective in the treatment of depression with a moderate to large effect size, and that its effect is still significant up to 12 months. The superiority of CBT over other psychotherapies does not emerge clearly from this meta-analysis. CBT appears to be as effective as pharmacotherapies at the short term, but more effective at the longer term. Combined treatment appears to be superior to pharmacotherapy alone but not to CBT alone. The efficacy of CBT in depression is documented across different delivery formats, ages, target groups, and settings.

ACKNOWLEDGEMENTS

Supplementary information on this study is available at http://osf.io/a6p3w.

REFERENCES

1. World Health Organization. World mental health report: transforming mental health for all. Geneva: World Health Organization, 2022.
2. Herrman H, Patel V, Kieling C et al. Time for united action on depression: a Lancet-World Psychiatric Association Commission. Lancet 2022;399:957-1022.
3. Cuijpers P, Vogelzangs N, Twisk J et al. Comprehensive meta-analysis of excess mortality in depression in the general community versus patients with specific illnesses. Am J Psychiatry 2014;171:453-62.
4. Chisholm D, Sweeny K, Sheehan P et al. Scaling-up treatment of depression and anxiety: a global return on investment analysis. Lancet Psychiatry 2016; 3:415-24.
5. Cipriani A, Furukawa TA, Salanti G et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. Lancet 2018;391:1357-66.
6. Cuijpers P, Quero S, Noma H et al. Psychotherapies for depression: a network meta-analysis covering efficacy, acceptability and long-term outcomes of all main treatment types. World Psychiatry 2021;20:283-93.
7. Cuijpers P, Karyotaki E, de Wit L et al. The effects of fifteen evidence-supported therapies for adult depression: a meta-analytic review. Psychother Res 2020; 30:279-93.
8. Cuijpers P, Berking M, Andersson G et al. A meta-analysis of cognitive behavior therapy for adult depression, alone and in comparison to other treatments. Can J Psychiatry 2013;58:376-85.
9. Cuijpers P, Noma H, Karyotaki E et al. A network meta-analysis of the effects of psychotherapies, pharmacotherapies and their combination in the treatment of adult depression. World Psychiatry 2020;19:92-107.
10. Cuijpers P, Karyotaki E. A meta-analytic database of randomised trials on psychotherapies for depression. https://doi.org/10.17605/OSF.IO/825C6.
11. Cuijpers P, van Straten A, Andersson G et al. Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. J Consult Clin Psychol 2008;76:909-22.
12. Cuijpers P, Noma H, Karyotaki E et al. Effectiveness and acceptability of cognitive behavior therapy delivery formats in adults with depression: a network meta-analysis. JAMA Psychiatry 2019;76:700-7.
13. Cuijpers P, Ciharova M, Miguel C et al. Psychological treatment of depression in institutional settings: a meta-analytic review. J Affect Disord 2021;286:340-50.
14. Higgins JPT, Altman DG, Gøtzsche PC et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928.
15. Sterne JAC, Savovic J, Page MJ et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019;366:l4898.
16. Hedges LV, Olkin I. Statistical methods for meta-analysis. San Diego: Academic Press, 1985.
17. Harrer M, Kuper P, Cuijpers P. metapsyTools: several R helper functions for the "metapsy" database. R package version 0.3.2, 2022. https://tools.metapsy.org.
18. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with r: a practical tutorial. Evid Based Ment Health 2019;22:153-60.
19. Viechtbauer W. Conducting meta-analyses in r with the metafor package. J Statist Softw 2010;36:1-48.

20. Harrer M, Cuijpers P, Furukawa T et al. dmetar: companion R package for the guide 'doing meta-analysis in R'. R package version 0.0.9000. http://dmetar.protectlab.org.
21. Pustejovsky JE, Tipton E. Meta-analysis with robust variance estimation: expanding the range of working models. Prev Sci 2022;23:425-38.
22. Viechtbauer W, Cheung MWL. Outlier and influence diagnostics for meta-analysis. Res Synth Meth 2010;1:112-25.
23. Maier M, VanderWeele TJ, Mathur MB. Using selection models to assess sensitivity to publication bias: a tutorial and call for more routine use. Campbell Syst Rev 2022;18:e1256.
24. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 2000;56:455-63.
25. Rücker G, Schwarzer G, Carpenter JR et al. Treatment-effect estimates adjusted for small-study effects via a limit meta-analysis. Biostatistics 2011;12:122-42.
26. McShane BB, Böckenholt U, Hansen KT. Adjusting for publication bias in meta-analysis: an evaluation of selection methods and some cautionary notes. Perspect Psychol Sci 2016;11:730-49.
27. Carter EC, Schönbrodt FD, Gervais WM et al. Correcting for bias in psychology: a comparison of meta-analytic methods. Adv Meth Pract Psychol Sci 2019;2:115-44.
28. IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. BMC Med Res Methodol 2014;14:25.
29. Higgins JPT, Thompson SG, Deeks J et al. Measuring inconsistency in meta-analyses. BMJ 2003; 327:557-60.
30. Cheung MWL. Modeling dependent effect sizes with three-level meta-analyses: a structural equation modeling approach. Psychol Methods 2014;19:211-29.
31. Borenstein M, Hedges LV, Higgins JPT et al. Introduction to meta-analysis. Chichester: Wiley, 2009.
32. Borenstein M, Higgins JPT, Hedges LV et al. Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity. Res Syn Meth 2017;8:5-18.
33. Furukawa TA. From effect size into number needed to treat. Lancet 1999;353:1680.
34. Cuijpers P, Karyotaki E, Ciharova M et al. The effects of psychotherapies for depression on response, remission, reliable change, and deterioration: a meta-analysis. Acta Psychiatr Scand 2021;144:288-99.
35. Furukawa TA, Cipriani A, Barbui C et al. Imputing response rates from means and standard deviations in metaanalyses. Int Clin Psychopharmacol 2005;20:49-52.
36. Frank E, Prien RF, Jarrett RB et al. Conceptualization and rationale for consensus definitions of terms in major depressive disorder: remission, recovery, relapse, and recurrence. Arch Gen Psychiatry 1991;48:851-5.
37. Johnsen TJ, Friborg O. The effects of cognitive behavioral therapy as an antidepressive treatment is falling: a meta-analysis. Psychol Bull 2015;141:747-68.

DOI:10.1002/wps.21069

Metabolic side effects in persons with schizophrenia during mid- to long-term treatment with antipsychotics: a network meta-analysis of randomized controlled trials

Angelika Burschinski[1], Johannes Schneider-Thoma[1], Virginia Chiocchia[2,3], Kristina Schestag[1], Dongfang Wang[1], Spyridon Siafis[1], Irene Bighelli[1], Hui Wu[1], Wulf-Peter Hansen[4], Josef Priller[1,5-7], John M. Davis[8,9], Georgia Salanti[2], Stefan Leucht[1]

[1]Department of Psychiatry and Psychotherapy, School of Medicine, Technical University of Munich, Munich, Germany; [2]Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland; [3]Graduate School for Health Sciences, University of Bern, Bern, Switzerland; [4]BASTA - Bündnis für psychisch erkrankte Menschen, Munich, Germany; [5]University of Edinburgh and UK Dementia Research Institute, Edinburgh, UK; [6]Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK; [7]Neuropsychiatrie, Charité Universitätsmedizin Berlin and German Center for Neurodegenerative Diseases, Berlin, Germany; [8]Psychiatric Institute, University of Illinois at Chicago, Chicago, IL, USA; [9]Maryland Psychiatric Research Center, Baltimore, MD, USA

Metabolic side effects of antipsychotic drugs can have serious health consequences and may increase mortality. Although persons with schizophrenia often take these drugs for a long time, their mid- to long-term metabolic effects have been studied little so far. This study aimed to evaluate the mid- to long-term metabolic side effects of 31 antipsychotics in persons with schizophrenia by applying a random-effects Bayesian network meta-analysis. We searched the Cochrane Schizophrenia Group's Study-Based Register of Trials (up to April 27, 2020) and PubMed (up to June 14, 2021). We included published and unpublished, open and blinded randomized controlled trials with a study duration >13 weeks which compared any antipsychotic in any form of administration with another antipsychotic or with placebo in participants diagnosed with schizophrenia. The primary outcome was weight gain measured in kilograms. Secondary outcomes included "number of participants with weight gain", fasting glucose, total cholesterol, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, and triglycerides. We identified 137 eligible trials (with 35,007 participants) on 31 antipsychotics, with a median follow-up of 45 weeks. Chlorpromazine produced the most weight gain (mean difference to placebo: 5.13 kg, 95% credible interval, CrI: 1.98 to 8.30), followed by clozapine (4.21 kg, 95% CrI: 3.03 to 5.42), olanzapine (3.82 kg, 95% CrI: 3.15 to 4.50), and zotepine (3.87 kg, 95% CrI: 2.14 to 5.58). The findings did not substantially change in sensitivity and network meta-regression analyses, although enriched design, drug company sponsorship, and the use of observed case instead of intention-to-treat data modified the mean difference in weight gain to some extent. Antipsychotics with more weight gain were often also among the drugs with worse outcome in fasting glucose and lipid parameters. The confidence in the evidence ranged from low to moderate. In conclusion, antipsychotic drugs differ in their propensity to induce metabolic side effects in mid- to long-term treatment. Given that schizophrenia is often a chronic disorder, these findings should be given more consideration than short-term data in drug choice.

Key words: Antipsychotic drugs, metabolic side effects, weight gain, glucose, cholesterol, triglycerides, chlorpromazine, clozapine, olanzapine, zotepine, schizophrenia

(World Psychiatry 2023;22:116–128)

Antipsychotic drugs are the core treatment for schizophrenia, because they are efficacious in acute episodes[1] and in preventing relapses[2]. Consequently, many persons with schizophrenia take antipsychotics for years, or even lifetime[3,4]. However, antipsychotics also have considerable side effects[1]. Metabolic side effects can manifest as weight gain, changes in cholesterol and triglyceride metabolism (dyslipidaemia), and dysregulation of glucose homeostasis (insulin resistance extending to diabetes)[5]. They are associated with cardiovascular diseases, including myocardial infarction and stroke[6-8]. Therefore, metabolic side effects of antipsychotics are likely to contribute to the average 14.5 years reduced life-span of individuals with schizophrenia[9]. Furthermore, weight gain is associated with decreased quality of life[10] and treatment non-adherence[11,12], the latter resulting in poor treatment outcome and psychotic relapses.

As antipsychotic drugs do not differ much in efficacy[13], guidelines recommend that the choice of the drug should be primarily informed by their side effects[14,15]. Recently, a network meta-analysis compared the metabolic effects of 18 antipsychotics during acute treatment of schizophrenia in studies with a median treatment duration of 6 weeks[16]. However, antipsychotics are also used for prevention of relapses, and individuals take them for much longer periods of time. Therefore, the aim of the current network meta-analysis was to investigate the mid- to long-term metabolic effects of these drugs in randomized controlled trials (RCTs). Such knowledge should be highly relevant for clinical practice and contribute to tailored drug choice.

METHODS

Inclusion criteria and search strategy

We report following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) extension statement for network meta-analyses. The study protocol was registered with PROSPERO (registration number: CRD42020175414) and published[17].

We included mid-term and long-term randomized controlled antipsychotic drug trials (>3 and >6 months, respectively), following the classification of the Cochrane Schizophrenia Group[18]. Trials were included irrespective of their blinding and study setting. However, trials conducted in mainland China were excluded due to raised quality concerns[19-21], and trials with a randomization process at high risk of bias were also excluded. Moreover, continuation studies in which only responders of the core trial could participate were excluded, because this corrupts

randomization.

Studies were included if at least 80% of the trial participants had a diagnosis of schizophrenia, schizophreniform or schizoaffective disorder, irrespectively of the diagnostic criteria. There were no restrictions concerning participants' stage of the disease, age, gender or ethnicity.

We included all second-generation antipsychotics (SGAs) available in Europe or the US, and a selection of first-generation antipsychotics (FGAs) informed by a survey of international schizophrenia experts[22], administered as monotherapy – namely, amisulpride, aripiprazole, asenapine, benperidol, brexpiprazole, cariprazine, chlorpromazine, clopenthixol, clozapine, flupentixol, fluphenazine, fluspirilene, haloperidol, iloperidone, levomepromazine, loxapine, lumateperone, lurasidone, molindone, olanzapine, paliperidone, penfluridol, perazine, perphenazine, pimozide, quetiapine, risperidone, sertindole, sulpiride, thioridazine, tiotixene, trifluoperazine, ziprasidone, zotepine, and zuclopenthixol.

Oral and long-acting injectable (LAI) formulations of one compound were considered as different interventions, because their side effect profile could differ due to pharmacokinetic or adherence issues[23,24], but were combined in a *post-hoc* sensitivity analysis. We included all study arms with doses within the target to maximum range according to the International Consensus Study on Antipsychotic Dosing[25]. Only for specific populations such as individuals with first episode or primarily negative symptoms, for which clinically different dosing regimens are recommended, we included lower doses.

We searched the Cochrane Schizophrenia Group's Study-Based Register of Trials (compiled by monthly searches in multiple electronic databases and trial registries up to April 27, 2020), PubMed (last update on June 14, 2021) and related systematic reviews[23,26-33] (see also supplementary information). Two reviewers (AB, DW) independently screened the searches; in case of disagreement, a third reviewer (JS-T or SL) was involved.

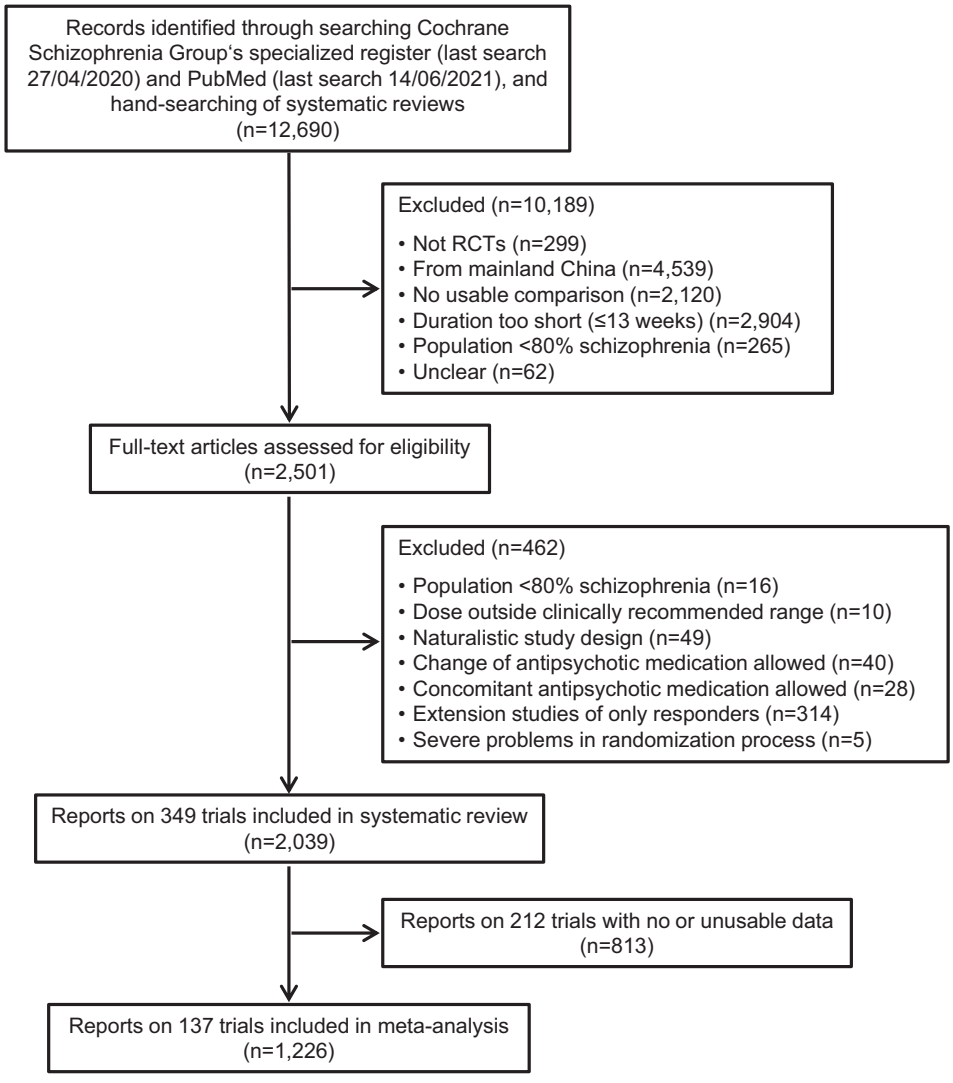

Figure 1 Flow chart of the study selection process. RCT – randomized controlled trial

Outcomes, data extraction and evaluation of study risk of bias

The primary outcome was weight gain in kilograms (kg). Secondary outcomes were the "number of participants with weight gain" (≥7% from baseline preferred to other definitions), and continuous measurements of fasting glucose, total cholesterol, low density lipoprotein (LDL) cholesterol, high density lipoprotein (HDL) cholesterol, and triglycerides. All outcomes were extracted at study endpoint.

Following our protocol, we also extracted data for infrequently reported outcomes – such as body mass index, waist circumference, hemoglobin A1c (HbA1c), homeostasis model assessment of insulin resistance (HOMA-IR), and insulin – but did not consider them for further analysis due to scarcity of data. Additionally, study characteristics (study duration, blinding, criteria used to diagnose schizophrenia), population characteristics (baseline weight, age, gender, ethnicity, and lifetime exposure to antipsychotics – if not available, duration of illness was used as a proxy), and treatment characteristics (drug company sponsorship, antipsychotic dose) were extracted.

Two reviewers (AB, KS) extracted data for each included study in specifically customized digital forms in a Microsoft Access database and evaluated risk of bias using Cochrane's Risk of Bias 2 tool[34]. Conflicting entries were automatically detected and discussed, if needed, with a third reviewer (JS-T or SL) or the original authors. Original authors and drug companies responsible for included studies published during the past 20 years were also contacted via e-mail by AB and SL for missing information.

Data synthesis and evaluation of confidence in the evidence

Pairwise meta-analyses were performed in a frequentist setting, while network meta-analyses were performed in a Bayesian setting, both using the random effects model. We synthesized continuous outcomes with mean differences (MDs) and dichotomous outcomes using odds ratios (ORs), both presented with 95% credible intervals (CrIs).

For each outcome, we assumed a common heterogeneity variance (τ^2) across comparisons. The magnitude of heterogeneity was judged by comparing τ^2 to its empirical distribution[35,36] and by considering the width of the prediction intervals. Statistical inconsistency was evaluated using the SIDE-test for each comparison[37] and the design-by-treatment interaction test for the overall network[38].

To assess the plausibility of the transitivity assumption, we compared the distribution of key study characteristics across studies

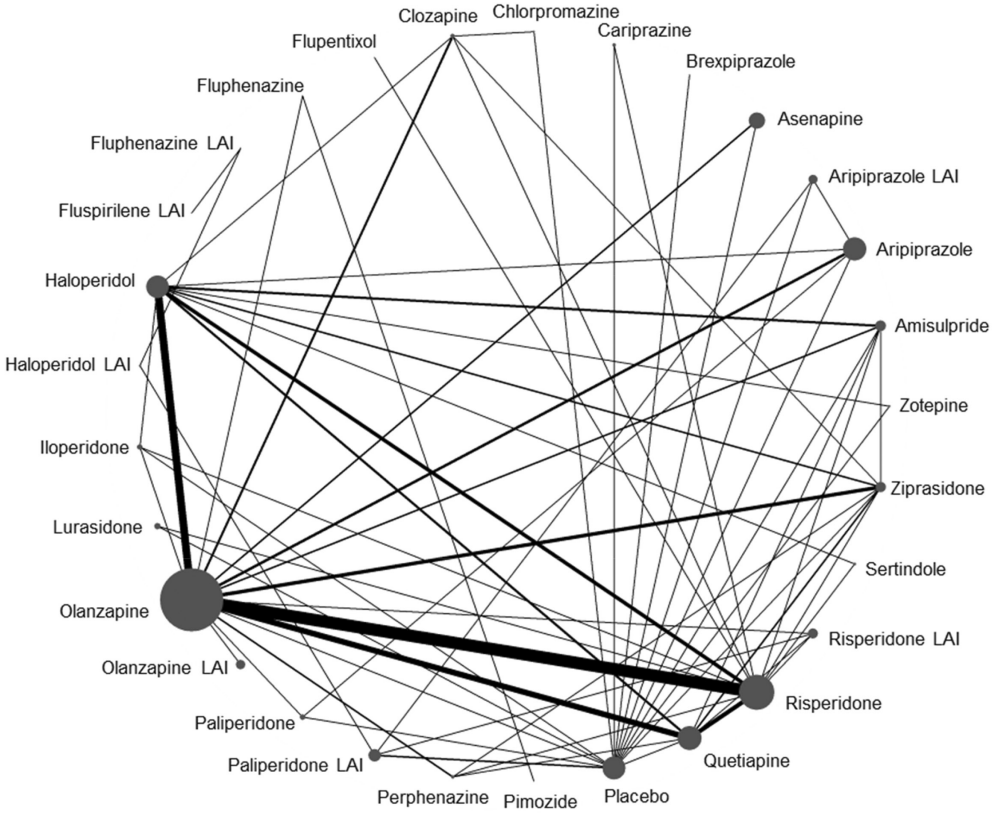

Figure 2 Network plot of primary outcome "weight gain". The lines link treatments that were directly compared in trials. The thickness of the lines corresponds to the number of trials evaluating the comparison. The size of the nodes corresponds to the number of participants assigned to the treatment. LAI – long-acting injectable.

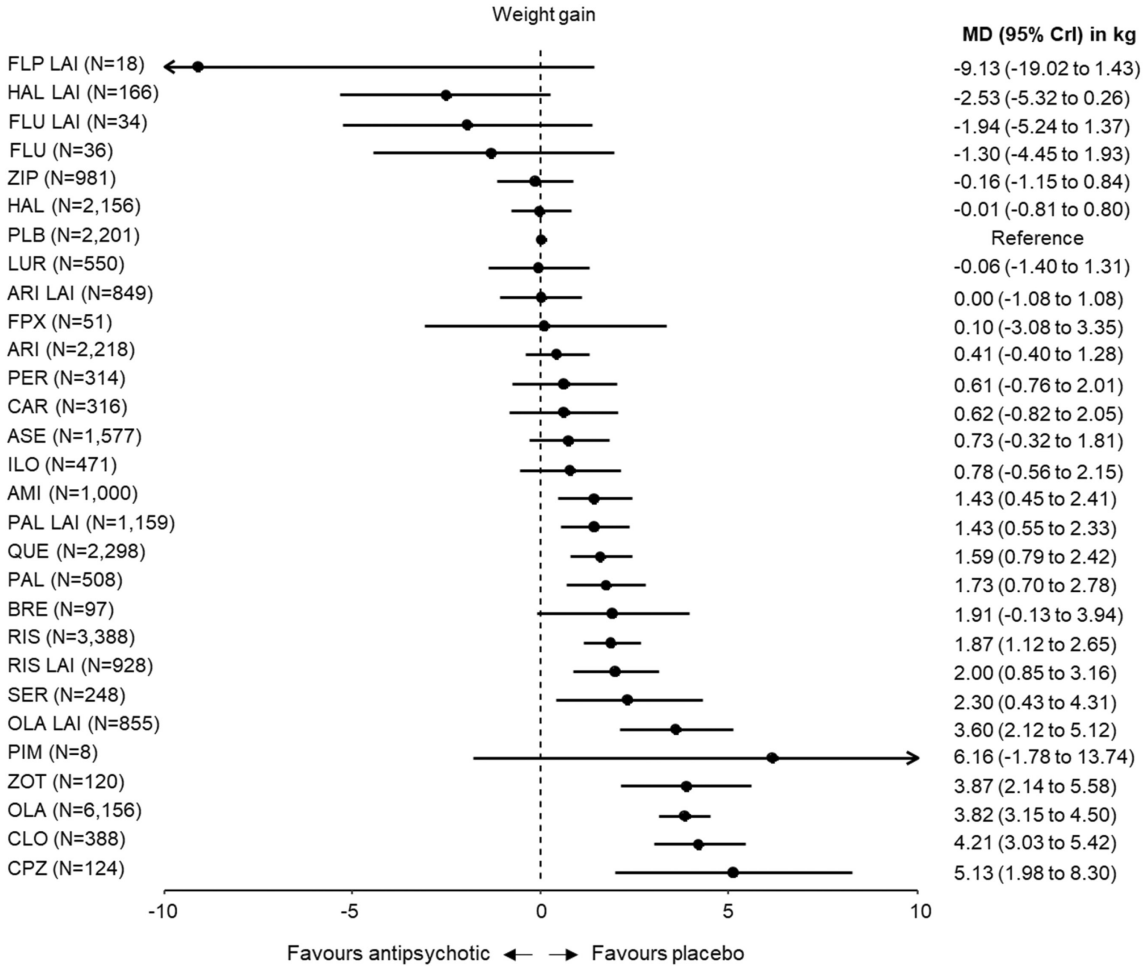

Figure 3 Forest plot of antipsychotic drugs vs. placebo for the primary outcome "weight gain". Network meta-analysis estimates of treatment effect of each drug vs. placebo are reported as mean differences (MDs) and 95% credible intervals (CrIs). The order of treatments is according to surface under the cumulative ranking curve (SUCRA) ranking. LAI – long-acting injectable, AMI – amisulpride, ARI – aripiprazole, ASE – asenapine, BRE – brexpiprazole, CAR – cariprazine, CLO – clozapine, CPZ – chlorpromazine, FLP – fluspirilene, FLU – fluphenazine, FPX – flupentixol, HAL – haloperidol, ILO – iloperidone, LUR – lurasidone, OLA – olanzapine, PAL – paliperidone, PER – perphenazine, PIM – pimozide, PLB – placebo, QUE – quetiapine, RIS – risperidone, SER – sertindole, ZIP – ziprasidone, ZOT – zotepine.

grouped by comparison. To explore sources of heterogeneity and inconsistency, we also planned network meta-regressions for baseline weight, age, gender, ethnicity, lifetime exposure to antipsychotics, drug company sponsorship, and study duration.

We performed sensitivity analyses by analyzing only observed cases, and by excluding non-double blind studies, studies with an overall assessment of high risk of bias, studies with enriched design, studies not using operationalized criteria to diagnose schizophrenia, and studies in which participants had minimal prior exposure to antipsychotics (e.g., children and first episode). We also performed a *post-hoc* analysis excluding doses at the lower and upper ends of the range recommended by the International Consensus Study on Antipsychotic Dosing[25].

To investigate the presence of small-study effects (potentially associated with publication bias), we performed – for the primary outcome – a comparison-adjusted funnel plot[39] and a contour-enhanced funnel plot of all drugs versus placebo[40].

All analyses were performed in R. We conducted Bayesian network meta-analyses using the BUGSnet package[41], and network meta-regression analyses using self-programmed routines with the rjags package[42]. Frequentist network and pairwise meta-analyses were performed with the netmeta and meta packages[43,44]. The confidence in the network meta-analysis estimates was evaluated for the primary outcome with the Confidence in Network Meta-Analysis (CINeMA) framework[45].

RESULTS

Description of included studies

We identified 12,690 references. After title/abstract screening, we assessed 2,501 full-text articles and included 2,039 reports on 349 trials (see Figure 1).

Table 1 Map of antipsychotics ranked according to associated alteration in weight gain and metabolic parameters

	Weight gain	Fasting glucose	Total cholesterol	LDL cholesterol	HDL cholesterol	Triglycerides
Fluspirilene LAI	–9.13 (–19.02 to 1.43)					
Haloperidol LAI	–2.53 (–5.32 to 0.26)	0.84 (–11.92 to 13.36)	7.68 (–3.09 to 18.90)	4.00 (–4.67 to 13.15)	0.49 (–2.08 to 3.04)	9.22 (–19.68 to 38.70)
Fluphenazine LAI	–1.94 (–5.24 to 1.37)					
Fluphenazine	–1.30 (–4.45 to 1.93)		15.45 (–19.43 to 51.29)			
Ziprasidone	–0.16 (–1.15 to 0.84)	–0.67 (–5.40 to 4.24)	–4.69 (–10.39 to 1.23)	–1.32 (–7.32 to 4.38)	–0.14 (–1.74 to 1.38)	–11.85 (–28.44 to 4.95)
Haloperidol	–0.01 (–0.81 to 0.80)	2.72 (–2.32 to 7.96)	2.77 (–3.21 to 8.69)	1.49 (–14.95 to 20.28)	–1.58 (–9.94 to 5.91)	6.90 (–13.58 to 27.06)
Placebo	0	0	0	0	0	0
Lurasidone	–0.06 (–1.40 to 1.31)	0.96 (–5.33 to 7.43)	3.88 (–3.11 to 11.07)	5.08 (–0.94 to 10.65)	0.70 (–0.97 to 2.54)	–13.09 (–33.06 to 7.51)
Aripiprazole LAI	–0.00 (–1.08 to 1.08)	2.35 (–1.51 to 6.53)	2.51 (–3.35 to 8.05)	0.60 (–4.06 to 5.49)	0.32 (–1.26 to 1.81)	–0.14 (–13.47 to 14.37)
Flupentixol	0.10 (–3.08 to 3.35)					
Aripiprazole	0.41 (–0.40 to 1.28)	0.35 (–2.40 to 3.28)	–0.75 (–4.90 to 3.21)	–1.92 (–5.64 to 1.96)	0.71 (–0.76 to 1.98)	–1.07 (–12.26 to 9.87)
Perphenazine	0.61 (–0.76 to 2.01)		4.46 (–3.72 to 12.73)		–0.18 (–1.98 to 1.66)	8.79 (–20.83 to 39.49)
Cariprazine	0.62 (–0.82 to 2.05)	1.76 (–2.82 to 6.42)	–0.55 (–8.18 to 7.52)	0.73 (–5.51 to 6.97)	–1.22 (–3.25 to 0.73)	–1.08 (–20.58 to 18.71)
Asenapine	0.73 (–0.32 to 1.81)	3.37 (–0.80 to 7.36)	4.86 (–1.25 to 11.32)	3.25 (–2.93 to 9.67)	–0.12 (–1.90 to 1.77)	4.22 (–14.87 to 22.67)
Iloperidone	0.78 (–0.56 to 2.15)	–0.24 (–4.40 to 4.57)	–0.59 (–9.37 to 8.02)	2.36 (–3.70 to 7.67)	–0.33 (–1.80 to 1.27)	14.66 (–3.01 to 29.36)
Amisulpride	1.43 (0.45 to 2.41)	2.13 (–2.72 to 7.04)	9.77 (–6.96 to 26.68)	9.72 (–6.90 to 26.88)	–5.24 (–8.94 to –2.05)	38.98 (12.66 to 66.49)
Paliperidone LAI	1.43 (0.55 to 2.33)	0.83 (–2.49 to 4.00)	3.31 (–1.18 to 8.13)	2.29 (–1.62 to 6.35)	–0.30 (–1.48 to 0.93)	–0.09 (–12.14 to 11.33)
Quetiapine	1.59 (0.79 to 2.42)	3.14 (0.09 to 6.33)	8.20 (3.33 to 13.30)	5.87 (1.33 to 10.51)	–1.59 (–2.91 to –0.27)	21.87 (7.79 to 35.81)
Paliperidone	1.73 (0.70 to 2.78)	1.85 (–1.89 to 5.64)	7.58 (2.21 to 13.17)	3.35 (–1.44 to 8.56)	0.15 (–1.42 to 1.75)	4.61 (–8.80 to 18.29)
Brexpiprazole	1.91 (–0.13 to 3.94)	3.62 (–4.37 to 11.71)	–0.28 (–14.06 to 13.51)	2.18 (–9.70 to 14.08)	–1.31 (–4.31 to 1.70)	2.18 (–24.34 to 28.63)
Risperidone	1.87 (1.12 to 2.65)	3.51 (0.21 to 6.80)	3.62 (–0.93 to 8.28)	4.02 (–0.91 to 9.04)	–1.20 (–2.45 to 0.15)	2.88 (–10.54 to 16.07)
Risperidone LAI	2.00 (0.85 to 3.16)	3.34 (–0.38 to 7.21)	7.58 (2.33 to 12.90)	5.84 (0.49 to 11.38)	–0.17 (–1.61 to 1.40)	8.40 (–6.63 to 23.83)
Sertindole	2.30 (0.43 to 4.31)	6.44 (–0.21 to 13.06)	9.07 (–6.01 to 24.54)	6.91 (–5.68 to 19.49)	0.24 (–3.42 to 4.61)	8.79 (–18.02 to 35.51)
Olanzapine LAI	3.60 (2.12 to 5.12)	7.64 (3.17 to 13.20)	12.02 (5.07 to 19.01)	9.59 (3.61 to 15.49)	–2.91 (–4.45 to –1.18)	20.46 (–0.40 to 41.68)
Pimozide	6.16 (–1.78 to 13.74)					
Zotepine	3.87 (2.14 to 5.58)					
Olanzapine	3.82 (3.15 to 4.50)	5.07 (2.44 to 7.98)	12.65 (8.73 to 16.51)	8.09 (4.32 to 11.89)	–2.59 (–3.71 to –1.44)	31.66 (20.32 to 42.84)
Clozapine	4.21 (3.03 to 5.42)	1.64 (–7.08 to 10.26)	15.83 (–2.44 to 32.73)			
Chlorpromazine	5.13 (1.98 to 8.30)	4.94 (–7.93 to 18.90)	13.00 (–2.21 to 29.08)			

Numbers present the mean differences (MDs) with their 95% credible intervals (CrIs) from the network meta-analysis compared to placebo. The order of treatments is according to surface under the cumulative ranking curve (SUCRA) value of the primary outcome "weight gain". The color gradient from grey to white represents the SUCRA value, with darker fields indicating a higher probability of being the worst drug. Empty cells indicate that no data are available. LAI – long-acting injectable, LDL – low density lipoprotein, HDL – high density lipoprotein.

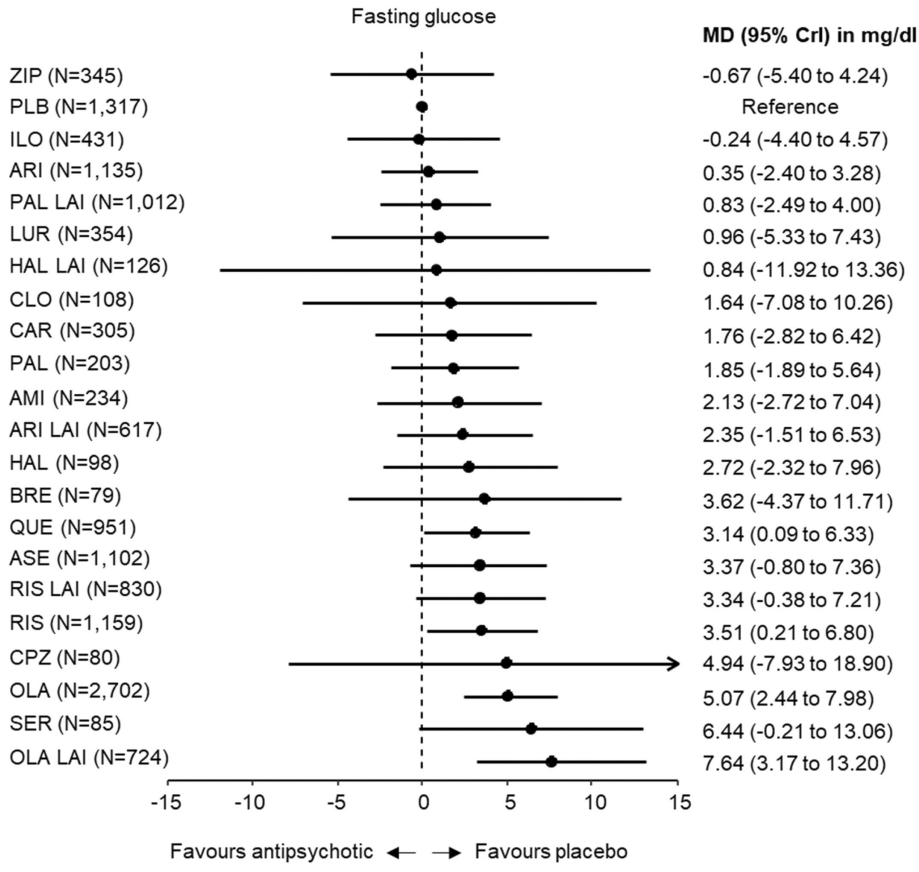

Figure 4 Forest plot of antipsychotic drugs vs. placebo for the secondary outcome "fasting glucose". Network meta-analysis estimates of treatment effect of each drug vs. placebo are reported as mean differences (MDs) and 95% credible intervals (CrIs). The order of treatments is according to surface under the cumulative ranking curve (SUCRA) ranking. LAI – long-acting injectable, AMI – amisulpride, ARI – aripiprazole, ASE – asenapine, BRE – brexpiprazole, CAR – cariprazine, CLO – clozapine, CPZ – chlorpromazine, HAL – haloperidol, ILO – iloperidone, LUR – lurasidone, OLA – olanzapine, PAL – paliperidone, PLB – placebo, QUE – quetiapine, RIS – risperidone, SER – sertindole, ZIP – ziprasidone.

One hundred thirty-seven trials with 35,007 participants and 31 different antipsychotics provided usable data. The median average age of participants was 38.9 years (interquartile range, IQR: 35.3-41.4); the median trial duration was 45 weeks (IQR: 26-52); the median percentage of women was 37% (IQR: 29-43); and 70% (96 of 137) of the trials were double-blind. We found no clear evidence of differences in study characteristics across comparisons. Although the number of studies per comparison was small, we judged that there were no clear violations of the transitivity assumption (see supplementary information).

Primary outcome: weight gain

One hundred ten trials on 28 antipsychotics (N=29,215 participants with an average baseline weight of 76.55 kg) contributed to the network meta-analysis for the primary outcome (weight gain). The network plot is provided in Figure 2.

The network estimates and corresponding 95% CrI for each drug versus placebo are reported in Figure 3. Medication administration is oral if not otherwise stated. Most drugs were associated with more weight gain than placebo. The following drugs produced on average more than 2 kg weight gain in excess to placebo: chlorpromazine (MD: 5.13), clozapine (MD: 4.21), olanzapine oral/LAI (MD: 3.82/3.60), zotepine (MD: 3.87), pimozide (MD: 6.16), and sertindole (MD: 2.30). The following drugs produced on average between 1 and 2 kg weight gain in excess to placebo: risperidone LAI/oral (MD: 2.00/1.87), brexpiprazole (MD: 1.91), paliperidone oral/LAI (MD: 1.73/1.43), quetiapine (MD: 1.59), and amisulpride (MD: 1.43). The following drugs produced on average less than 1 kg weight gain in excess to placebo: iloperidone (MD: 0.78), asenapine (MD: 0.73), cariprazine (MD: 0.62), perphenazine (MD: 0.61), and aripiprazole (MD: 0.41). The following drugs were similar to placebo: flupentixol (MD: 0.10), aripiprazole LAI (MD: 0.00), lurasidone (MD: –0.06), haloperidol (MD: –0.01), and ziprasidone (MD: –0.16). Three drugs produced on average a weight loss compared to placebo: fluspirilene LAI (MD: –9.13), haloperidol LAI (MD: –2.53), and fluphenazine LAI/oral (MD: –1.94/–1.30). However, their 95% CrIs were wide (i.e., the estimates were imprecise) and even include the possibility of small weight gain. MDs between drugs and results of relevant pairwise meta-analyses are provided in the supplementary infor-

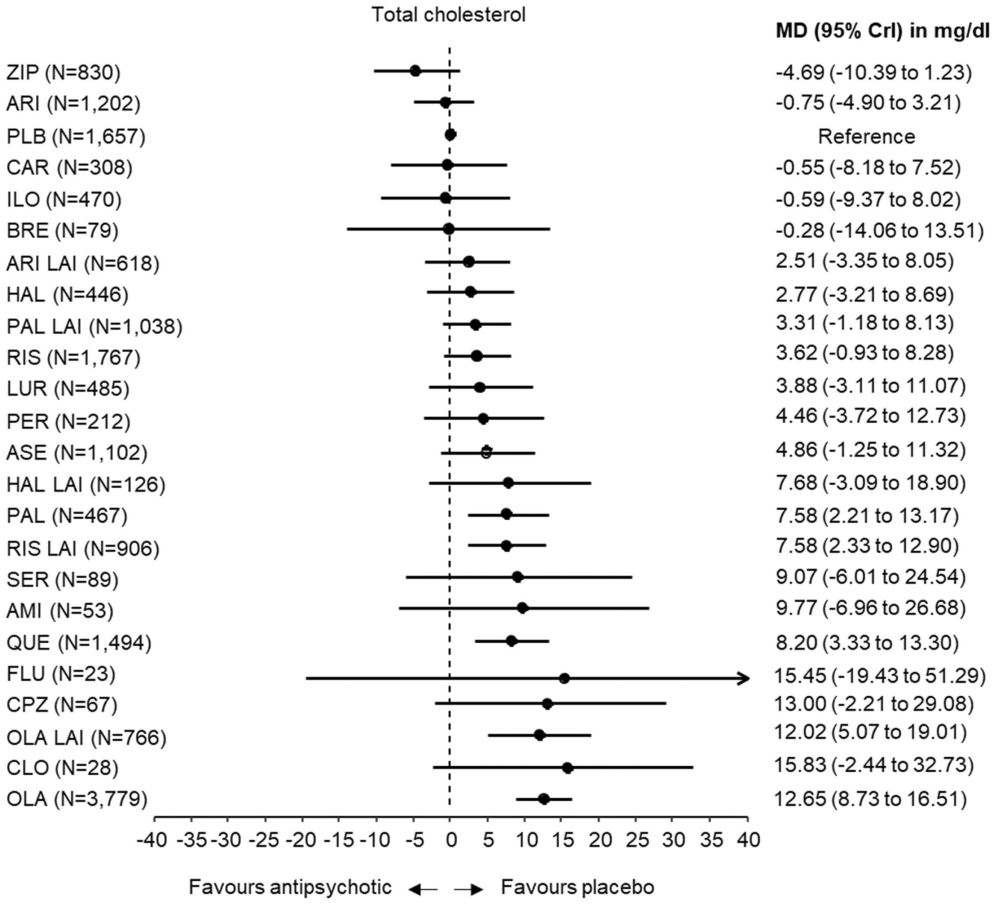

Figure 5 Forest plot of antipsychotic drugs vs. placebo for the secondary outcome "total cholesterol". Network meta-analysis estimates of treatment effect of each drug vs. placebo are reported as mean differences (MDs) and 95% credible intervals (CrIs). The order of treatments is according to surface under the cumulative ranking curve (SUCRA) ranking. LAI – long-acting injectable, AMI – amisulpride, ARI – aripiprazole, ASE – asenapine, BRE – brexpiprazole, CAR – cariprazine, CLO – clozapine, CPZ – chlorpromazine, FLU – fluphenazine, HAL – haloperidol, ILO – iloperidone, LUR – lurasidone, OLA – olanzapine, PAL – paliperidone, PER – perphenazine, PLB – placebo, QUE – quetiapine, RIS – risperidone, SER – sertindole, ZIP – ziprasidone.

mation.

No evidence of inconsistency was found (see supplementary information). The heterogeneity standard deviation common-τ was 0.82 for MD and 0.15 on the standardized mean difference (SMD) scale, which can be interpreted as low to moderate when considering empirical distributions and prediction intervals (see also supplementary information).

In network meta-regressions, we found that the MD of any antipsychotic versus placebo was on average 0.45 kg (95% CrI: 0.01 to 0.89) higher in sponsored than in non-sponsored study arms. Adjusting for drug company sponsoring reduced common-τ from 0.82 to 0.65. Other possible effect modifiers showed no clear effect (see also supplementary information).

In sensitivity analyses, when studies with enriched design were excluded, all antipsychotics showed larger MDs (on average 0.63 kg) compared to the main analysis; and observed cases (available for 21 antipsychotics) yielded more pronounced differences in MDs versus placebo, ranging from –10.63 to 6.42 kg (see also supplementary information).

Despite these observed effects on treatment results, the rankings remained similar in all network meta-regressions and sensitivity analyses.

We found no clear indication of small-study effects and publication bias. The overall risk of bias was "some concerns" for 72% (79 of 110) and "high" for 28% (31 of 110) of studies. The confidence in the network meta-analysis estimates was low in 276, moderate in 123 and very low in 7 comparisons (see also supplementary information).

Secondary metabolic outcomes

The results for "number of participants with weight gain" were very similar to the primary outcome weight gain (see supplementary information). For lipid and glucose outcomes, less data were available for most antipsychotics, with no data for zotepine and the older antipsychotics except haloperidol and perphenazine.

Drugs associated with weight gain were often also associated with worse outcomes in fasting glucose and lipid parameters (see Table 1). The ranges of the MDs in mg/dl compared to placebo

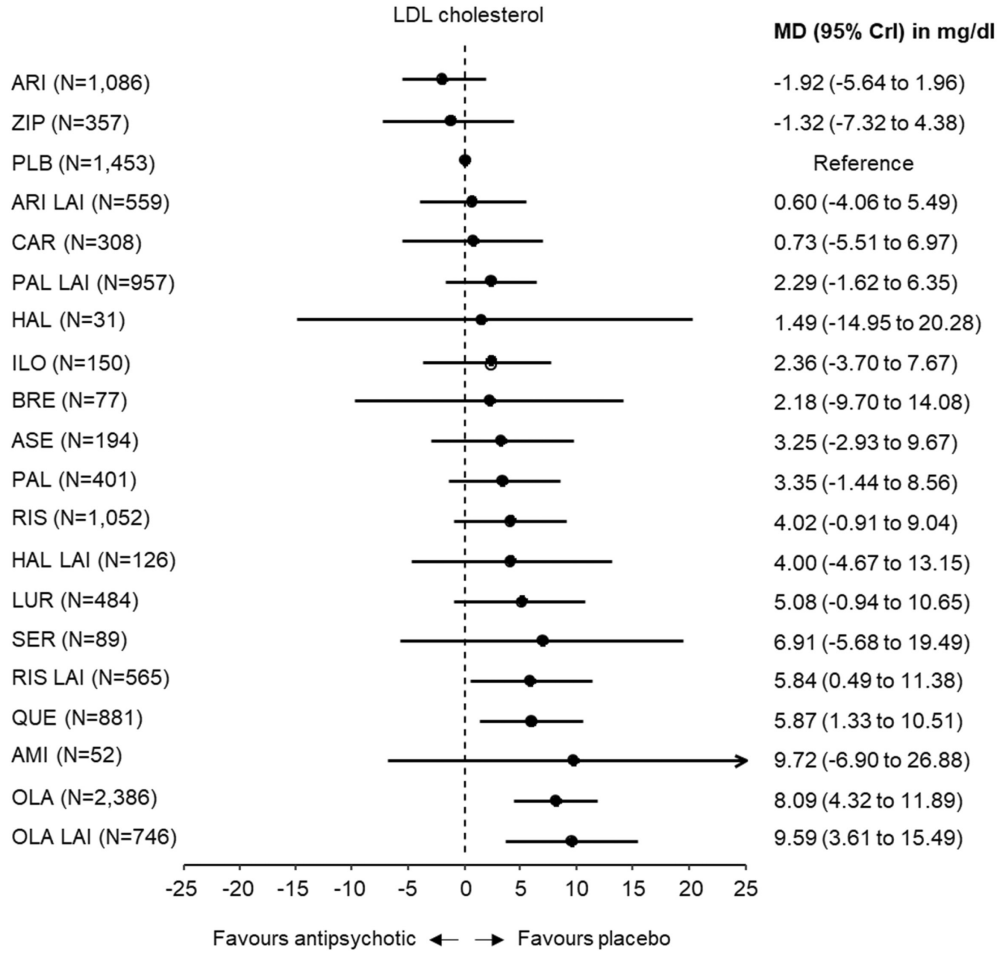

Figure 6 Forest plot of antipsychotic drugs vs. placebo for the secondary outcome "low-density lipoprotein (LDL) cholesterol". Network meta-analysis estimates of treatment effect of each drug vs. placebo are reported as mean differences (MDs) and 95% credible intervals (CrIs). The order of treatments is according to surface under the cumulative ranking curve (SUCRA) ranking. LAI – long-acting injectable, AMI – amisulpride, ARI – aripiprazole, ASE – asenapine, BRE – brexpiprazole, CAR – cariprazine, HAL – haloperidol, ILO – iloperidone, LUR – lurasidone, OLA – olanzapine, PAL – paliperidone, PLB – placebo, QUE – quetiapine, RIS – risperidone, SER – sertindole, ZIP – ziprasidone.

were as follows: from 7.64 (95% CrI: 3.17 to 13.20) for olanzapine LAI to –0.67 (95% CrI: –5.40 to 4.24) for ziprasidone concerning fasting glucose (see also Figure 4); from 12.65 (95% CrI: 8.73 to 16.51) for olanzapine to –4.69 (95% CrI: –10.39 to 1.23) for ziprasidone concerning total cholesterol (see also Figure 5); from 9.59 (95% CrI: 3.61 to 15.49) for olanzapine LAI to –1.92 (95% CrI: –5.64 to 1.96) for aripiprazole concerning LDL cholesterol (see also Figure 6); from –5.24 (95% CrI: –8.94 to –2.05) for amisulpride to 0.71 (95% CrI: –0.76 to 1.98) for aripiprazole concerning HDL cholesterol (see also Figure 7); and from 38.98 (95% CrI: 12.66 to 66.49) for amisulpride to –11.85 (95% CrI: –28.44 to 4.95) for ziprasidone concerning triglycerides (see also Figure 8).

No evidence of inconsistency was detected for total cholesterol, LDL and HDL cholesterol; little evidence of inconsistency was present for "number of participants with weight gain", fasting glucose and triglycerides. Heterogeneity for the secondary outcomes ranged between low and low to moderate (see supplementary information).

DISCUSSION

We, for the first time, synthesized the mid- to long-term (median: 45 weeks) evidence on metabolic side effects of 31 antipsychotics in people with schizophrenia, using a network meta-analysis based on 137 RCTs including 35,007 participants. As antipsychotic drugs are often taken for long periods of time, our results represent more valuable clinical information on these health consequences than previous analyses based on short-term studies which on average only lasted 6 weeks[1,16].

Every kilogram increase in body weight (our primary outcome) increases the risk of cardiovascular disease by 3.1%[16,46]. We found that antipsychotics differ in their propensity to cause weight gain (see Figure 3). For some antipsychotics, the average weight gain was comparable with placebo, in the sense that there was a tendency to either weight loss (fluspirilene LAI, haloperidol LAI and oral, fluphenazine LAI and oral, and ziprasidone) or an average weight gain of up to 1 kg (lurasidone, aripiprazole LAI and oral,

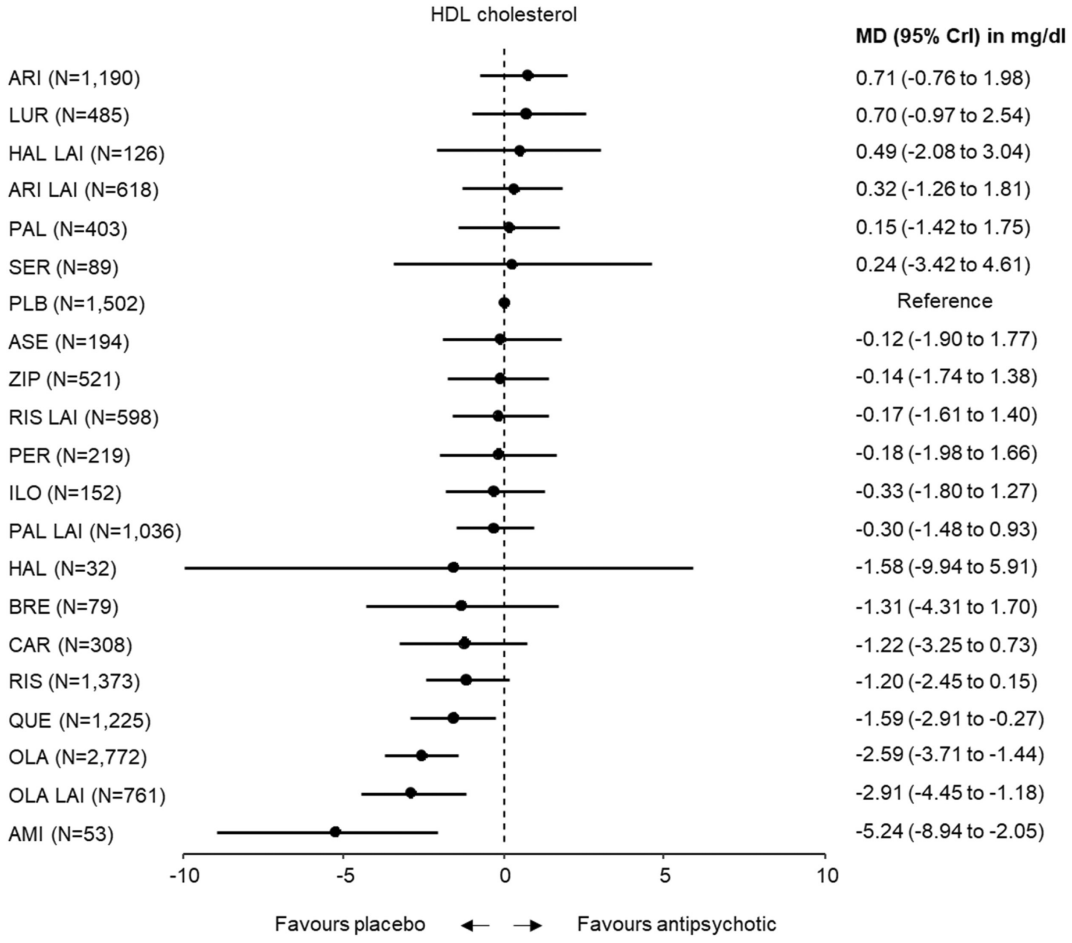

Figure 7 Forest plot of antipsychotic drugs vs. placebo for the secondary outcome "high-density lipoprotein (HDL) cholesterol". Network meta-analysis estimates of treatment effect of each drug vs. placebo are reported as mean differences (MDs) and 95% credible intervals (CrIs). The order of treatments is according to surface under the cumulative ranking curve (SUCRA) ranking. LAI – long-acting injectable, AMI – amisulpride, ARI – aripiprazole, ASE – asenapine, BRE – brexpiprazole, CAR – cariprazine, HAL – haloperidol, ILO – iloperidone, LUR – lurasidone, OLA – olanzapine, PAL – paliperidone, PER – perphenazine, PLB – placebo, QUE – quetiapine, RIS – risperidone, SER – sertindole, ZIP – ziprasidone.

flupentixol, perphenazine, cariprazine, asenapine and iloperidone). All CrIs of these drugs included zero, indicating that some weight loss or weight gain is possible.

An average weight gain between 1 and 2 kg compared to placebo was observed for amisulpride, paliperidone LAI and oral, quetiapine, brexpiprazole, and risperidone oral and LAI. An average weight gain higher than 2 kg compared to placebo was estimated for sertindole, olanzapine LAI and oral, pimozide, zotepine, clozapine and chlorpromazine. These drugs with substantial weight gain were also associated with more glucose and lipid disturbances, with olanzapine showing the most pronounced alterations. Of note, for some drugs, the estimates are very uncertain due to small sample sizes, particularly for fluspirilene LAI and pimozide.

In network meta-regressions, we found no moderating effect for baseline weight, gender, age and ethnicity. Sponsored study arms showed more weight gain compared to non-sponsored ones, which gives no indication for bias by drug company sponsorship because the effect is not in favour of sponsored drugs.

We found no substantial difference between oral and LAI formulations, and the hierarchy in the sensitivity analysis pooling oral and LAI formulations was similar to the main analysis (see Figure 9). Haloperidol is an exception, since weight loss was observed with its LAI formulation, while the oral formulation was weight neutral. However, haloperidol LAI is only connected to the main network by one study with extreme results[47], meaning that for this drug, as well as for fluphenazine LAI and fluspirilene LAI, control by indirect evidence is lacking.

The ranking of antipsychotics in all outcomes was comparable with short-term findings[16] (median treatment duration: 6 weeks vs. 45 weeks here). For fasting glucose and lipid parameters, the magnitude of the effect was also similar. This suggests that the effects on these parameters occur rapidly, and then remain stable.

Weight gain was more pronounced in our mid- to long-term data compared to the reported short-term data[16], but not as much as expected, with the highest difference (approximately +1 kg) seen for olanzapine. However, this result is in line with those of other studies: in a pairwise meta-analysis[26], a significant addi-

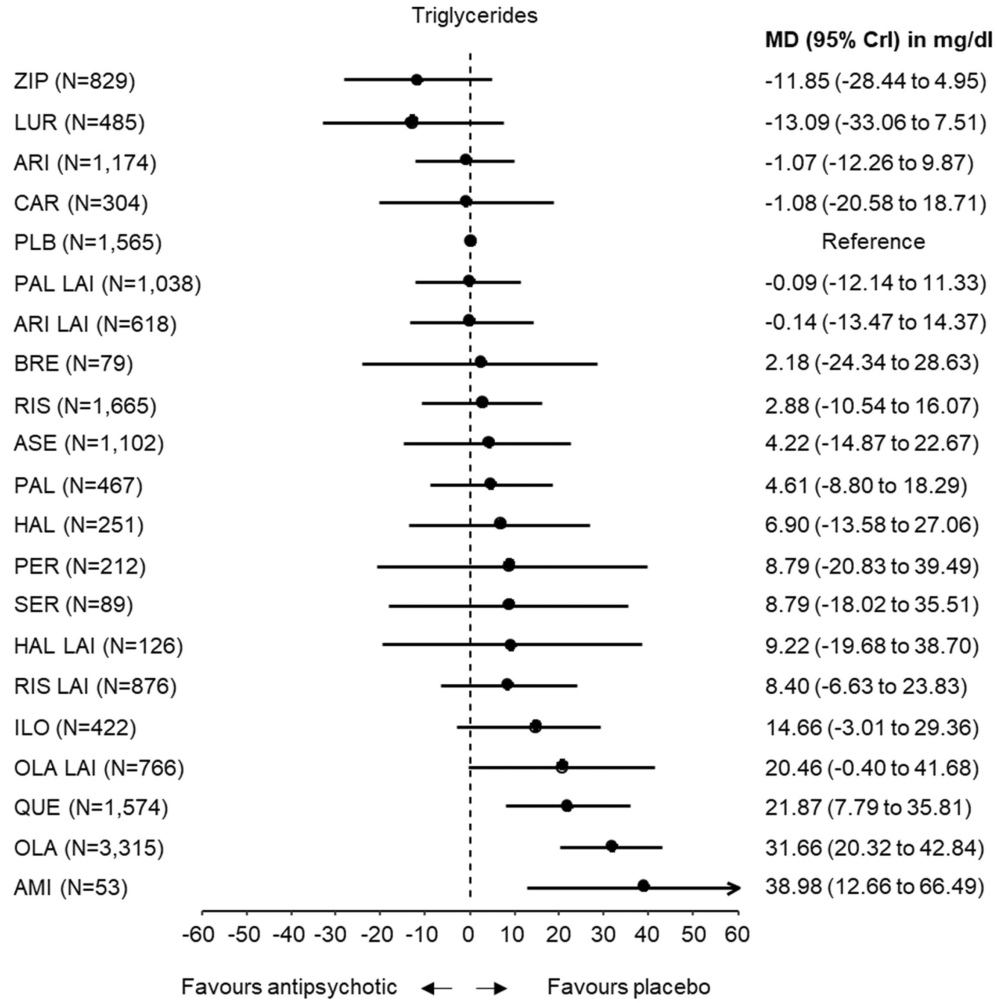

Figure 8 Forest plot of antipsychotic drugs vs. placebo for the secondary outcome "triglycerides". Network meta-analysis estimates of treatment effect of each drug vs. placebo are reported as mean differences (MDs) and 95% credible intervals (CrIs). The order of treatments is according to surface under the cumulative ranking curve (SUCRA) ranking. LAI – long-acting injectable, AMI – amisulpride, ARI – aripiprazole, ASE – asenapine, BRE – brexpiprazole, CAR – cariprazine, HAL – haloperidol, ILO – iloperidone, LUR – lurasidone, OLA – olanzapine, PAL – paliperidone, PER – perphenazine, PLB – placebo, QUE – quetiapine, RIS – risperidone, SER – sertindole, ZIP – ziprasidone.

tional weight increase after 6 weeks was found only for olanzapine and first-generation antipsychotics as a group. In a population-based cohort study in UK primary care[48], more weight gain was observed during the first 6 weeks of treatment than in the following 4 years. For example, men treated with olanzapine (>5 mg/day) gained on average 4.5 kg in the first six weeks, but only 1.4 kg thereafter. In 573 patients treated with olanzapine for a median of 2.5 years[49], an average weight gain of 6.26 kg was observed, which plateaued at 39 weeks (compared to 3.82 kg in our meta-analysis with 6,156 study participants treated with olanzapine for a median of 26 weeks). Taken together, these results suggest that antipsychotic-induced weight gain stagnates over time[5,48-51].

Several considerations and limitations need to be taken into account when interpreting our results. First, there is evidence that antipsychotic-naïve individuals are more vulnerable to weight gain[26,52], but only 11% of our studies included participants with minimal prior exposure to antipsychotics, although excluding these subjects in a sensitivity analysis did not materially change the results.

Second, 140 (older) studies on FGAs and 99 studies on at least one SGA met our inclusion criteria, but did not report weight gain. This missing information led to downgrading the certainty in results with CINeMA, regardless the primary study aim and publication year, although without the original protocol we cannot state whether these outcomes were not measured or not reported.

Third, enriched designs in which patients are stabilized on the drug under investigation before randomization may lead to ceiling effects. Excluding these studies (22/110, 20%) in a sensitivity analysis led to 0.63 kg more weight gain on average, with the most extreme result for iloperidone (1.97 kg versus 0.78 kg in the primary analysis).

Finally, the high dropout rates in long-term studies are a major concern (42% here). The classical last-observation-carried-

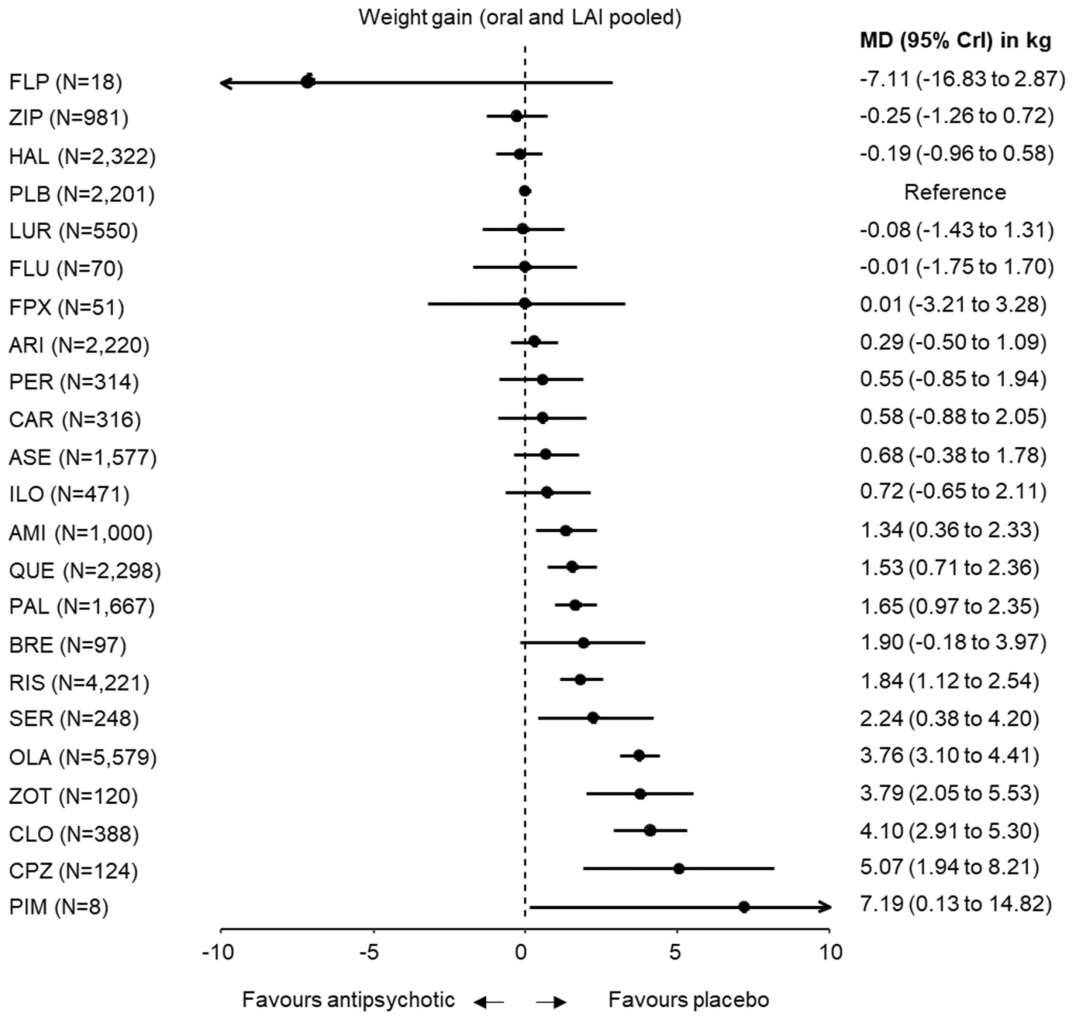

Figure 9 Forest plot of antipsychotic drugs vs. placebo for the post-hoc sensitivity analysis on weight gain pooling long-acting injectable (LAI) and oral formulations. MD – mean difference, 95% CrI – 95% credible interval, LAI – long-acting injectable, AMI – amisulpride, ARI – aripiprazole, ASE – asenapine, BRE – brexpiprazole, CAR – cariprazine, CLO – clozapine, CPZ – chlorpromazine, FLP – fluspirilene, FLU – fluphenazine, FPX – flupentixol, HAL – haloperidol, ILO – iloperidone, LUR – lurasidone, OLA – olanzapine, PAL – paliperidone, PER – perphenazine, PIM – pimozide, PLB – placebo, QUE – quetiapine, RIS – risperidone, SER – sertindole, ZIP – ziprasidone, ZOT – zotepine.

forward (LOCF) method underestimates the total weight gain, because the last measurement before dropout is used, which reflects an exposure period shorter than the planned study duration. More sophisticated models such as mixed models of repeated measures (MMRM) try to implement missing data (used by 6/110 studies included here). In our sensitivity analysis including only observed cases, antipsychotics with substantial weight gain in the primary analysis had a somewhat more pronounced effect. Nevertheless, this analysis cannot account for patients who dropped out due to weight gain.

We conclude that antipsychotics differ clearly in weight gain and metabolic parameters in mid- to long-term treatment. The magnitude of the differences in fasting glucose and lipid parameters was approximately the same as previously reported for short-term studies, suggesting that these effects occur quickly. Differences in weight gain were more pronounced compared to previously published short-term data. However, the overall evidence seems to suggest that weight gain is most pronounced at the beginning of treatment and then remains somewhat stable. Long-term studies with initially antipsychotic-naïve participants are needed.

Although the results were robust to several potential confounders, there was substantial interindividual variability, which could be explored by individual participant data meta-analysis, and should be considered in treatment decisions.

ACKNOWLEDGEMENTS

This work was supported by the German Ministry for Education and Research (grant no. FKZ 01KG1904). V. Chiocchia and G. Salanti have been supported by the Swiss National Science Foundation (grant no. 179185). The funding sources had no role in the design and conduct of the study, and the preparation, approval and decision to submit the manuscript. The authors thank F. Shokraneh,

who conducted the first literature search, F. Krayer for technical support, and all the authors of the included studies, particularly R. Emsley, Y. Koshikawa, L. San and G.D. Kotzalidis, as well as Janssen/Johnson & Johnson (via YODA Project #2020-4517), Eli Lilly, Vanda and Gedeon Richter, for providing additional data. The interpretation and reporting of data are solely the responsibility of the authors and do not necessarily represent the views of the data sharing agencies. A. Burschinski and J. Schneider-Thoma contributed equally to this work. Supplementary information on the study is available at https://ebmpp.org/fileadmin/resources/files/Appendix_Metabolic.pdf.

REFERENCES

1. Huhn M, Nikolakopoulou A, Schneider-Thoma J et al. Comparative efficacy and tolerability of 32 oral antipsychotics for the acute treatment of adults with multi-episode schizophrenia: a systematic review and network meta-analysis. Lancet 2019;394:939-51.
2. Leucht S, Tardy M, Komossa K et al. Antipsychotic drugs versus placebo for relapse prevention in schizophrenia: a systematic review and meta-analysis. Lancet 2012;379:2063-71.
3. De Hert M, Sermon J, Geerts P et al. The use of continuous treatment versus placebo or intermittent treatment strategies in stabilized patients with schizophrenia: a systematic review and meta-analysis of randomized controlled trials with first- and second-generation antipsychotics. CNS Drugs 2015;29:637-58.
4. Takeuchi H, Suzuki T, Uchida H et al. Antipsychotic treatment for schizophrenia in the maintenance phase: a systematic review of the guidelines and algorithms. Schizophr Res 2012;134:219-25.
5. De Hert M, Detraux J, van Winkel R et al. Metabolic and cardiovascular adverse effects associated with antipsychotic drugs. Nat Rev Endocrinol 2011;8: 114-26.
6. Kaur J. A comprehensive review on metabolic syndrome. Cardiol Res Pract 2014;2014:943162.
7. Mottillo S, Filion KB, Genest J et al. The metabolic syndrome and cardiovascular risk a systematic review and meta-analysis. J Am Coll Cardiol 2010; 56:1113-32.
8. Correll CU, Solmi M, Veronese N et al. Prevalence, incidence and mortality from cardiovascular disease in patients with pooled and specific severe mental illness: a large-scale meta-analysis of 3,211,768 patients and 113,383,368 controls. World Psychiatry 2017;16:163-80.
9. Hjorthøj C, Stürup AE, McGrath JJ et al. Years of potential life lost and life expectancy in schizophrenia: a systematic review and meta-analysis. Lancet Psychiatry 2017;4:295-301.
10. Allison DB, Mackell JA, McDonnell DD. The impact of weight gain on quality of life among persons with schizophrenia. Psychiatr Serv 2003;54:565-7.
11. Perkins DO. Predictors of noncompliance in patients with schizophrenia. J Clin Psychiatry 2002;63:1121-8.
12. Weiden PJ, Mackell JA, McDonnell DD. Obesity as a risk factor for antipsychotic noncompliance. Schizophr Res 2004;66:51-7.
13. Tandon R, Belmaker RH, Gattaz WF et al. World Psychiatric Association Pharmacopsychiatry Section statement on comparative effectiveness of antipsychotics in the treatment of schizophrenia. Schizophr Res 2008;100:20-38.
14. National Collaborating Centre for Mental Health. Psychosis and schizophrenia in adults: treatment and management. London: National Institute for Health and Care Excellence, 2014.
15. American Psychiatric Association. The American Psychiatric Association practice guideline for the treatment of patients with schizophrenia. Washington: American Psychiatric Association Publishing, 2020.
16. Pillinger T, McCutcheon RA, Vano L et al. Comparative effects of 18 antipsychotics on metabolic function in patients with schizophrenia, predictors of metabolic dysregulation, and association with psychopathology: a systematic review and network meta-analysis. Lancet Psychiatry 2020;7:64-77.
17. Schneider-Thoma J, Kapfhammer A, Wang D et al. Metabolic side effects of antipsychotic drugs in individuals with schizophrenia during medium- to long-term treatment: protocol for a systematic review and network meta-analysis of randomized controlled trials. Syst Rev 2021;10:214.
18. Cochrane Schizophrenia. https://schizophrenia.cochrane.org/.
19. Tong Z, Li F, Ogawa Y et al. Quality of randomized controlled trials of new generation antidepressants and antipsychotics identified in the China National Knowledge Infrastructure (CNKI): a literature and telephone interview study. BMC Med Res Methodol 2018;18:96.
20. Woodhead M. 80% of China's clinical trial data are fraudulent, investigation finds. BMJ 2016;355:i5396.
21. Parry J. China vows to clamp down on academic fraud amid medical journal scandal. BMJ 2017;357:j2970.
22. Leucht S, Davis JM. Which first-generation antipsychotics should be "repurposed" for the treatment of schizophrenia. Eur Arch Psychiatry Clin Neurosci 2022;272:1-3.
23. Misawa F, Kishimoto T, Hagi K et al. Safety and tolerability of long-acting injectable versus oral antipsychotics: a meta-analysis of randomized controlled studies comparing the same antipsychotics. Schizophr Res 2016;176:220-30.
24. Correll CU, Lencz T, Malhotra AK. Antipsychotic drugs and obesity. Trends Mol Med 2011;17:97-107.
25. Gardner DM, Murphy AL, O'Donnell H et al. International consensus study of antipsychotic dosing. Am J Psychiatry 2010;167:686-93.
26. Bak M, Fransen A, Janssen J et al. Almost all antipsychotics result in weight gain: a meta-analysis. PLoS One 2014;9:e94112.
27. Zhang J-P, Gallego JA, Robinson DG et al. Efficacy and safety of individual second-generation vs. first-generation antipsychotics in first-episode psychosis: a systematic review and meta-analysis. Int J Neuropsychopharmacol 2013;16:1205-18.
28. Spertus J, Horvitz-Lennon M, Abing H et al. Risk of weight gain for specific antipsychotic drugs: a meta-analysis. NPJ Schizophr 2018;4:12.
29. Barton BB, Segger F, Fischer K et al. Update on weight-gain caused by antipsychotics: a systematic review and meta-analysis. Expert Opin Drug Saf 2020;19:295-314.
30. Rummel-Kluge C, Komossa K, Schwarz S et al. Head-to-head comparisons of metabolic side effects of second generation antipsychotics in the treatment of schizophrenia: a systematic review and meta-analysis. Schizophr Res 2010;123:225-33.
31. Pagsberg AK, Tarp S, Glintborg D et al. Acute antipsychotic treatment of children and adolescents with schizophrenia-spectrum disorders: a systematic review and network meta-analysis. J Am Acad Child Adolesc Psychiatry 2017;56:191-202.
32. De Hert M, Yu W, Detraux J et al. Body weight and metabolic adverse effects of asenapine, iloperidone, lurasidone and paliperidone in the treatment of schizophrenia and bipolar disorder: a systematic review and exploratory meta-analysis. CNS Drugs 2012;26:733-59.
33. Zhang Y, Liu Y, Su Y et al. The metabolic side effects of 12 antipsychotic drugs used for the treatment of schizophrenia on glucose: a network meta-analysis. BMC Psychiatry 2017;17:373.
34. Sterne JAC, Savović J, Page MJ et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019;2:l4898.
35. Turner RM, Jackson D, Wei Y et al. Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. Stat Med 2015;34:984-98.
36. Rhodes KM, Turner RM, Higgins JPT. Predictive distributions were developed for the extent of heterogeneity in meta-analyses of continuous outcome data. J Clin Epidemiol 2015;68:52-60.
37. Dias S, Welton NJ, Caldwell DM et al. Checking consistency in mixed treatment comparison meta-analysis. Stat Med 2010;29:932-44.
38. Higgins JPT, Jackson D, Barrett JK et al. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. Res Synth Methods 2012;3:98-110.
39. Chaimani A, Salanti G. Using network meta-analysis to evaluate the existence of small-study effects in a network of interventions. Res Synth Methods 2012;3:161-76.
40. Peters JL, Sutton AJ, Jones DR et al. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. J Clin Epidemiol 2008;61:991-6.
41. Béliveau A, Boyne DJ, Slater J et al. BUGSnet: an R package to facilitate the conduct and reporting of Bayesian network meta-analyses. BMC Med Res Methodol 2019;19:196.
42. Plummer M. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. Presented at the 3rd International Workshop on Distributed Statistical Computing, Vienna, March 2003.
43. Rücker G, König J, Efthimiou O et al. Package 'netmeta': network meta-analysis using frequentist methods. https://cran.r-project.org/web/packages/netmeta/netmeta.pdf.
44. Schwarzer G. Package 'meta': general package for meta-analysis. https://cran.r-project.org/web/packages/meta/meta.pdf.
45. Nikolakopoulou A, Higgins JPT, Papakonstantinou T et al. CINeMA: an approach for assessing confidence in the results of a network meta-analysis. PLoS Med 2020;17:e1003082.
46. Willett WC, Manson JE, Stampfer MJ et al. Weight, weight change, and coronary heart disease in women. Risk within the 'normal' weight range. JAMA 1995;273:461-5.

47. McEvoy JP, Byerly M, Hamer RM et al. Effectiveness of paliperidone palmitate vs haloperidol decanoate for maintenance treatment of schizophrenia: a randomized clinical trial. JAMA 2014;311:1978-87.
48. Bazo-Alvarez JC, Morris TP, Carpenter JR et al. Effects of long-term antipsychotics treatment on body weight: a population-based cohort study. J Psychopharmacol 2020;34:79-85.
49. Kinon BJ, Basson BR, Gilmore JA et al. Long-term olanzapine treatment: weight change and weight-related health factors in schizophrenia. J Clin Psychiatry 2001;62:92-100.
50. Millen BA, Campbell GM, Beasley CM. Weight changes over time in adults treated with the oral or depot formulations of olanzapine: a pooled analysis of 86 clinical trials. J Psychopharmacol 2011;25:639-45.
51. Pérez-Iglesias R, Martínez-García O, Pardo-Garcia G et al. Course of weight gain and metabolic abnormalities in first treated episode of psychosis: the first year is a critical period for development of cardiovascular risk factors. Int J Neuropsychopharmacol 2014;17:41-51.
52. Alvarez-Jiménez M, González-Blanch C, Crespo-Facorro B et al. Antipsychotic-induced weight gain in chronic and first-episode psychotic disorders: a systematic critical reappraisal. CNS Drugs 2008;22:547-62.

DOI:10.1002/wps.21036

Candidate diagnostic biomarkers for neurodevelopmental disorders in children and adolescents: a systematic review

Samuele Cortese[1-5], Marco Solmi[1,6-9], Giorgia Michelini[10,11], Alessio Bellato[12], Christina Blanner[13], Andrea Canozzi[14], Luis Eudave[15], Luis C. Farhat[16], Mikkel Højlund[17,18], Ole Köhler-Forsberg[19,20], Douglas Teixeira Leffa[21,22], Christopher Rohde[20,23], Gonzalo Salazar de Pablo[24-27], Giovanni Vita[14], Rikke Wesselhoeft[18,28], Joanna Martin[29], Sarah Baumeister[30], Natali S. Bozhilova[24,31], Christina O. Carlisi[32], Virginia Carter Leno[33], Dorothea L. Floris[34,35], Nathalie E. Holz[30,35-37], Eline J. Kraaijenvanger[30], Seda Sacu[30], Isabella Vainieri[38], Giovanni Ostuzzi[14], Corrado Barbui[14], Christoph U. Correll[9,39-41]

[1]Centre for Innovation in Mental Health, School of Psychology, Faculty of Environmental and Life Sciences, University of Southampton, Southampton, UK; [2]Clinical and Experimental Sciences (CNS and Psychiatry), Faculty of Medicine, University of Southampton, Southampton, UK; [3]Solent NHS Trust, Southampton, UK; [4]Hassenfeld Children's Hospital at NYU Langone, New York University Child Study Center, New York, NY, USA; [5]Division of Psychiatry and Applied Psychology, School of Medicine, University of Nottingham, Nottingham, UK; [6]Department of Psychiatry, University of Ottawa, Ottawa, ON, Canada; [7]Department of Mental Health, Ottawa Hospital, Ottawa, ON, Canada; [8]Ottawa Hospital Research Institute (OHRI) Clinical Epidemiology Program, University of Ottawa, Ottawa, ON, Canada; [9]Department of Child and Adolescent Psychiatry, Charité Universitätsmedizin, Berlin, Germany; [10]Department of Biological & Experimental Psychology, School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK; [11]Department of Psychiatry and Biobehavioral Sciences, Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles (UCLA), Los Angeles, CA, USA; [12]School of Psychology, University of Nottingham, Semenyih, Malaysia; [13]Mental Health Center, Glostrup, Copenhagen University Hospital - Mental Health Services CPH, Copenhagen, Denmark; [14]Department of Neuroscience, Biomedicine, and Movement Sciences, Section of Psychiatry, University of Verona, Verona, Italy; [15]Faculty of Education and Psychology, University of Navarra, Pamplona, Spain; [16]Department of Psychiatry, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brazil; [17]Department of Psychiatry Aabenraa, Mental Health Services in the Region of Southern Denmark, Aabenraa, Denmark; [18]Clinical Pharmacology, Pharmacy, and Environmental Medicine, Department of Public Health, University of Southern Denmark, Odense, Denmark; [19]Psychosis Research Unit, Aarhus University Hospital - Psychiatry, Aarhus, Denmark; [20]Department of Clinical Medicine, Aarhus University, Aarhus, Denmark; [21]ADHD Outpatient Program & Development Psychiatry Program, Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul, Brazil; [22]Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA, USA; [23]Department of Affective Disorders, Aarhus University Hospital - Psychiatry, Aarhus, Denmark; [24]Department of Child and Adolescent Psychiatry, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK; [25]Child and Adolescent Mental Health Services, South London and Maudsley NHS Foundation Trust, London, UK; [26]Early Psychosis: Interventions and Clinical-detection (EPIC) Lab, Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK; [27]Institute of Psychiatry and Mental Health, Department of Child and Adolescent Psychiatry, Hospital General Universitario Gregorio Marañón School of Medicine, Universidad Complutense, CIBERSAM, Madrid, Spain; [28]Child and Adolescent Mental Health Odense, Mental Health Services in the Region of Southern Denmark, Odense, Denmark; [29]MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, UK; [30]Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health, University of Heidelberg, Mannheim, Germany; [31]School of Psychology, University of Surrey, Guilford, UK; [32]Division of Psychology and Language Sciences, University College London, London, UK; [33]Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK; [34]Department of Psychology, University of Zurich, Zurich, Switzerland; [35]Donders Institute for Brain, Cognition, and Behavior, Radboud University Nijmegen, Nijmegen, The Netherlands; [36]Department for Cognitive Neuroscience, Radboud University Medical Center Nijmegen, Nijmegen, The Netherlands; [37]Institute of Medical Psychology and Medical Sociology, University Medical Center Schleswig Holstein, Kiel University, Kiel, Germany; [38]Department of Clinical, Educational and Health Psychology, University College London, London, UK; [39]Psychiatry Research, Northwell Health, Zucker Hillside Hospital, New York, NY, USA; [40]Department of Psychiatry and Molecular Medicine, Zucker School of Medicine, Hempstead, NY, USA; [41]Center for Neuroscience, Feinstein Institute for Medical Research, Manhasset, NY, USA

Neurodevelopmental disorders – including attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorder, communication disorders, intellectual disability, motor disorders, specific learning disorders, and tic disorders – manifest themselves early in development. Valid, reliable and broadly usable biomarkers supporting a timely diagnosis of these disorders would be highly relevant from a clinical and public health standpoint. We conducted the first systematic review of studies on candidate diagnostic biomarkers for these disorders in children and adolescents. We searched Medline and Embase + Embase Classic with terms relating to biomarkers until April 6, 2022, and conducted additional targeted searches for genome-wide association studies (GWAS) and neuroimaging or neurophysiological studies carried out by international consortia. We considered a candidate biomarker as promising if it was reported in at least two independent studies providing evidence of sensitivity and specificity of at least 80%. After screening 10,625 references, we retained 780 studies (374 biochemical, 203 neuroimaging, 133 neurophysiological and 65 neuropsychological studies, and five GWAS), including a total of approximately 120,000 cases and 176,000 controls. While the majority of the studies focused simply on associations, we could not find any biomarker for which there was evidence – from two or more studies from independent research groups, with results going into the same direction – of specificity and sensitivity of at least 80%. Other important metrics to assess the validity of a candidate biomarker, such as positive predictive value and negative predictive value, were infrequently reported. Limitations of the currently available studies include mostly small sample size, heterogeneous approaches and candidate biomarker targets, undue focus on single instead of joint biomarker signatures, and incomplete accounting for potential confounding factors. Future multivariable and multi-level approaches may be best suited to find valid candidate biomarkers, which will then need to be validated in external, independent samples and then, importantly, tested in terms of feasibility and cost-effectiveness, before they can be implemented in daily clinical practice.

Key words: Biological markers, neurodevelopmental disorders, ADHD, autism spectrum disorder, communication disorders, intellectual disability, motor disorders, specific learning disorders, tic disorders, genome-wide association studies, neuroimaging, neurophysiology

(World Psychiatry 2023;22:129–149)

Limitations related to the subjective nature of psychiatric diagnoses have prompted, in the past decades, several lines of investigation aimed at identifying valid biomarkers that can assist in the diagnosis, prediction, prognosis and management of mental health conditions.

According to the US Food and Drug Administration (FDA) - National Institute of Health (NIH) Biomarker Working Group, a *biomarker* is defined as "a characteristic that is measured as an indicator of normal biological processes, pathogenic processes or responses to an exposure or intervention"[1]. Based on their main clinical application, biomarkers can be grouped as: a) *diagnostic*, used to detect or confirm the presence of a disease or medical condition or to identify homogeneous subtypes of the disease; b) *monitoring*, to monitor the status of a disease and the

response to a treatment; c) *pharmacodynamic*, to evaluate the response to a clinical intervention; d) *predictive*, to predict the probability to develop any effect following a clinical intervention; e) *prognostic*, to identify the probability of developing a clinical event in individuals with a disease or a clinical condition; f) *safety*, to evaluate the probability of developing an adverse event following an intervention; and g) *susceptibility/risk*, to quantify the risk of an individual to develop a disease or medical condition[2].

Valid and usable at scale biomarkers, if identified, promise to allow the clinical implementation of precision medicine in psychiatry[2-7], whereby: a) individual patients would receive the proper diagnosis, and therefore proper treatment, more quickly; b) they would be matched more accurately to the treatments they are most likely to respond to; c) treatment could be started before symptoms reach a severe level and/or lead to dysfunction, increasing the likelihood of expedited recovery; d) clinicians could more easily identify who is most at risk for relapse and recurrence.

However, the path for the identification of a biological characteristic as a valid biomarker in real-world clinical settings is a long one, and needs to follow rigorous steps. The biomarker needs first to be *sensitive*, i.e., accurately identify as positive those individuals who have the outcome of interest, and *specific*, namely, accurately label as negative those individuals who do not have the outcome of interest. Although there are no established benchmarks for these metrics, quantitative measures that allow diagnostic accuracy with at least 80% sensitivity and 80% specificity are often considered as clinically useful[8].

The consensus report by the American Psychiatric Association (APA) Work Group on Neuroimaging Markers of Psychiatric Disorders suggested that a promising biomarker should have two or more independent well-powered studies providing evidence of sensitivity and specificity at least of 80%[9]. In addition, a biomarker would need to: a) have good *positive predictive value* (PPV), which refers to the proportion of individuals who have the outcome of interest among those who tested positive; b) have good *negative predictive value* (NPV), indicating the proportion of individuals who do not have the outcome of interest among those who tested negative; c) have good *internal validity*, i.e., measure the intended feature in an unbiased way, without relevant influence of confounding factors; d) be *externally valid*, so that the results of the studies assessing the candidate biomarker are generalizable to the population of interest in real-life clinical settings; and e) be reliable, in terms of *test-retest reliability* (i.e., being consistent with itself when measured on several occasions) and *inter-rater reliability* (i.e., being consistent when measured across different raters)[10]. Furthermore, a biomarker should change in a dynamic and reliable way in relation to the progress/change of the clinical condition[2].

Steps for biomarker discovery should therefore include an initial phase where a clinically relevant question is identified; a phase testing internal validity, ruling out the possible role of confounding factors; a subsequent phase where external validity is tested, assessing PPV and NPV in independent, targeted samples; and a last phase where the biomarker is tested to assess whether it brings a significant benefit in relation to standard clinical practice, with acceptable number needed to assess (NNA) and number needed to treat (NNT), i.e. the number of individuals that should be assessed or treated in order to benefit one additional individual compared to those who are not assessed or treated. Crucially, this last phase should also assess if the biomarker is cost-effective in relation to standard practice[10].

Based on the pathophysiological overlap across disorders, it has been suggested that at least some of the candidate biomarkers may have a transdiagnostic nature across mental health conditions[11]. However, for at least some peripheral biomarkers, it is possible that their transdiagnostic nature be related to the chronic stress or allostatic load associated with a variety of psychiatric conditions[12]. The notion of *transdiagnosticity* of peripheral biomarkers has been supported by a systematic review showing that, out of the six molecules most commonly referred to as "biomarkers" in studies of schizophrenia, major depressive disorder and bipolar disorder, five – brain-derived neurotrophic factor (BDNF), tumor necrosis factor (TNF)-alpha, interleukin (IL)-6, C-reactive protein (CRP), and cortisol – were proposed across these disorders[12], even though without a rigorous transdiagnostic framework. Furthermore, a systematic review and meta-analysis of electrophysiological correlates of performance monitoring in four common childhood disorders – attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorder (ASD), Tourette's syndrome, and obsessive-compulsive disorder – found a significant overlap in electrophysiological correlates across these disorders[13,14].

Recent umbrella reviews have shown that, in the case of many putative biomarkers for ASD and ADHD, most meta-analyses claiming significant associations were likely inflated by high risk of bias, including excess of significance bias[15-17]. By pooling different studies and increasing power, meta-analyses frequently find significant results. However, in this specific field, what determines the credibility of a diagnostic biomarker is replication of findings in terms of specificity, sensitivity, accuracy, and predictive value[9], rather than a pooled effect size of association. Hence, a systematic review accounting for these variables is needed. In the present systematic review, we focus on diagnostic biomarkers of neurodevelopmental disorders, alongside oppositional defiant disorder (ODD) and conduct disorders (CD), in children and adolescents.

Neurodevelopmental disorders is an umbrella term encompassing a broad range of conditions characterized by impaired development of cognitive, social or motor functions, or atypical functioning, usually manifesting themselves from early childhood, and having a steady course without marked remissions or relapses[18,19]. The conceptualization and grouping of these disorders have changed over time and are still a matter of debate. Currently, the ICD-11[20] includes ADHD, ASD, communication disorders, intellectual disability, motor disorders, specific learning disorders (involving reading, writing and arithmetic), and tic disorders.

Neurodevelopmental disorders are highly heterogeneous in terms of their epidemiology[21], clinical characteristics, causes[22], burden, treatment responses and tolerability[23,24], and outcomes[25]. Notably, ODD and CD are often comorbid with neurodevelopmental disorders, in particular ADHD[26].

The level of overlap between neurodevelopmental disorders

and their symptom dimensions is substantial. This is accounted for by shared or correlated risk factors, and common or overlapping molecular and neuronal mechanisms. While this co-occurrence supports the rationale for grouping these disorders together, from a clinical standpoint it is also relevant to recognize them as individual entities. Indeed, specific, distinct diagnostic categories allow clinicians to communicate about patients' characteristics with each other and with the patients and their family members/caregivers. Furthermore, patients with different categorical diagnoses respond to different treatments. For instance, psychostimulants are effective for ADHD, and so-called antipsychotics can decrease the severity of tics, but psychostimulants are not effective for tics, and antipsychotics do not improve attention regulation difficulties of ADHD.

While previous systematic reviews, meta-analyses or umbrella reviews have provided a synthesis of the evidence on specific biomarkers in specific disorders, for example on peripheral biomarkers in ADHD[16,27] or ASD[15], no systematic review has been conducted so far covering a broad range of biomarkers across neurodevelopmental disorders.

We aimed to fill this gap by conducting a systematic review of studies on promising candidate diagnostic biomarkers in children and/or adolescents with any neurodevelopmental disorder or with ODD or CD. We aimed to assess: a) which are the candidate biochemical, genetic, neuroimaging, neurophysiological and neuropsychological biomarkers that have been replicated across studies as being significantly associated with the diagnosis of specific neurodevelopmental disorders; b) how many of these biomarkers could be defined as *promising*, based on specificity and sensitivity at least of 80% in two or more independent studies; and c) for how many of these candidate biomarkers, internal as well as external validation – assessing sensitivity, specificity, PPV and NPV – have been implemented, alongside an evaluation of the cost-effectiveness of the biomarker; and d) to what extent biomarkers are disorder-specific or transdiagnostic.

METHODS

This systematic review was based on a pre-registered protocol (available at https://osf.io/wp4je/?view_only=8c349f45a9ac441490981acf946c8d9a) and was conducted in accordance with the 2020 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement[28].

Search

We searched Medline and Embase + Embase Classic, from inception until April 6, 2022. We did not apply any limit in terms of language or type of document. We used terms related to neurodevelopmental disorders (alongside ODD and CD) and "biomarker" or equivalent ("marker", "diagnostic test", and "endophenotype"), in order to retrieve studies assessing what the study authors deemed to be a potential biomarker. The exact search syntax is reported in the supplementary information.

Additionally, we searched for the largest genome-wide association studies (GWAS), as GWAS are typically based on meta-analyses of increasing numbers of samples and, as such, many previous smaller studies are sub-samples of the largest available GWAS, which will be best powered and use the latest methodologies and best practices. We also searched for neuroimaging or neurophysiological studies conducted by international consortia.

Inclusion/exclusion criteria

We included any observational study with a comparison group, assessing children or adolescents (mean age: 18 years or less) presenting with any (one or more) of the following disorders (reported here according to the ICD-11), provided that they were diagnosed using the ICD (9, 10 or 11) or the DSM (III, III-R, IV, IV-TR or 5): 6A00 Disorders of Intellectual Development; 6A01 Developmental Speech or Language Disorders; 6A01.0 Developmental Speech Sound Disorder; 6A01.1 Developmental Speech Fluency Disorder; 6A01.2 Developmental Language Disorder; 6A02 Autism Spectrum Disorder; 6A03 Developmental Learning Disorder; 6A04 Developmental Motor Coordination Disorder; 6A05 Attention Deficit Hyperactivity Disorder; 6A06 Stereotyped Movement Disorder; 6A0Y Other Specified Neurodevelopmental Disorder; 8A05.00 Tourette Syndrome; 8A05.01 Chronic Motor Tic Disorder; 8A05.02 Chronic Phonic Tic Disorder; 6C90 Oppositional Defiant Disorder; 6C91 Conduct-Dissocial Disorder.

For ASD, we also included studies with a diagnosis based on the Autism Diagnostic Observation Schedule (ADOS), that has shown acceptable diagnostic accuracy in research settings[29].

Study selection and data extraction

Two authors independently screened titles and abstracts, and any conflicts were resolved by a third senior author. All selected articles underwent full text screening by two authors independently, with conflicts resolved by consultation with a third senior author.

For each retained study, we extracted the following variables: first author, year of publication, design (cross-sectional or longitudinal), specific disorder(s) included, diagnostic criteria, number and age of cases and controls, percentage of males, percentage of White ethnicity individuals, type of biomarker(s), most adjusted effect size or p value, and inclusion of any of the following, when available: sensitivity, specificity, PPV, NPV, and receiver operating characteristic area under the curve (ROC AUC).

Study quality appraisal

We rated the quality of cross-sectional studies using BIOCROSS, an appraisal tool for cross-sectional studies using biomarker data (no tools for longitudinal studies of biomarkers are available)[30]. The following items were selected as the most appropriate for

the appraisal of studies of biochemical biomarkers: item 3 (3.1: "Was the sampling frame reported (study population source)?"; 3.2: "Was the participation rate reported (i.e., eligible persons at least 50%)?"; 3.3: "Was sample size justification or power description provided?"); item 4 (4.1: "Were the study population characteristics (i.e., demographic, clinical and social) presented?"; 4.2: "Were the exposures and potential confounders described?"; 4.3: "Were any missing values and strategies to deal with missing data reported?"); item 5 (5.1: "Did the authors clearly report statistical methods used to calculate estimates (e.g., Spearman, Pearson, linear regression)?"; 5.2: "Were key potential confounding variables measured and adjusted statistically in reported analyses?"; 5.3: "Was the raw effect size estimate (correlation coefficient, beta coefficient) or measure of study precision provided (e.g., confidence intervals, precise p value)?"); item 8 (8.1: "Were the measurement methods described (assay methods, preservation and storage, detailed protocol, including specific reagents or kits used)?"; 8.2: "Were the reproducibility assessments performed for evaluating biomarker stability?"; 8.3: "Were the quantitation methods well described?"); item 9 (9.1: "Was the laboratory/place of measurement mentioned?"; 9.2: "Were any quality control procedures and results reported (e.g., reported coefficient of variation)?"; 9.3: "Were the analyses blinded for laboratory staff?"). We selected items 3, 4, 5, 8 and 9, with exclusion of sub-items 4.2, 8.2, 9.1 and 9.3, for neuroimaging, neurophysiological and neuropsychological studies. We selected items 3, 4, 5 and 8, with exclusion of sub-item 8.3, for GWAS.

Synthesis of the evidence

We provided a qualitative synthesis of the included studies and of the level of transdiagnosticity. To assess promising biomarkers, we indicated first, when possible, the number and frequency of positive and negative replications (with the direction of the association, i.e. *increased* or *decreased*) for each biomarker assessed in at least two studies, with at least one positive finding in terms of significant associations. We then identified the biomarkers for which at least two studies reported on sensitivity, specificity, PPV, NPV and/or ROC AUC, and the biomarkers with a sensitivity and specificity of at least 80% replicated in at least two studies.

RESULTS

From an initial pool of 10,625 references, we retained 780 studies (see Figure 1, reporting the PRISMA 2020 flow chart[31]). The lists of included references and of those excluded, with reasons for exclusion after checking the full text, are reported in the supplementary information.

We present the findings in relation to each type of candidate biomarker (now onwards, for simplicity, referred to as "biomarker"), based on the primary outcome of the study (for instance, a study assessing a neurophysiological biomarker as primary outcome but including also biochemical biomarkers is reported under the section "Neurophysiology").

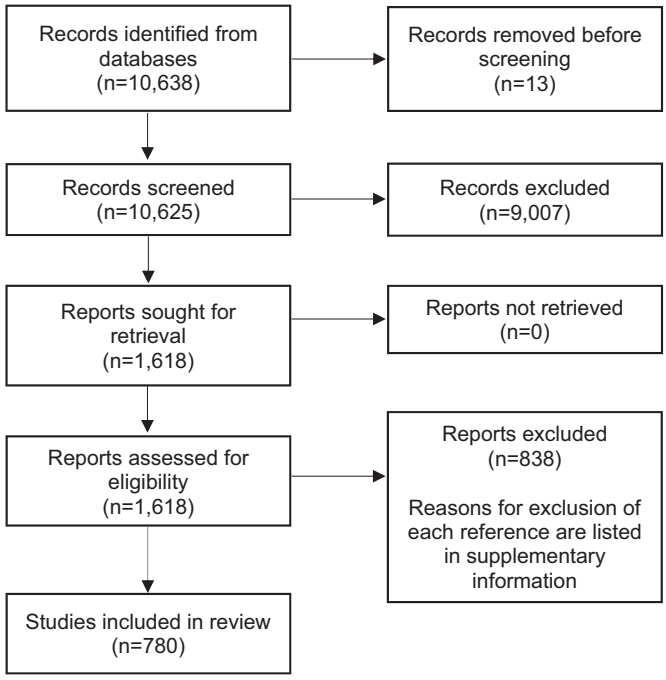

Figure 1 PRISMA 2020 flow diagram showing selection of studies for inclusion

Biochemical biomarkers

We included a total of 374 studies (359 cross-sectional and 15 longitudinal), 370 of which conducted in 58 individual countries and four in multiple countries, encompassing a total of 26,715 cases and 41,903 controls, and investigating 1,427 biomarkers (see supplementary information).

The average total BIOCROSS score (for cross-sectional studies) was 5.1 (out of 10). The average scores were 0.7 for item 3; 1.1 for item 4; 1.5 for item 5; 1.4 for item 8; and 0.5 for item 9. Therefore, the most concerning methodological issues of the included studies were related to the lack of reporting of sampling frame, participation rate and power calculation, as well as of quality procedures and blinding of the laboratory staff.

The included studies focused on a variety of biochemical biomarkers, including neurotransmitters (e.g., dopamine), hormones (e.g., oxytocin), inflammatory markers (e.g., IL-6), heavy metals (e.g., iron), antioxidants (e.g., vitamin E), and detoxifying agents (e.g., cytochrome P450 oxidase). We summarize below the findings for each neurodevelopmental disorder.

ADHD

We retained 53 studies (51 cross-sectional and two longitudinal), reported in 54 papers, from 19 countries, including a total of 4,164 participants with ADHD and 7,363 controls.

The average total BIOCROSS score was 4.9 (out of 10). The average scores were 0.8 for item 3; 1.0 for item 4; 1.3 for item 5; 1.2

for item 8; and 0.4 for item 9. Therefore, in line with the ratings across all studies of biochemical markers, the most concerning aspects were in relation to the lack of reporting of sampling frame, participation rate and power calculation, as well as of quality procedures and blinding of the laboratory staff.

The included studies assessed, collectively, 229 biomarkers (see supplementary information). Of these, 24 biomarkers were investigated in at least two studies, with at least one positive finding (see Table 1). Biomarkers with positive replications only, without negative findings, in the same direction (i.e., increased in ADHD vs. controls, or decreased in ADHD vs. controls) included: copper (two studies, increased in ADHD compared to neurotypical participants); malondialdehyde, one of the final products of polyunsaturated fatty acids peroxidation in the cells (two studies, increased); mean platelet volume (three studies, increased); and zinc (two studies, decreased).

For 28 biomarkers, one or more of the following metrics were investigated: specificity, sensitivity, PPV, NPV, and ROC AUC. However, only for mean platelet volume these metrics were available from at least two studies. In both studies, specificity and sensitivity were less than 80%, and ROC AUC values were less than 0.8. Therefore, none of the biomarkers for which a significant association with ADHD was detected and replicated, without negative associations, had evidence of a specificity and sensitivity at least of 80% and ROC AUC at least of 0.8 (see also supplementary information).

Autism spectrum disorder

We included 300 studies (289 cross-sectional and 11 longitudinal), reported in 303 papers, from 55 countries, encompassing a total of 20,583 participants with ASD and 33,450 controls. The average total BIOCROSS score was 5.2 (out of 10). The average scores were 0.8 for item 3; 1.0 for item 4; 1.3 for item 5; 1.3 for item 8; and 0.7 for item 9.

The included studies evaluated, overall, 1,298 biomarkers (see supplementary information). Of these, 73 biomarkers were in-

Table 1 Candidate biochemical biomarkers investigated in at least two studies, with at least one positive finding, for attention-deficit/hyperactivity disorder (ADHD)

Biomarker	Number of studies with significant finding	Number of studies with non-significant finding	Direction	Frequency of replication (%)
Copper (urine, hair)	2	0	Increased	100
Malondialdehyde (plasma)	2	0	Increased	100
Mean platelet volume (blood)	3	0	Increased	100
Zinc (urine, hair)	2	0	Decreased	100
Cortisol (saliva, serum)	2	1	Decreased	67
Neutrophil/lymphocyte ratio (blood)	2	1	Increased	67
Oxytocin (serum)	2	1	Decreased	67
Platelet/lymphocyte ratio (blood)	2	1	Increased	67
Folate (blood)	1	1	Decreased	50
Gamma-aminobutyric acid (serum)	1	1	Increased	50
Glial cell line-derived neurotrophic factor (plasma)	1	1	Increased	50
Glutamate (serum)	1	1	Increased	50
Interleukin-6 (plasma)	1	1	Increased	50
Lymphocytes (blood)	1	1	Decreased	50
Melatonin (saliva)	1	1	Decreased	50
Monocyte/lymphocyte ratio (blood)	1	1	Increased	50
Red blood cell distribution width (blood)	1	1	Increased	50
Soluble vascular cell adhesion molecule-1 (plasma)	1	1	Increased	50
Tumor necrosis factor-alpha (plasma)	1	1	Decreased	50
Vitamin B12 (serum)	1	1	Decreased	50
Brain-derived neurotrophic factor (plasma)	2	3	Decreased	40
Neutrophils (blood)	2	1	One increased, one decreased	33
8-hydroxy-2-deoxyguanosine (serum)	1	2	Increased	33
Ferritin (serum)	1	2	Decreased	33

Table 2 Candidate biochemical biomarkers investigated in at least two studies, with at least one positive finding and more than 50% frequency of replication, for autism spectrum disorder

Biomarker	Number of studies with significant finding	Number of studies with non-significant finding	Direction	Frequency of replication (%)
2-aminobutyric acid (urine, plasma)	2	0	Increased	100
2-hydroxybutyric acid (urine)	2	0	Increased	100
8-isoprostane (urine, plasma)	3	0	Increased	100
Adrenic acid (plasma)	2	0	Decreased	100
Alanine (urine, serum)	2	0	Decreased	100
Alpha-1-antitrypsin (plasma)	2	0	Increased	100
Anandamide (serum, plasma)	2	0	Decreased	100
Arachidic acid (serum, plasma)	2	0	Increased	100
Aspartic acid (urine, plasma)	2	0	Decreased	100
Parabacteroides (gut microbiota)	2	0	Increased	100
Creatine kinase (serum, urine)	2	0	Increased	100
Coproporphyrin (urine)	4	0	Increased	100
Cysteine (serum, plasma, urine)	3	0	Decreased	100
Glutamine (blood, serum)	4	0	Decreased	100
Glutathione/oxidized glutathione ratio (serum)	3	0	Decreased	100
High-density lipoprotein (serum)	2	0	Decreased	100
Hippuric acid (urine)	2	0	Increased	100
High-sensitivity C-reactive protein (serum)	2	0	Increased	100
Heat shock protein 70 (serum, plasma)	2	0	Increased	100
Interferon-gamma-inducible protein 16 (serum)	2	0	Increased	100
Kynurenic acid (serum, urine)	2	0	Decreased	100
Lactic acid (urine)	2	0	Increased	100
Lead (urine, hair, red blood cells)	3	0	Increased	100
Neurotensin (serum)	3	0	Increased	100
Para-cresol (urine)	3	0	Increased	100
Peroxiredoxin 1 (serum, plasma)	2	0	Increased	100
Phosphatidylcholine (serum)	2	0	Decreased	100
Pregnenolone sulfate (plasma)	2	0	Decreased	100
Secreted amyloid precursor protein alpha (plasma)	3	0	Increased	100
Succinic acid (urine, plasma)	3	0	Increased	100
Transforming growth factor beta (serum, blood)	3	0	Increased	100
Thiol (serum, urine)	2	0	Decreased	100
Triglycerides (plasma)	2	0	Increased	100
Gamma-aminobutyric acid (blood, plasma, serum)	7	0	Six increased, one decreased	85
Melatonin (serum, plasma, urine)	5	0	One increased, four decreased	80
Dopamine (plasma, blood)	4	0	Three increased, one decreased	75
Glial fibrillary acidic protein (serum)	3	1	Increased	75
Glutathione (serum, plasma)	7	1	One increased, six decreased	75
Potassium (serum)	4	0	One increased, three decreased	75

Table 2 Candidate biochemical biomarkers investigated in at least two studies, with at least one positive finding and more than 50% frequency of replication, for autism spectrum disorder *(continued)*

Biomarker	Number of studies with significant finding	Number of studies with non-significant finding	Direction	Frequency of replication (%)
Leucine (serum)	3	0	Two increased, one decreased	67
Sodium (serum, plasma)	3	0	Two increased, one decreased	67
Antioxidant capacity (urine)	3	0	Two decreased, one increased	67
Arginine vasopressin (cerebrospinal fluid)	2	1	Decreased	67
Catalase (urine, plasma)	2	1	Increased	67
Citric acid (urine, plasma)	3	0	Two increased, one decreased	67
Citrulline (blood, urine)	2	1	Increased	67
Docosahexaeonic acid/arachidonic acid (plasma)	2	1	Increased	67
Epidermal growth factor (plasma)	2	1	Decreased	67
Epinephrine (plasma, blood, gut metabolites)	3	0	Two increased, one decreased	67
Glutamate (serum, blood)	2	1	Increased	67
Hexanol-lysine (urine)	2	1	Increased	67
Hypoxanthine (urine)	2	1	Increased	67
Interleukin-17-A (plasma, serum)	2	1	Increased	67
Indole-3-acetic acid (urine)	2	1	Increased	67
Oxalic acid (urine)	2	1	Increased	67
Oxidized glutathione (plasma)	2	1	Increased	67
Pentacarboxyporphyrin (urine)	2	1	Increased	67
Phosphoric acid (urine)	2	1	Decreased	67
S100 calcium-binding protein B (serum, plasma)	4	2	Increased	67
Tumor necrosis factor-alpha (saliva, serum)	2	1	Increased	67
Thyroid stimulating hormone (serum)	2	1	Decreased	67
Uric acid (serum, urine)	3	0	Two increased, one decreased	67
Vitamin E (plasma)	2	1	Decreased	67
Glutathione S-transferase (serum, plasma)	3	0	One increased, two decreased	67
Brain-derived neurotrophic factor (serum, plasma, blood)	9	1	Six increased, three decreased	67
Cortisol (saliva, plasma, gut metabolites)	3	2	Increased	60
Eicosapentaenoic acid (serum)	3	2	Increased	60
Ferritin (serum)	3	2	Decreased	60
Homocysteine (serum, urine, plasma)	9	1	Six increased, three decreased	60
Interleukin-8 (serum, plasma)	6	4	Increased	60
Creatinine (urine)	4	3	Increased	57
Mercury (blood cells, serum, urine, hair)	4	3	Increased	57
Interleukin-1-beta (plasma)	7	4	Six increased, one decreased	54

vestigated in at least two studies, with at least one positive finding and more than 50% frequency of replication (see Table 2). Biomarkers with positive replications only, without negative findings, in the same direction (i.e., increased in ASD vs. controls, or decreased in ASD vs. controls) included: 2-aminobutyric acid (two studies, increased); 2-hydroxybutyric acid (two studies, increased); 8-isoprostane, a prostaglandin isomer (three studies, increased); adrenic acid (two studies, decreased); alanine (two studies, decreased); alpha-1-antitrypsin, an enzyme inhibitor that acts as a protector against enzymes of inflammatory cells (two studies, increased); anandamide, an endocannabinoid (two studies, decreased); arachidic acid (two studies, increased); aspartic acid (two studies, decreased); parabacteroides (two studies, increased); creatine kinase, an enzyme catalyzing the conversion of creatine (two studies, increased); coproporphyrin, a product of heme synthesis (four studies, increased); cysteine (three studies, decreased); glutamine (four studies, decreased); glutathione/oxidized glutathione ratio (three studies, decreased); high-density lipoprotein (two studies, decreased); hippuric acid (two studies, increased); high-sensitivity C-reactive protein, a marker of inflammation (two studies, increased); heat shock protein 70, a molecular chaperone that stabilizes protein substrates against denaturation (two studies, increased); interferon-gamma-inducible protein 16 (two studies, increased); kynurenic acid (two studies, decreased); lactic acid (two studies, increased); lead (three studies, increased); neurotensin, a neurotransmitter/modulator (three studies, increased); para-cresol or 4-methylphenol, a phenol derivative that can be converted in an antioxidant (three studies, increased); peroxiredoxin 1, an antioxidant (two studies, increased); phosphatidylcholine, a phospholipid (two studies, decreased); pregnenolone sulfate (two studies, decreased); secreted amyloid precursor protein alpha, a neuroprotective and neurotrophic protein (three studies, increased); succinic acid (three studies, increased); human transforming growth factor beta (three studies, increased); thiol, an organosulfur protecting against oxidative stress (two studies, decreased); and triglycerides (two studies, increased).

Specificity, sensitivity, PPV, NPV and/or ROC AUC were assessed for 303 candidate biomarkers or combinations of biomarkers. When considering biomarkers reported in more than one study, with at least one study showing specificity of 80% or higher, we found 15 biomarkers. Likewise, we located 15 biomarkers reported in more than one study, with at least one study showing sensitivity of 80% or higher. Additionally, 16 biomarkers reported in more than one study had at least one study showing ROC AUC of at least 0.8 (see Table 3). There were no compounds for which PPV or NPV were reported in more than one study.

The only biomarkers showing a specificity of at least 80% in two or more studies, without studies where specificity was less than 80%, with the same direction (i.e., biomarker increased or decreased in all studies) were oxytocin (decreased, two studies) and vitamin E (decreased, two studies). Heat shock protein 70 (increased, two studies), interferon-gamma-inducible protein-16 (increased, two studies), interferon-gamma (increased, two studies), and vitamin E (decreased, two studies) showed a sensitivity of at least 80% in two or more studies, with no studies where sensitivity was less than 80%, with the same direction. Of note, the two studies on specificity and sensitivity in relation to vitamin E derived from non-independent research groups.

In relation to ROC AUC, the following candidate biomarkers showed values of at least 0.8 in two or more studies, without studies where ROC AUC was less than 0.8, with the same direction: heat shock protein 70 (increased, 2 studies), interferon-gamma (increased, two studies), mercury (increased, two studies), and vitamin E (decreased, three studies).

Therefore, similarly to ADHD, none of the biomarkers for which a significant association with ASD was detected and replicated, without negative associations, had evidence of specificity *and* sensitivity of 80% or higher, alongside ROC AUC of 0.8 or higher.

Of note, we also found studies exploring diagnostic classification based on models including a broad array of metabolites or microbiota, and four of these (all from China) provided a ROC AUC of at least 0.8, but none of these models was tested in additional independent studies.

Conduct disorder

We retained only five studies (three cross-sectional and two longitudinal), reported in five papers, three conducted in the US, one in Croatia and one in multiple countries, including a total of 298 participants with conduct disorder and 362 controls.

The average total BIOCROSS score was 6.3 (out of 10). The average scores were 1.0 for item 3; 1.0 for item 4; 1.7 for item 5; 1.7 for item 8; and 1.0 for item 9. So, the BIOCROSS scores were in general higher than those found for ADHD and ASD, even though deriving from a much smaller number of studies.

Overall, 13 unique biomarkers were assessed. Cortisol was the only biomarker tested in more than one study (n=2), and was found significantly associated with conduct disorder in one study but not in the other one. No values of sensitivity and specificity were reported for any biomarker in two or more independent studies.

Global developmental delay/Intellectual disability

We included only five studies (all cross-sectional), reported in six papers, one conducted in China, one in France, one in South Korea, one in Iran, and one in Turkey, encompassing a total of 954 cases of intellectual disability and 189 controls.

Our rating of the quality of the studies was lower compared to the other disorders, but this should be considered cautiously, being based on a limited number of studies. The average total BIOCROSS score was 4.0 (out of 10). The average scores were 0.7 for item 3; 0.7 for item 4; 1.3 for item 5; 1.3 for item 8; and 0.5 for item 9.

Overall, 14 unique biomarkers were assessed. BDNF was the only biomarker tested in more than one study (n=2), and was found significantly associated with intellectual disability in one study but not in the other one. No biomarkers had values of sensitivity and specificity from two or more independent studies.

Table 3 Specificity and sensitivity of at least 80% and receiver operating characteristic area under the curve (ROC AUC) of at least 0.8 in relation to diagnostic biomarkers for autism spectrum disorder investigated in at least two studies, with at least one positive finding

Biomarker	Number of studies with metrics above threshold	Number of studies with metrics below threshold	Direction of the association in studies with metrics above threshold	Frequency of replication (%)
Specificity ≥ 80%				
Oxytocin (serum, plasma)	2	0	Decreased	100
Vitamin E (plasma)	2	0	Decreased	100
Gamma-aminobutyric acid (plasma)	4	0	Three increased, one decreased	75
Brain-derived neurotrophic factor (serum)	2	1	Increased	67
Tumor necrosis factor-alpha (plasma)	3	0	Two decreased, one increased	67
Catalase (blood)	1	1	Increased	50
Glutamate (plasma)	1	1	Increased	50
Homocysteine (serum, plasma)	2	0	One increased, one decreased	50
Heat shock protein 70 (plasma)	1	1	Increased	50
Interferon-gamma (plasma)	1	1	Increased	50
Methionine (plasma)	1	1	Increased	50
Potassium (serum)	1	1	Increased	50
Interleukin-6 (plasma)	3	1	Two decreased, one increased	50
Glutathione S-transferase (plasma)	1	2	Decreased	33
Serotonin (plasma)	2	2	One increased, one decreased	25
Sensitivity ≥ 80%				
Heat shock protein 70 (plasma)	2	0	Increased	100
Interferon-gamma-inducible protein 16 (plasma)	2	0	Increased	100
Interferon-gamma (plasma)	2	0	Increased	100
Vitamin E (plasma)	2	0	Decreased	100
Sodium (plasma)	1	0	Increased	100
Gamma-aminobutyric acid (plasma)	4	0	Three increased, one decreased	75
Catalase (blood)	2	0	One increased, one decreased	50
Glutamate (plasma)	1	1	Increased	50
Potassium (serum)	1	1	Increased	50
Oxytocin (serum)	1	1	Decreased	50
Brain-derived neurotrophic factor (serum)	1	2	Increased	33
Glutathione S-transferase (plasma)	1	2	Decreased	33
Tumor necrosis factor-alpha (plasma)	1	2	Increased	33
Interleukin-6 (plasma)	3	4	Two decreased, one increased	28.5
Serotonin (plasma)	1	3	Decreased	25
ROC AUC ≥ 0.8				
Heat shock protein 70 (plasma)	2	0	Increased	100
Interferon-gamma (plasma)	2	0	Increased	100
Mercury (serum, plasma)	2	0	Increased	100
Vitamin E (plasma)	3	0	Decreased	100
Gamma-aminobutyric acid (plasma)	4	0	Three increased, one decreased	75
Glutathione S-transferase (plasma)	3	0	Two decreased, one increased	67
Interferon-gamma-inducible protein 16 (plasma)	2	1	Increased	67

Table 3 Specificity and sensitivity of at least 80% and receiver operating characteristic area under the curve (ROC AUC) of at least 0.8 in relation to diagnostic biomarkers for autism spectrum disorder investigated in at least two studies, with at least one positive finding *(continued)*

Biomarker	Number of studies with metrics above threshold	Number of studies with metrics below threshold	Direction of the association in studies with metrics above threshold	Frequency of replication (%)
Potassium (serum)	3	0	Two decreased, one increased	67
Tumor necrosis factor-alpha (plasma)	3	0	Two decreased, one increased	67
Brain-derived neurotrophic factor (serum)	1	1	Increased	50
Catalase (blood)	1	1	Increased	50
Glutamate (plasma)	1	1	Increased	50
Interleukin-6 (plasma)	4	2	Two increased, two decreased	33
Melatonin (serum)	1	2	Decreased	33
Oxytocin (serum, plasma)	2	1	Decreased	33
Serotonin (plasma, blood)	2	3	One decreased, one increased	20

Tic disorder/Tourette's syndrome

We found seven eligible studies (all cross-sectional), reported in seven papers; two conducted in China, two in the Netherlands, one in Israel, one in the US, and one in multiple countries; including a total of 569 cases of tic disorder/Tourette's syndrome and 425 controls.

The average total BIOCROSS score was 4.4 (out of 10). The average scores were 0.6 for item 3; 0.9 for item 4; 1.3 for item 5; 1.0 for item 8; and 0.7 for item 9. So, the most concerning aspects, in terms of study quality, were in relation to the lack of reporting of sampling frame, participation rate and power calculation.

Overall, 50 unique biomarkers were assessed. None was tested in more than one study.

Other or combined disorders

We found only one study for coordination developmental disorder. Only three studies included cases with more than one diagnosis, i.e., two studies assessing participants with ADHD plus ASD, reporting on non-overlapping biomarkers across the two studies, and one study including individuals with ADHD and conduct disorder/oppositional defiant disorder.

Genetics

We included five GWAS (see Table 4), covering ADHD, ASD, global developmental delay and autism, tic disorder and Tourette's syndrome, and speech/language impairment. They were conducted in the UK or US or by multinational consortia, encompassing a total of 51,083 participants with neurodevelopmental disorders and 81,918 controls.

Twelve single nucleotide polymorphisms (SNPs) were found to be significantly associated with ADHD, five with ASD, one with tic disorder/Tourette's syndrome, and none with global developmental delay or speech/language impairment. There was no overlap of significant SNPs across disorders (see Table 4).

Despite this limited number of robustly identified genetic biomarkers, several of the studies estimated the total contribution of common genetic risk factors linked to each phenotype (i.e., the "SNP-based heritability" or SNP-h^2). SNP-h^2 was estimated to be approximately 21.6% for ADHD, 11.8% for ASD, 7.7% for global developmental delay, and 21.0% for tic disorder/Tourette's syndrome.

In terms of study quality, according to the selected BIOCROSS criteria, the studies of ADHD, ASD and global developmental delay scored highly (total score: 7 out of 8), while those of tic disorder/Tourette's syndrome and speech/language impairment had moderate scores (6 out of 8, and 5 out of 8, respectively), indicating that the studies were largely well-conducted.

Of note, whereas these GWAS provided an estimate of the degree of association, none of them assessed specificity, sensitivity, PPV, NPV or ROC AUC.

We could not locate any GWAS study focusing on ODD or CD as diagnostic entities. However, there have been several GWAS related to ODD/CD which focused on a broad concept of "externalizing" problems (including, for example, substance use disorder) and consisted of primarily adult samples. The largest relevant GWAS in children[37] operationalized "aggression" and was based on symptoms in the general population, rather than disorder/diagnosis.

Neuroimaging

We included a total of 203 studies (198 cross-sectional and 5 longitudinal), 176 of which conducted in 22 individual countries and 27 in multiple countries, encompassing a total of 28,636 cases and 39,508 controls (see supplementary information).

Retained studies encompassed a variety of brain imaging techniques and measures. At the structural level, magnetic resonance imaging (MRI) morphometric measures – i.e., brain volume, sur-

Table 4 Characteristics of genome-wide association studies (GWAS) of diagnostic biomarkers in neurodevelopmental disorders

Study	Country	Design	Disorder/s	Diagnosis	N probands	N controls	Biomarker(s)	Most adjusted effect size or p value
Demontis et al[32]	Multiple	Cross-sectional	ADHD	Various (DSM/ICD)	20,183	35,191	Global SNP-h^2 rs11420276 rs1222063 rs9677504 rs4858241 rs28411770 rs4916723 rs5886709 rs74760947 rs11591402 rs1427829 rs281324 rs212178	SNP-h^2 = 0.216±0.014 All SNPs: $p<5x10^{-8}$, OR range = 0.835-0.928 and 1.079-1.124
Grove et al[33]	Multiple	Cross-sectional	ASD	Various (DSM/ICD)	18,381	27,969	Global SNP-h^2 rs910805 rs10099100 rs201910565 rs71190156 rs111931861	SNP-h^2 = 0.118±0.010 All SNPs: $p<5x10^{-8}$
Niemi et al[34]	UK and Ireland	Cross-sectional	Global developmental delay and autism	Various	6,987	9,270	Global SNP-h^2 No robust genome-wide significant SNPs	SNP-h^2 = 0.077±0.021
Yu et al[35]	Multiple	Cross-sectional	Tic disorder and Tourette's syndrome	Various	4,819	9,488	Global SNP-h^2 rs2504235	SNP-h^2 = 0.21±0.024 OR=1.16, $p=2.1\times10^{-8}$
Nudel et al[36]	UK	Cross-sectional	Speech/language impairment	Various	278	Not applicable (family based study)	No robust genome-wide significant SNPs	

ADHD – attention-deficit/hyperactivity disorder, ASD – autism spectrum disorder, SNP-h^2 – single nucleotide polymorphism-based heritability, OR – odds ratio

face area, cortical thickness (region-specific and whole-brain) – as well as structural connectivity (via diffusion tensor imaging, DTI) were included. At the functional level, different levels of functional connectivity (including effective connectivity, whole-brain connectivity, network-based connectivity, global/local efficiency, and low frequency fluctuations) were measured with task-based or resting state functional MRI. In addition, a few studies reported less commonly measured functional phenotypes, such as wavelet coherence or entropy, other measures (e.g., brain iron content in ADHD), or used imaging modalities other than MRI, e.g. functional near-infrared spectroscopy (fNIRS) (see also supplementary information).

The average total BIOCROSS score was 4.86 (out of 8). The average scores were 0.98 for item 3; 1.03 for item 4; 1.40 for item 5; and 1.44 for item 8. Therefore, the main concerns were around study population source, reporting of participation rate, and sample size justification.

Four studies included two or more neurodevelopmental disorders compared to controls; the rest focused on individual disorders. Of note, only five studies tested the candidate biomarker in an external, independent sample.

ADHD

We included 66 studies (64 cross-sectional and 2 longitudinal), 61 conducted in 17 countries and five in multiple countries, encompassing a total of 10,273 cases and 20,518 controls.

The average total BIOCROSS score was 5.14 (out of 8). The average scores were 1.00 for item 3; 1.12 for item 4; 1.50 for item 5; and 1.56 for item 8.

More than half of the studies (53%) reported results only as p values, which are poorly informative as significance depends on sample size. Reported effect sizes (d) were lower than 1, and frequently low (around 0.2-0.4). Of note, both specificity and sensitivity were at least 80% for four studies only. These studies were based, respectively, on a semi-supervised learning algorithm that discovers natural groupings of brains based on the spatial patterns of variation in the morphology of the cerebral cortex and other brain regions; fNIRS functional connectivity; a support vector machine (SVM) model including prefrontal cortex activity (fNIRS) during interference with inhibitory control; and cortical thickness and volume features (see supplementary information). However, importantly, there were no other studies replicating these findings. Other

measures such as PPV and NPV were reported only inconsistently.

Autism spectrum disorder

We retained 115 studies (112 cross-sectional and 3 longitudinal), 94 conducted in 14 countries and 21 in multiple countries, including a total of 17,632 cases and 18,254 controls.

The average total BIOCROSS score was 4.72 (out of 8). The average scores were 0.97 for item 3; 0.99 for item 4; 1.36 for item 5; and 1.40 for item 8.

Nearly half of the studies (47%) reported only p values. In seven studies, both specificity and sensitivity were higher than 80%: one assessing wavelet-based coherence in resting state across larger-scale functional networks; four assessing resting-state functional connectivity in different networks; and two evaluating different DTI parameters. In one study only, specificity and sensitivity were higher than 80% and ROC AUC higher than 0.8; that study used a SVM model including ten critical functional resting-state sub-networks (see supplementary information).

Conduct disorder

We found six eligible studies (including 197 cases and 194 controls), all cross-sectional, five conducted in China and one in the UK.

The average total BIOCROSS score was 4.60 (out of 8). The average scores were 1.00 for item 3; 1.00 for item 4; 1.16 for item 5; and 1.50 for item 8.

Three studies reported only p values. Sensitivity and specificity were equal to or higher than 80% in one study only, based on a convolutional neural network (CNN) model to automatically extract multi-layer high dimensional features of structural MRI (see supplementary information).

Tic disorder/Tourette's syndrome

Eight studies (196 cases and 211 controls), all cross-sectional, six conducted in China and two in the US, were retained.

The average total BIOCROSS score was 4.50 (out of 8). The average scores were 1.00 for item 3; 1.00 for item 4; 1.20 for item 5; and 1.50 for item 8.

Four of the studies (50.0%) reported p values only. Both sensitivity and specificity were at least 80% in three of the included studies. The first of these studies focused on inter-hemispheric intrinsic functional connectivity for the bilateral orbitofrontal gyrus, bilateral midbrain, and bilateral ventral striatum; the second on global functional network properties; and the third on multiscale entropy. In all these studies, ROC AUC was higher than 0.8, but no replication of the results was found.

Other disorders

We found only one eligible study on developmental delay, one on dyslexia, and one on dyslexia/learning disorders. In none of these studies, specificity and sensitivity were higher than 80%.

Neurophysiology

A total of 133 studies were retained, 121 cross-sectional, 11 longitudinal, and 1 cross-sectional plus longitudinal, 128 conducted in a total of 24 countries and five in multiple countries, including a total of 7,045 cases and 6,923 controls (see supplementary information).

The average total BIOCROSS score was 4.87 (out of 8). The average scores were 0.97 for item 3; 1.11 for item 4; 1.32 for item 5; and 1.52 for item 8. Therefore, the most critical items were related to sampling frame, participation rate, and sample size justification.

Biomarkers tested in the retained studies included electroencephalogram (EEG), magnetoencephalography (MEG), cardiovascular, acoustic startle reflex, oculomotor, actigraphy and pupillometry measures.

ADHD

N2 amplitude, contingent negative variation (CNV) amplitude, mismatch negativity (MMN) latency, gamma coherence, and activity levels had a replication rate of 100%, albeit in a small number of studies (four for N2 amplitude and two for the other measures) (see Table 5).

The average total BIOCROSS score was 4.88 (out of 8). The average scores were 0.97 for item 3; 1.12 for item 4; 1.33 for item 5; and 1.52 for item 8.

There were no biomarkers for which sensitivity, specificity, PPV, NPV and ROC AUC have been tested in more than one study per biomarker (see supplementary information).

Autism spectrum disorder

The only biomarker with a replication rate of 100% was acoustic eye-blink startle latency (see Table 6). Sensitivity, specificity, PPV, NPV or ROC AUC were not tested in more than one study per biomarker (see supplementary information).

The average total BIOCROSS score was 4.87 (out of 8). The average scores were 0.97 for item 3; 1.11 for item 4; 1.32 for item 5; and 1.51 for item 8.

Other disorders

We could not assess replication rates of biomarkers in other disorders, due to paucity of data.

Neuropsychology

We included 65 studies, 61 cross-sectional, three longitudinal, and one cross-sectional plus longitudinal, 61 conducted in a total

Table 5 Candidate neurophysiological biomarkers investigated in at least two studies, with at least one positive finding, in relation to attention-deficit/hyperactivity disorder (ADHD)

Biomarker	Number of significant effects	Number of non-significant effects	Direction	Rate of replication (%)
MEG/EEG measures				
N2 amplitude	4	0	Four increased	100
Contingent negative variation (CNV) amplitude	2	0	Two increased	100
Mismatch negativity (MMN) latency	2	0	Two increased	100
Gamma coherence	2	0	Two decreased	100
P3 amplitude	6	3	Six decreased	67
Mismatch negativity (MMN) amplitude	2	1	Two increased	67
Alpha clustering coefficient	2	1	Two decreased	67
Alpha path length	2	1	Two decreased	66
Delta power	10	2	Six increased, four decreased	50
Alpha coherence	2	0	One increased, one decreased	50
Theta/beta ratio	5	7	Five increased	42
Alpha power	13	6	Five increased, eight decreased	42
Theta power	5	9	Five increased	36
P3 latency	1	2	One increased	33
Gamma power	2	4	Two decreased	33
Alpha peak frequency	1	2	One decreased	33
Alpha asymmetry	2	4	Two increased	33
Theta coherence	3	0	One increased, two decreased	33
Beta power	9	11	Four increased, five decreased	25
Actigraphy				
Activity level	2	0	Increased	100
Oculomotor measures and visual attention				
Exploration of social information	1	2	One increased	33
Visual attention orienting	3	5	Two increased, one decreased	25
Pupillometry				
Pupil diameter changes	1	1	One decreased	50

MEG – magnetoencephalography, EEG – electroencephalography

of 24 countries and four in multiple countries, including a total of 7,335 cases and 6,341 controls (see supplementary information).

The average total BIOCROSS score was 5.09 (out of 8). The average scores were 1.04 for item 3; 1.19 for item 4; 1.69 for item 5; and 1.16 for item 8.

ADHD

Long-term and short-term memory were characterized by replication rates of 100%, but across a small number of studies (two and five, respectively) (see Table 7).

The average total BIOCROSS score was 5 (out of 8). The average scores were 0.95 for item 3; 1.14 for item 4; 1.67 for item 5; and 1.24 for item 8.

In no instance, sensitivity, specificity, PPV, NPV or ROC AUC have been tested in more than one study per biomarker (see supplementary information).

Autism spectrum disorder

Long-term and short-term memory had replication rates of 100%, but across a small number of studies (two and five, respectively) (see Table 8).

The average total BIOCROSS score was 5.17 (out of 8). The average scores were 1.09 for item 3; 1.21 for item 4; 1.74 for item 5;

Table 6 Candidate neurophysiological biomarkers investigated in at least two studies, with at least one positive finding, in relation to autism spectrum disorder

Biomarker	Number of significant effects	Number of non-significant effects	Direction	Rate of replication (%)
MEG/EEG measures				
P3 amplitude	3	1	Three increased	75
Alpha power	5	3	Five decreased	62.5
N1 amplitude	5	1	Three increased, two decreased	50
N170 amplitude	1	1	One decreased	50
N2 amplitude	2	2	Two increased	50
Mismatch negativity (MMN) amplitude	4	3	Three increased, one decreased	43
Gamma power	22	11	Thirteen increased, nine decreased	39
P1 amplitude	1	2	One increased	33
P2 amplitude	1	2	One decreased	33
Theta power	1	2	One decreased	33
Delta power	1	3	One decreased	25
Beta power	1	10	One decreased	9
Cardiovascular measures				
Heart rate	3	0	One increased, two decreased	67
Heart rate variability - high frequency	3	0	One increased, two decreased	67
Acoustic startle reflex				
Acoustic eye-blink startle latency	3	0	Three increased	100
Acoustic eye-blink startle magnitude	10	5	Ten increased	66
Acoustic eye-blink startle habituation	1	8	One decreased	11
Oculomotor measures and visual attention				
Exploration of visual stimuli	4	0	One increased, three decreased	75
Visual attention - biological motion	4	1	One increased, three decreased	60
Perseveration on visual stimuli	8	4	Six increased, two decreased	50
Visual attention - social	22	33	Eight increased, 19 decreased	34
Visual attention - non-social	11	10	Five increased, six decreased	28
Pupillometry				
Pupil light reflex - dilation	3	1	Two slower	75
Pupil light reflex - constriction	7	3	Six slower, one faster	60
Pupil diameter	4	4	Two increased, two decreased	25

MEG – magnetoencephalography, EEG – electroencephalography

and 1.12 for item 8.

We could not locate any biomarkers for which sensitivity, specificity, PPV, NPV or ROC AUC have been tested in more than one study per biomarker (see supplementary information).

Tourette's syndrome

No replication, for any biomarkers, was found in relation to Tourette's syndrome.

Are there promising biomarkers which are transdiagnostic?

As we did not find any promising biomarker according to the criteria that we set, we could not address our additional aim, i.e., to assess to what extent promising biomarkers are transdiagnostic across neurodevelopmental disorders.

However, replication rates of associations, when available, did not suggest the transdiagnostic nature of any candidate biomarkers, with the possible exception of long-term and short-term mem-

Table 7 Candidate neuropsychological biomarkers investigated in at least two studies, with at least one positive finding, in relation to attention-deficit/hyperactivity disorder (ADHD)

Biomarker	Number of significant effects	Number of non-significant effects	Direction	Rate of replication (%)
Long-term memory	2	0	Two decreased	100
Short-term memory	5	0	Five decreased	100
IQ	6	1	Six decreased	86
Other task accuracy measures	13	2	Thirteen decreased	86
Working memory	20	4	Twenty decreased	83
Sustained attention omission errors	8	2	Eight increased	80
Reaction time variability	17	5	Seventeen increased	77
Ex-Gaussian sigma	3	1	Three increased	75
Response inhibition commission errors	8	5	Eight increased	62
Interference accuracy (e.g., Stroop test)	5	3	Five decreased	62
Mean reaction time	11	7	Eleven increased	61
Ex-Gaussian tau	3	2	Three increased	60
Delay aversion	3	2	Three increased	60
Timing task variability	2	2	Two increased	50
Face/emotion recognition accuracy	1	1	One decreased	50
Face/emotion recognition speed	1	1	One decreased	50
Set shifting accuracy	3	5	Three decreased	37.5
Other memory measures	3	7	Three decreased	30
Reaction time frequency measures	4	8	Three increased, one decreased	25
Wisconsin Card Sorting Test accuracy	1	3	One decreased	25

Table 8 Candidate neuropsychological biomarkers investigated in at least two studies, with at least one positive finding, in relation to autism spectrum disorder

Biomarker	Number of significant effects	Number of non-significant effects	Direction	Rate of replication (%)
Long-term memory	2	0	Two decreased	100
Short-term memory	5	0	Five decreased	100
Working memory	4	1	Four decreased	80
Face/emotion recognition accuracy	3	1	Three decreased	75
Reaction time variability	5	2	Five increased	71
Ex-Gaussian tau	2	1	Two increased	67
Motor coordination	2	1	Two decreased	67
Other memory measures	3	2	Three decreased	60
Other task accuracy measures	3	3	Three decreased	50
Reaction time frequency measures	2	4	Two increased	33
Face/emotion recognition speed	2	1	One increased, one decreased	33
Mean reaction time	1	8	One increased	11

ory, that had 100% replication for ADHD and ASD, and of working memory, that had ~80% replication for these disorders. Similarly, there was no overlap across SNPs across neurodevelopmental disorders in the included GWAS.

DISCUSSION

We conducted the first systematic review of studies on candidate diagnostic biomarkers for neurodevelopmental disorders, including 780 studies encompassing biochemical, genetic, neuroimaging, neurophysiological and neuropsychological measures.

In principle, finding valid, reliable and broadly usable biomarkers to detect or confirm the presence of any neurodevelopmental disorder would be highly valuable. Indeed, as these disorders manifest themselves early in development, an accurate and early diagnosis is crucial from a clinical and public health standpoint. However, despite decades of research and hundreds of publications, we could not find any biomarker that could be defined as promising based on evidence from two or more independent studies with specificity and sensitivity of at least 80%. Other important metrics to assess the validity of a biomarker, such as PPV and NPV, were unfrequently reported. We could not find any cost-effectiveness study.

Findings across the different areas included in this systematic review suggest that, while it is unlikely for a single candidate biomarker to become promising in terms of clinical translation, models including multiple biomarkers, converging on the same or related biological pathways, might be more successful. An additional aim of this review was to assess if promising biomarkers are transdiagnostic across neurodevelopmental disorders. We could not find evidence for this across any combination of the included disorders, but this negative finding was likely due to the absence of promising biomarkers in individual disorders in the first place.

While the body of research considered in this systematic review may seem impressive, the majority of included studies have simply focused on associations, reporting mainly p values, which are poorly informative as they are strongly affected by sample size. Whenever effect sizes were reported, these were generally in the low or moderate range, and certainly not in the range of an effect size of d=1.66 that would be needed to lead to a sensitivity and specificity of 80%[8].

Even when statistically significant associations have been reported, the way candidate biomarkers relate to the symptoms and the pathophysiology of a given disorder is unclear. Moreover, a large number of biomarkers have been significantly related with a given disorder, but in opposite directions, with equally plausible explanations, at least theoretically. For instance, a significant decrease of melatonin in ASD has been interpreted as a reflection of the genetically determined disruption of the serotonin-N-acetylserotonin-melatonin pathway[38]; by contrast, increased levels of melatonin have been explained as a consequence of a putative disruption of the blood-brain barrier in ASD[39].

Furthermore, the role of possible confounding effects when interpreting associations is crucial. Indeed, some markers may be influenced by factors such as diet, abnormal weight, stress, activity levels, smoking, or pharmacological treatment[40]. Our quality appraisal via the BIOCROSS tool indicated that controlling for confounding effects was inconsistent across studies. Importantly, the type of factors adjusted for varied substantially across studies.

Longitudinal studies may help in gaining better insight into the possible causal role of candidate biomarkers. However, only a few (n=36, 4.6%) of the included studies used a longitudinal design. This finding is consistent with evidence in relation to candidate biomarkers for other mental health conditions. For instance, a systematic review of studies on peripheral biomarkers for major psychiatric disorders found that only 34% of the included studies used a longitudinal design[12].

Beyond associations, a minority of studies focused on metrics that are crucial in order to assess to which extent a biomarker is promising, mainly including specificity, sensitivity or ROC AUC. Other important metrics, such as PPV or NPV values, were only rarely assessed. Of note, we could not find any biomarker with evidence from two or more studies with acceptable specificity and sensitivity, or evidence of acceptable PPV, NPV and ROC AUC.

Beyond the methodological issues related to small sample size, poor replicability, lack of standardization, and confounding factors, the main issue that seems to hamper the successful discovery of biomarkers is the very nature of the current psychiatric diagnoses, including the diagnosis of neurodevelopmental disorders, which are based on heterogeneous clusters of symptoms rather than underlying neurobiology. While different conceptualizations exist[41-46], clinical characterizations and delineations of psychiatric diagnoses remain problematic. Stratification of patients based on more homogeneous characteristics may move the field forward leading to more valid biomarkers. As Kapur et al[47] noted, the field of breast cancer faced a similar issue until bumps could be classified with histological tools. The Research Domain Criteria framework[48], aimed at establishing underpinning dimensions from the micro (i.e., genetic) to the macro (i.e., self-reported symptoms) levels, thus appears as a remarkable opportunity for stratification of patients with neurodevelopmental disorders and, hence, the discovery of valid diagnostic biomarkers.

Arguably, given the complexity and heterogeneity of neurodevelopmental disorders in terms of pathophysiology, it is highly unlikely that biomarker applications based on a single parameter will be meaningful in clinical practice[49-52]. Indeed, we found that models based on multiple parameters were in general associated with higher specificity, sensitivity and ROC AUC, although there was no replication of such models yet. In this regard, the scientific community focusing on neurodevelopmental disorders should be inspired by initiatives in other fields integrating several modalities in the same study, such as the Canadian Biomarker Integration Network on Depression (CAN-BIND), connecting clinical information with neuroimaging (e.g., brain structure), molecular (e.g., genetic, hormonal) and electrophysiological (e.g., response to transcranial magnetic stimulation) data[53].

However, even once biomarkers with good specificity, sensitivity and other metrics are found, they will need to be first validated in external, independent samples and then, importantly, also assessed in terms of feasibility and cost-effectiveness in daily clinical practice. Strikingly, we found only a limited number of studies with external validation, mainly limited to neuroimaging studies, and, in an additional search, no replication of studies testing the cost-effectiveness of any biomarker for neurodevel-

opmental disorders. Until this path is completed, any suggestion about the clinical relevance of candidate biomarkers would be misleading. Indeed, there have been reports of court cases where neuroimaging findings and genetic polymorphisms have been used to argue that the accused had a mitigating psychiatric disorder[40]. Our findings do not provide any evidence to support a similar approach for neurodevelopmental disorders[40].

While it is highly unlikely that diagnostic biomarkers will replace clinical assessment, they may eventually support clinical decision making. For instance, preliminary evidence from a randomized, parallel, single-blind, controlled trial showed that the diagnosis of ADHD with the support of a computerized test of attention and activity (QbTest), compared to the standard clinical diagnosis, led to an appointment length reduced by 15% (time ratio: 0.85, 95% CI: 0.77-0.93) and an increased clinicians' confidence in their diagnostic decisions (odds ratio: 1.77, 95% CI: 1.09-2.89)[54]. However, since attention is at the core of the clinical symptoms defining the diagnosis, it is debatable to what degree the measurement of attention is a candidate biomarker of ADHD or a standardized symptom assessment.

The possible future clinical implementation of diagnostic biomarkers will also need to consider important ethical aspects. Patients, lay people and some professionals are concerned that biomarkers may increase mental health stigma and discrimination. Indeed, as a reaction to the Human Genome project, fuelled by historical concerns about eugenics, national legislation has been developed in some countries to prevent genomic discrimination[55]. We argue that educational campaigns will be crucial to address issues around stigma while supporting the discovery of biomarkers.

The lack of evidence for a transdiagnostic nature of the biomarkers that have been explored in neurodevelopmental disorders so far is at odds with the conclusions of another systematic review[12], supporting a transdiagnostic nature of peripheral biomarkers across several mental health conditions (major depressive disorder, bipolar disorder and schizophrenia), as well as evidence from neurophysiological studies in children and adolescents[13]. However, the conclusions of that systematic review were based on the type of key words retrieved from relevant papers as well as on the variation (increase or decrease) of the biomarkers across disorders. By contrast, we focused on replication patterns, in line with the Report of the APA Work Group on Neuroimaging Markers of Psychiatric Disorders recommendations[9].

Moreover, the lack of evidence of transdiagnosticity from GWAS should be considered with caution, given the small sample size for neurodevelopmental disorders (particularly learning disorders) and meta-analytic evidence indicating large genetic correlations between most neurodevelopmental disorders[56]. Indeed, cross-disorder genetic correlation estimates clearly show that there are substantial shared common genetic risks (e.g., across ADHD and ASD) and therefore future studies of specific SNPs that are implicated in multiple disorders will need to be identified through multi-disorder analyses[32]. Similarly, previous large scale studies and meta-analyses of neuroimaging, neurophysiological and neuropsychological impairments have highlighted areas of overlap, particularly between ADHD and ASD[57-60].

It is worth noting that the vast majority of studies have focused on cases of one neurodevelopmental disorder in comparison to neurotypical or population controls – a design that can determine whether a measure may be a good diagnostic biomarker. Should promising diagnostic biomarkers emerge from this literature, their potential clinical utility may be to aid diagnostic decisions when it is unclear whether a child meets criteria for a given disorder. However, a much more likely scenario in clinical practice is the need for objective tools that can augment the valid differential diagnosis between different neurodevelopmental disorders or to determine whether a child should receive a diagnosis of one or more comorbid neurodevelopmental disorders. Yet, a low number of studies have conducted comparisons across different neurodevelopmental disorders.

Biochemical biomarkers

Biochemical biomarkers contributed the largest pool of studies included in the present systematic review. This fact may not be surprising, as, compared to other modalities (e.g., brain imaging), it is arguably less challenging, from a logistic and financial standpoint, to conduct studies on biochemical biomarkers. However, despite a plethora of studies in the field, replications are rare, and at times coming from the same research group.

In addition to the general issues that we have discussed above, there are issues, but also opportunities, that are specific to biochemical biomarkers. Biochemical substances analyzed in the studies retained in the present review were generally collected from blood, plasma, serum or urine samples. Collection from cerebrospinal fluid (CSF) is considered to be of particular interest, due to its proximity to the brain. However, this collection is very complex, due to the invasive procedure. Furthermore, CSF contains far less proteins than plasma, contributing to a reduction of chances to identify proteomic biomarkers[2].

An alternative approach would be the use of post-mortem brain tissues, which would boost the translational links between animal models of neurodevelopmental disorders and studies in living humans, although it should be considered that such studies are not informative on brain activity[61]. Overall, the use of post-mortem tissues for neurodevelopmental disorders is still in its infancy, and mainly limited to ASD. A recent systematic review[62] focusing on ASD and related disorders identified only three post-mortem studies assessing proteins and metabolites, without replicated findings[62]. Efforts in this field, such as the post-mortem brain tissue Autism BrainNet collection from the Simons Foundation[63], are therefore laudable and mirror a trend for other psychiatric disorders, such as the setting-up of the Douglas-Bell Canada Brain Bank[64], or the Netherlands Brain Bank for Psychiatry[65].

Another aspect relates to the type of biochemical biomarker. While a broad range of substances have been investigated, some in the field argue that metabolites ("metabolomics") should be particularly promising as, differently from genomics, they capture the dynamic nature of a disease and, in contrast to proteins ("proteomics"), they provide information on the final product of

complex interactions between proteins, signalling cascades and cellular environments[2]. However, there is usually a high degree of heterogeneity in terms of metabolite panels across studies.

Finally, the procedure to collect data is also highly relevant. Factors including time of day or length of time since last meal are known to impact the levels of certain biomarkers (e.g., cytokines, gene expression, or cortisol)[61]. Therefore, future studies should endeavour to follow standardized procedures, both within and across studies.

Genetic biomarkers

Compared to GWAS of other psychiatric disorders in adults (e.g., major depressive disorder with more than 135,000 cases[66], or schizophrenia with more than 76,000 cases[67]), the five retained GWAS of child neurodevelopmental disorders are relatively small and underpowered to detect robustly associated common genetic risk factors related to these disorders. However, the results of the available GWAS suggest that these disorders are highly polygenic, with thousands of common genetic variants that collectively contribute to an increased disorder risk.

It should be noted that GWAS of child disorders often include adults as well, and further work is needed to understand the degree to which the same genetic risk factors are implicated in childhood/remitting vs. persistent forms of disorder. This type of research has already been undertaken for some neurodevelopmental disorders, for instance ADHD[68].

Furthermore, for many child neurodevelopmental phenotypes, the largest available genetic analyses have focused on continuously distributed symptoms/traits in general population cohorts of children (e.g., the Avon Longitudinal Study of Parents and Children[69]), which only include a small number of diagnosed "cases". These studies were not included in this review, due to being beyond its scope, but it is plausible that biological insights which are gained from GWAS of traits/symptoms may also be relevant to diagnosed disorders, due to a large degree of shared genetic risks across disorders and traits for many neurodevelopmental conditions[70]. It should be also considered that, in addition to GWAS, studies have begun to uncover rare genetic variants, such as copy number variants or protein truncating mutations, especially in ASD[71,72], which should be assessed as possible diagnostic biomarkers.

Overall, although genetic discovery still has a long way to go to be potentially informative for neurodevelopmental disorders in children, existing GWAS can already be applied via polygenic risk score methods to gain insights into phenotypic heterogeneity, and thus inform research on diagnostic biomarkers.

Neuroimaging biomarkers

From a methodological standpoint, we highlight three important aspects that have hampered biomarker discovery and that are particularly applicable to the neuroimaging field. First, it has been noted that this field has mainly been in a mechanistic discovery phase, whereby the main focus has been on detecting alterations in brain imaging measures rather than on searching promising biomarkers[10]. Some in the field have suggested that although, ideally, biomarkers would be based on neurobiologically and mechanistically interpretable findings, this might not always be necessary, as long as biomarkers are rigorously validated. In a parallel with drug development, serendipitously discovered medications with proven clinical effectiveness were incorporated into clinical practice before their biological mechanisms were fully elucidated[10].

Second, brain development is significantly affecting case-control comparisons, and differences in developmental stage could account for greater heterogeneity during childhood and adolescence. Even if biomarkers are found, the lack of reference models of brain development renders the interpretation of certain patterns as a maturational delay or acceleration in neurodevelopmental disorders very difficult. In this context, machine learning approaches have just recently embraced advances that allow the characterization of normative trajectories and parsing of the heterogeneity at the individual level[73]. Notably, these individual-level statistics have revealed a higher predictive power of functionality when compared to unmodelled raw data[74]. Likewise, in line with the complexity of processes and mechanisms underpinning most psychiatric disorders, advanced modelling techniques[75] allow for the integration of multimodal, multivariate imaging features in neurodevelopmental disorders, which hopefully will advance biomarker discovery.

Third, neuroimaging studies included in this review, and in general across neuroimaging literature, provided effect sizes as Cohen's d. However, this metric may not be interpretable if derived out of non-normal distributions, as is often encountered in neuroimaging[8].

In terms of translation/implementation in clinical practice, it is often reported that neuroimaging biomarkers present the disadvantage of higher costs in relation to other modalities (e.g., EEG). However, it should be noted that costs may decrease over time, and the focus should be on cost-effectiveness, rather than cost *per se*. It would be worthwhile to assess to what extent neuroimaging biomarkers could avoid additional expenses, related to delayed or wrong diagnosis, to the health care system.

Neurophysiological and neuropsychological biomarkers

Several neurophysiological and neuropsychological measures have only been investigated in a small number of studies, and mainly in children with ADHD or ASD. Findings for these modalities are highly mixed and suggest very few promising biomarkers. With the exception of markers of memory performance (decreased in both ADHD and ASD), highest replication rates were generally evident for measures that have been investigated to a lesser extent.

Findings appeared more consistent for neuropsychological than for neurophysiological biomarkers. This is likely because the

ceiling/floor effects of neuropsychological measures mean that impaired profiles for a given measure are more likely to emerge consistently in the same direction (e.g., decreased working memory accuracy in children with ADHD)[76]. In contrast, atypical profiles may represent either increases or decreases relative to neurotypical controls for most neurophysiological measures (e.g., increased or decreased EEG connectivity or power).

Of note, previous studies indicate that neurophysiological profiles are highly heterogeneous in children with neurodevelopmental disorders, particularly with ADHD[77] and ASD[78], meaning that the lack of replication on these measures may not be solely attributable to methodological limitations of original studies (e.g., unrepresentative and underpowered samples). This is demonstrated by studies identifying data-driven subgroups of patients characterized by different EEG profiles, which appear associated with various clinical characteristics[79] and different rates of treatment response[80,81].

Another important consideration to make for this type of measures is that, similar to the neuroimaging literature, most of the research on neurophysiological and neuropsychological markers has focused on identifying possible mechanisms implicated in neurodevelopmental disorders (mechanistic discovery phase), rather than on developing biomarkers. Our search explicitly focused on potential biomarkers (or similar terms), and thus did not retrieve studies that investigated relevant measures, but without identifying them with these terms. The limited focus on biomarker development from this literature is also reflected in the very limited number of studies reporting diagnostic metrics (e.g., ROC AUC, sensitivity, specificity) required for establishing whether potential case-control differences at the group level can point to viable biomarkers. Future studies combining data-driven subgrouping techniques to parse heterogeneity with formal tests of biomarker properties may be particularly promising for identifying candidate biomarkers from neurophysiological and neuropsychological assessments.

Limitations

The findings of this systematic review should be considered in the light of some limitations. First, we used the term "biomarker" or equivalent terms (marker, diagnostic test, endophenotype) to retrieve studies in which the authors themselves had labeled their measure(s) as a "(bio)marker", but we could not search for all possible (bio)markers individually, which would have not been feasible. Other systematic reviews[e.g.,12] on biomarkers have used the same strategy. This limitation is particularly relevant for neuroimaging, neurophysiological and neuropsychological studies, of which only a portion used the term "biomarker" or equivalents in the article.

A meta-analytic synthesis was beyond the scope of this review. However, given the generally limited number of studies for each specific biomarker, it would have not been possible to explore sources of heterogeneity in relation to meta-analytic estimates. Therefore, our approach in terms of a narrative presentation of the data is preferable and appropriate for the current stage of the field. Moreover, we could not locate any specific tool for the quality appraisal of longitudinal studies. Rather than adapting the current BIOCROSS for cross-sectional studies, we took a more conservative and cautious approach and we did not rate the quality of longitudinal studies; however, they were only 4.6% of the total number of studies.

Even though we were careful in determining the number of positive and negative replications for each biomarker, it is possible that some studies selectively reported only positive findings, thus biasing our estimates. Furthermore, while we endeavoured to count participants from the same sample only once, the total numbers of participants reported in this systematic review are approximate, because some research groups reported results with partially overlapping samples. Finally, we focused on child-related biomarkers, but we did not include environmental biomarkers, or maternal biomarkers during pregnancy, which were beyond the scope of this work and would require an additional, specific systematic review.

CONCLUSIONS

The present work is the most comprehensive systematic review of candidate diagnostic biomarkers for neurodevelopmental disorders in children and adolescents, and should guide future research in the field. Results point to the need for well-powered studies, replication, standardization of the procedures, use of multimodal approaches in the same study, focus on metrics that are relevant for the validity of a biomarker – as opposed to assessing and reporting mere associations – and an increased focus on disorders less well investigated, such as tic disorder/Tourette's syndrome, intellectual disability, learning and language disorders, as well as a design comparing two or more neurodevelopmental disorders.

It is hoped that in the future the biomarker research in youth with neurodevelopmental disorders will benefit from larger samples, consistent methods, concerted efforts focusing on replication, building on recent consortia and other promising ongoing efforts[82,83]. This research should follow the lead of biomarker research in adults with severe mental disorders[84,85] and of other areas of medicine[86,87], that can inform appropriate assessment techniques. Future research should focus on machine learning and other advanced data analytic techniques as well as multivariable and multi-level biomarker approaches that may arguably be best suited to match the complexities of mental disorders.

ACKNOWLEDGEMENTS

S. Cortese is supported by the UK National Institute for Health and Care Research, NIHR (grant nos. NIHR203035, RP-PG-0618-20003, NIHR203684 and NIHR130077). G. Michelini is supported by a Klingenstein Third Generation Foundation fellowship; L.C. Farhat by the São Paulo Research Foundation; D. Teixeira Leffa by a NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation; G. Salazar de Pablo by the Alicia Koplowitz Foundation: C. Carlisi and V. Carter Leno by a Wellcome Trust Postdoctoral Fellowship; D. Floris by the European Union's Horizon 2020 research and innovation pro-

gramme (Marie Skłodowska-Curie grant agreement no. 101025785); N. Holz by the German Research Foundation and the Radboud Excellence Fellowship. S. Cortese and M. Solmi have contributed equally to this work. Supplementary information on the study is available at https://osf.io/wp4je/?view_only=8c349f45a9ac441490981acf946c8d9a.

REFERENCES

1. FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) resource. Bethesda: National Institutes of Health, 2016.
2. García-Gutiérrez MS, Navarrete F, Sala F et al. Biomarkers in psychiatry: concept, definition, types and relevance to the clinical reality. Front Psychiatry 2020;11:432.
3. McGorry P, Keshavan M, Goldstone S et al. Biomarkers and clinical staging in psychiatry. World Psychiatry 2014;13:211-23.
4. Lutz W, Schwartz B. Trans-theoretical clinical models and the implementation of precision mental health care. World Psychiatry 2021;20:380-1.
5. Maj M, van Os J, De Hert M et al. The clinical characterization of the patient with primary psychosis aimed at personalization of management. World Psychiatry 2021;20:4-33.
6. Stein DJ, Craske MG, Rothbaum BO et al. The clinical characterization of the adult patient with an anxiety or related disorder aimed at personalization of management. World Psychiatry 2021;L20:336-56.
7. McIntyre RS, Alda M, Baldessarini RJ et al. The clinical characterization of the adult patient with bipolar disorder aimed at personalization of management. World Psychiatry 2022;21:364-87.
8. Loth E, Ahmad J, Chatham C et al. The meaning of significant mean group differences for biomarker discovery. PLoS Comput Biol 2021;17:e1009477.
9. First M, Botteron K, Carter C et al. Consensus Report of the APA Work Group on Neuroimaging Markers of Psychiatric Disorders. https://www.psychiatry.org.
10. Abi-Dargham A, Horga G. The search for imaging biomarkers in psychiatric disorders. Nat Med 2016;22:1248-55.
11. Kapczinski F, Vieta E, Andreazza AC et al. Allostatic load in bipolar disorder: implications for pathophysiology and treatment. Neurosci Biobehav Rev 2008; 32:675-92.
12. Pinto JV, Moulin TC, Amaral OB. On the transdiagnostic nature of peripheral biomarkers in major psychiatric disorders: a systematic review. Neurosci Biobehav Rev 2017;83:97-108.
13. Bellato A, Norman L, Idrees I et al. A systematic review and meta-analysis of altered electrophysiological markers of performance monitoring in obsessive-compulsive disorder (OCD), Gilles de la Tourette syndrome (GTS), attention-deficit/hyperactivity disorder (ADHD) and autism. Neurosci Biobehav Rev 2021;131:964-87.
14. Fusar-Poli P, Solmi M, Brondino N et al. Transdiagnostic psychiatry: a systematic review. World Psychiatry 2019;18:192-207.
15. Kim JY, Son MJ, Son CY et al. Environmental risk factors and biomarkers for autism spectrum disorder: an umbrella review of the evidence. Lancet Psychiatry 2019;6:590-600.
16. Kim JH, Kim JY, Lee J et al. Environmental risk factors, protective factors, and peripheral biomarkers for ADHD: an umbrella review. Lancet Psychiatry 2020;7:955-70.
17. Ioannidis JP, Trikalinos TA. An exploratory test for an excess of significant findings. Clin Trials 2007;4:245-53.
18. Thapar A, Cooper M, Rutter M. Neurodevelopmental disorders. Lancet Psychiatry 2017;4:339-46.
19. Solmi M, Radua J, Olivola M et al. Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. Mol Psychiatry 2022;27:281-95.
20. World Health Organization. International classification of diseases, 11th revision (ICD-11). https://icd.who.int/browse11.
21. Solmi M, Song M, Yon DK et al. Incidence, prevalence, and global burden of autism spectrum disorder from 1990 to 2019 across 204 countries. Mol Psychiatry 2022; doi: 10.1038/s41380-022-01630-7.
22. Uher R, Zwicker A. Etiology in psychiatry: embracing the reality of poly-gene-environmental causation of mental illness. World Psychiatry 2017;16:121-9.
23. Solmi M, Fornaro M, Ostinelli EG et al. Safety of 80 antidepressants, antipsychotics, anti-attention-deficit/hyperactivity medications and mood stabilizers in children and adolescents with psychiatric disorders: a large scale systematic meta-review of 78 adverse effects. World Psychiatry 2020;19:214-32.
24. Correll CU, Cortese S, Croatto G et al. Efficacy and acceptability of pharmacological, psychosocial, and brain stimulation interventions in children and adolescents with mental disorders: an umbrella review. World Psychiatry 2021; 20:244-75.
25. Godoy PBG, Sumiya FM, Seda L et al. A systematic review of observational, naturalistic, and neurophysiological outcome measures of nonpharmacological interventions for autism. Braz J Psychiatry 2022; doi: 10.47626/1516-4446-2021-2222.
26. Fairchild G, Hawes DJ, Frick PJ et al. Conduct disorder. Nat Rev Dis Primers 2019;5:43.
27. Scassellati C, Bonvicini C, Faraone SV et al. Biomarkers and attention-deficit/hyperactivity disorder: a systematic review and meta-analyses. J Am Acad Child Adolesc Psychiatry 2012;51:1003-19.e20.
28. Page MJ, McKenzie JE, Bossuyt PM et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. PLoS Med 2021;18:e1003583.
29. Kamp-Becker I, Albertowski K, Becker J et al. Diagnostic accuracy of the ADOS and ADOS-2 in clinical practice. Eur Child Adolesc Psychiatry 2018;27: 1193-207.
30. Wirsching J, Graßmann S, Eichelmann F et al. Development and reliability assessment of a new quality appraisal tool for cross-sectional studies using biomarker data (BIOCROSS). BMC Med Res Methodol 2018;18:122.
31. Page MJ, McKenzie JE, Bossuyt PM et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71.
32. Demontis D, Walters RK, Martin J et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat Genet 2019;51:63-75.
33. Grove J, Ripke S, Als TD et al. Identification of common genetic risk variants for autism spectrum disorder. Nat Genet 2019;51:431-44.
34. Niemi MEK, Martin HC, Rice DL et al. Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. Nature 2018;562:268-71.
35. Yu D, Sul JH, Tsetsos F et al. Interrogating the genetic determinants of Tourette's syndrome and other tic disorders through genome-wide association studies. Am J Psychiatry 2019;176:217-27.
36. Nudel R, Simpson NH, Baird G et al. Genome-wide association analyses of child genotype effects and parent-of-origin effects in specific language impairment. Genes Brain Behav 2014;13:418-29.
37. Ip HF, van der Laan CM, Krapohl EML et al. Genetic association study of childhood aggression across raters, instruments, and age. Transl Psychiatry 2021;11:413.
38. Pagan C, Delorme R, Callebert J et al. The serotonin-N-acetylserotonin-melatonin pathway as a biomarker for autism spectrum disorders. Transl Psychiatry 2014;4:e479.
39. El-Ansary A, Hassan WM, Daghestani M et al. Preliminary evaluation of a novel nine-biomarker profile for the prediction of autism spectrum disorder. PLoS One 2020;15:e0227626.
40. Boksa P. A way forward for research on biomarkers for psychiatric disorders. J Psychiatry Neurosci 2013;38:75-7.
41. Watson D, Levin-Aspenson HF, Waszczuk MA et al. Validity and utility of Hierarchical Taxonomy of Psychopathology (HiTOP): III. Emotional dysfunction superspectrum. World Psychiatry 2022;21:26-54.
42. Krueger RF, Hobbs KA, Conway CC et al. Validity and utility of Hierarchical Taxonomy of Psychopathology (HiTOP): II. Externalizing superspectrum. World Psychiatry 2021;20:171-93.
43. Kotov R, Jonas KG, Carpenter WT et al. Validity and utility of Hierarchical Taxonomy of Psychopathology (HiTOP): I. Psychosis superspectrum. World Psychiatry 2020;19:151-72.
44. Waszczuk MA. The utility of hierarchical models of psychopathology in genetics and biomarker research. World Psychiatry 2021;20:65-6.
45. Astle DE, Holmes J, Kievit R et al. Annual Research Review: The transdiagnostic revolution in neurodevelopmental disorders. J Child Psychol Psychiatry 2022;63:397-417.
46. Michelini G, Barch DM, Tian Y et al. Delineating and validating higher-order dimensions of psychopathology in the Adolescent Brain Cognitive Development (ABCD) study. Transl Psychiatry 2019;9:261.
47. Kapur S, Phillips AG, Insel TR. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? Mol Psychiatry 2012;17: 1174-9.
48. Sanislow CA. RDoC at 10: changing the discourse for psychopathology. World Psychiatry 2020;19:311-2.
49. Paulus MP, Thompson WK. The challenges and opportunities of small effects: the new normal in academic psychiatry. JAMA Psychiatry 2019;76:353-4.
50. Loo SK, McGough JJ, McCracken JT et al. Parsing heterogeneity in attention-deficit hyperactivity disorder using EEG-based subgroups. J Child Psychol Psychiatry 2018;59:223-31.
51. Pulini AA, Kerr WT, Loo SK et al. Classification accuracy of neuroimaging bio-

markers in attention-deficit/hyperactivity disorder: effects of sample size and circular analysis. Biol Psychiatry Cogn Neurosci Neuroimaging 2019;4:108-20.
52. Michelini G, Cheung CHM, Kitsune V et al. The etiological structure of cognitive-neurophysiological impairments in ADHD in adolescence and young adulthood. J Atten Disord 2021;25:91-104.
53. MacQueen GM, Hassel S, Arnott SR et al. The Canadian Biomarker Integration Network in Depression (CAN-BIND): magnetic resonance imaging protocols. J Psychiatry Neurosci 2019;44:223-36.
54. Hollis C, Hall CL, Guo B et al. The impact of a computerised test of attention and activity (QbTest) on diagnostic decision-making in children and young people with suspected attention deficit hyperactivity disorder: single-blind randomised controlled trial. J Child Psychol Psychiatry 2018;59:1298-308.
55. Brannan C, Foulkes AL, Lázaro-Muñoz G. Preventing discrimination based on psychiatric risk biomarkers. Am J Med Genet B Neuropsychiatr Genet 2019;180:159-71.
56. Gidziela A, Ahmadzadeh Y, Micheline G et al. Genetic influences on neurodevelopmental disorders and their overlap with co-occurring conditions in childhood and adolescence: a meta-analysis. medRxiv 2022; doi: 10.1101/2022.02.17.22271089.
57. Karcher NR, Michelini G, Kotov R et al. Associations between resting-state functional connectivity and a hierarchical dimensional structure of psychopathology in middle childhood. Biol Psychiatry Cogn Neurosci Neuroimaging 2021;6:508-17.
58. Lukito S, Norman L, Carlisi C et al. Comparative meta-analyses of brain structural and functional abnormalities during cognitive control in attention-deficit/hyperactivity disorder and autism spectrum disorder. Psychol Med 2020;50:894-919.
59. Kushki A, Cardy RE, Panahandeh S et al. Cross-diagnosis structural correlates of autistic-like social communication differences. Cereb Cortex 2021;31:5067-76.
60. Tung YH, Lin HY, Chen CL et al. Whole brain white matter tract deviation and idiosyncrasy from normative development in autism and ADHD and unaffected siblings link with dimensions of psychopathology and cognition. Am J Psychiatry 2021;178:730-43.
61. Kirkpatrick RH, Munoz DP, Khalid-Khan S et al. Methodological and clinical challenges associated with biomarkers for psychiatric disease: a scoping review. J Psychiatr Res 2021;143:572-9.
62. Fetit R, Hillary RF, Price DJ et al. The neuropathology of autism: a systematic review of post-mortem studies of autism and related disorders. Neurosci Biobehav Rev 2021;129:35-62.
63. Anderson MP, Quinton R, Kelly K et al. Autism BrainNet: a collaboration between medical examiners, pathologists, researchers, and families to advance the understanding and treatment of autism spectrum disorder. Arch Pathol Lab Med 2021;145:494-501.
64. Almeida D, Turecki G. A slice of the suicidal brain: what have postmortem molecular studies taught us? Curr Psychiatry Rep 2016;18:98.
65. Rademaker MC, de Lange GM, Palmen S. The Netherlands Brain Bank for Psychiatry. Handb Clin Neurol 2018;150:3-16.
66. Wray NR, Ripke S, Mattheisen M et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat Genet 2018;50:668-81.
67. Trubetskoy V, Pardiñas AF, Qi T et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature 2022;604:502-8.
68. Rovira P, Demontis D, Sánchez-Mora C et al. Shared genetic background between children and adults with attention deficit/hyperactivity disorder. Neuropsychopharmacology 2020;45:1617-26.
69. Ruisch IH, Dietrich A, Glennon JC et al. Interplay between genome-wide implicated genetic variants and environmental factors related to childhood antisocial behavior in the UK ALSPAC cohort. Eur Arch Psychiatry Clin Neurosci 2019;269:741-52.
70. Taylor MJ, Martin J, Lu Y et al. Association of genetic risk factors for psychiatric disorders and traits of these disorders in a Swedish population twin sample. JAMA Psychiatry 2019;76:280-9.
71. Sanders SJ, He X, Willsey AJ et al. Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. Neuron 2015;87:1215-33.
72. Satterstrom FK, Kosmicki JA, Wang J et al. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. Cell 2020;180:568-84.e23.
73. Marquand AF, Kia SM, Zabihi M et al. Conceptualizing mental disorders as deviations from normative functioning. Mol Psychiatry 2019;24:1415-24.
74. Holz NE, Floris DL, Llera A et al. Age-related brain deviations and aggression. Psychol Med 2022; doi: 10.1017/S003329172200068X.
75. Groves AR, Beckmann CF, Smith SM et al. Linked independent component analysis for multimodal data fusion. Neuroimage 2011;54:2198-217.
76. Kuntsi J, Rogers H, Swinard G et al. Reaction time, inhibition, working memory and 'delay aversion' performance: genetic influences and their interpretation. Psychol Med 2006;36:1613-24.
77. Karalunas SL, Nigg JT. Heterogeneity and subtyping in attention-deficit/hyperactivity disorder – considerations for emerging research using person-centered computational approaches. Biol Psychiatry 2020;88:103-10.
78. Hobson H, Petty S. Moving forwards not backwards: heterogeneity in autism spectrum disorders. Mol Psychiatry 2021;26:7100-1.
79. Dwyer P, Wang X, De Meo-Monteil R et al. Defining clusters of young autistic and typically developing children based on loudness-dependent auditory electrophysiological responses. Mol Autism 2020;11:48.
80. Voetterl H, van Wingen G, Michelini G et al. Brainmarker-I differentially predicts remission to various attention-deficit/hyperactivity disorder treatments: a discovery, transfer, and blinded validation study. Biol Psychiatry Cogn Neurosci Neuroimaging 2022; doi: 10.1016/j.bpsc.2022.02.007.
81. Juarez-Martinez EL, Sprengers JJ, Cristian G et al. Prediction of behavioral improvement through resting-state electroencephalography and clinical severity in a randomized controlled trial testing bumetanide in autism spectrum disorder. Biol Psychiatry Cogn Neurosci Neuroimaging 2021; doi: 10.1016/j.bpsc.2021.08.009.
82. Shic F, Naples AJ, Barney EC et al. The autism biomarkers consortium for clinical trials: evaluation of a battery of candidate eye-tracking biomarkers for use in autism clinical trials. Mol Autism 2022;13:15.
83. Charman T, Loth E, Tillmann J et al. The EU-AIMS Longitudinal European Autism Project (LEAP): clinical characterisation. Mol Autism 2017;8:27.
84. Onitsuka T, Hirano Y, Nemoto K et al. Trends in big data analyses by multicenter collaborative translational research in psychiatry. Psychiatry Clin Neurosci 2022;76:1-14.
85. Kato H, Kimura H, Kushima I et al. The genetic architecture of schizophrenia: review of large-scale genetic studies. J Hum Genet 2022; doi: 10.1038/s10038-022-01059-4.
86. Veyssière H, Bidet Y, Penault-Llorca F et al. Circulating proteins as predictive and prognostic biomarkers in breast cancer. Clin Proteomics 2022;19:25.
87. Hu X, Li G, Wu S. Advances in diagnosis and therapy for bladder cancer. Cancers 2022;14:3181.

DOI:10.1002/wps.21037

Protective and compensatory childhood experiences and their impact on adult mental health

Adult mental health is influenced by childhood exposure to both adverse and protective experiences.

The landmark Adverse Childhood Experiences (ACE) study[1] supported an association between ten categories of adversity, experienced from birth to 18 years, and subsequent physical and mental health problems. These ten types of adversity (physical, emotional and sexual abuse; physical and emotional neglect; parental divorce; domestic violence; living with household members with alcohol or substance abuse, mental illness, or criminality) were found to be common, with more than two-thirds of individuals reporting at least one. Moreover, ACEs had a cumulative or dose-response effect on multiple measures of health and well-being.

Research conducted across the globe and in many populations has consistently found that exposure to ACEs between birth and 18 years alters neurobiological adaptation to stress, increasing the likelihood of difficulties in emotion regulation, impulse control, attention, and social attachments, all of which contribute to mental health problems[2]. A cumulative ACE score of 4 or more increases the likelihood (using adjusted odds ratios) of panic reactions by 2.5 times, depression by 3.6 times, anxiety by 2.4 times, and hallucinations by 2.7 times[3].

While trauma and adversity are well-established risk factors for mental illness, the protective factors that promote resilience are less well known. Research on resilience was initially focused on identifying the qualities of children who succeeded as adults in spite of childhood poverty, abuse or neglect[4]. More recently, researchers have begun characterizing the resilience-promoting qualities of children's environments, identifying the types of supportive relationships and resources that mitigate the effects of ACEs.

Just as ACEs appear to have cumulative negative effects, protective experiences also appear to have a cumulative effect on adult functioning, lessening negative impacts. For example, in a large sample in the American Midwest, positive childhood experiences predicted less depression and better mental health among adults even after accounting for exposure to ACEs[5]. Much of the research on positive experiences has been limited to the presence of supportive relationships, emphasizing the importance of children feeling supported and safe as a counterbalance to the feelings of stress associated with ACEs[6].

Numerous studies indicate that positive experiences during childhood set the foundation for adult mental health. We have identified ten specific protective and compensatory experiences (PACEs) that promote positive outcomes in the face of adversity, as well as overall healthy development[2,7]. Like ACEs, we assess PACEs as experiences that occur prior to age 18.

PACEs are categorized into two domains: supportive relationships and enriching resources. Supportive relationships include unconditional love from a caregiver; having a best friend; volunteering in the community; being part of a group; and having a mentor. Positive parenting, social support, and belongingness have been found to facilitate the development of children's empathy, self-regulation and social skills. Our second domain, enriching resources, include living in a safe home where needs are met; getting a quality education; having a hobby; being physically active; and having rules and routines. Both animal and human studies point to the importance of enriched environments for learning, managing stress, and avoiding risky behaviors.[8]

Research on PACEs specifically indicates that adults who report more PACEs typically report fewer ACEs, suggesting that protective relationships and resources are less available among children who experience family dysfunction and maltreatment. More PACEs are related to less depression, anxiety, substance use, difficulties in emotion regulation, and life stress. Moreover, PACEs protect adults from depressive symptoms, such that greater PACEs weaken the link between ACEs and depression, acting as a protective factor in adulthood[2].

There is also evidence that PACEs can affect parenting attitudes and behaviors. For example, PACEs have been found to act as a buffer between negative parenting attitudes and adverse childhood experiences[2,8]. Similarly, PACEs have been associated with greater resilience and less stress during pregnancy (e.g., future worries about parenting[9]). Taken together, these findings suggest that PACEs buffer the deleterious effects of ACEs on adult functioning and mental health.

We have identified specific PACEs for different age groups (infants and toddlers, teens and young adults, school-aged children[2]). However, the foundation for PACEs remains the same – relationship and resources – and the basic idea of each PACE is similar. For example, having a best friend in early childhood is having opportunities to play with a child or a sibling of a similar age.

PACEs can be used as a tool for adults to help children handle stress, and this may be particularly important during times of chronic and extreme stress, such as the COVID-19 pandemic. On the other hand, parents' stress and mental health are largely influenced by their children's well-being and mental health, and strategies that promote optimal parenting can have a major impact on parents' own functioning[2].

The PACEs Heart model corresponds to the ACEs pyramid model, which posits that ACEs lead to disrupted neurological development; social, neurological and cognitive impairment; adoption of health-risk behaviors; disease, disability and social problems; and early death[1]. The PACEs Heart model posits that supportive relationships and resources lead to optimal neurological development; social, emotional and cognitive functioning; healthy behaviors; achievement of developmental milestones; and health and longevity[8]. These models integrate developmental science, clinical psychology, and mental and physical health research, by detailing possible life course trajectories that stem from childhood experiences.

Fairy tales, folklore and myths from around the world are re-

plete with examples of the youthful hero or heroine's journey from adversity and despair to triumph and success, supporting the empirical evidence that the path to resilience is paved with protective relationships and resources. What is lacking from many trauma-focused interventions is an acknowledgement that PACEs are powerful elements of everyday life that already exist, or can be engineered to occur routinely and frequently, and can be leveraged to support treatment goals and activities.

Our research indicates that adults can benefit from current PACEs as well as previous experiences from childhood. We have created an Adult PACEs Plan that encourages adults to choose one or two PACEs to work on each month with a group of adults. As with PACEs for children of different ages, adult PACEs focus on relationships and enriching experiences[2]. Anecdotally, we have found that individuals benefit from focusing on simple activities that strengthen relationships and impose structure and routine.

In summary, PACEs are often overlooked but powerful tools, that can support therapeutic interventions and mental health throughout the life course.

Amanda Sheffield Morris[1], Jennifer Hays-Grudo[2]
[1]Department of Psychology, Oklahoma State University, Stillwater, OK, USA; [2]Department of Psychiatry and Behavioral Sciences, Oklahoma State University Center for Health Sciences, Tulsa, OK, USA

1. Felitti VJ, Anda RF, Nordenberg D et al. Am J Prev Med 1998;14:245-58.
2. Hays-Grudo J, Morris AS. Adverse and protective childhood experiences: a developmental perspective. Washington: American Psychological Association, 2020.
3. Anda RF, Felitti VJ, Bremner JD et al. Eur Arch Psychiatry Clin Neurosci 2006; 256:174-86.
4. Masten AS. Am Psychol 2001;56:227-38.
5. Bethell C, Jones J, Gombojav N et al. JAMA Pediatr 2019;173:e193007.
6. Narayan AJ, Lieberman AF, Masten AS. Clin Psychol Rev 2021;85:101997.
7. Hays-Grudo J, Morris AS, Beasley L et al. Am Psychol 2021;76:203-15.
8. Morris AS, Hays-Grudo J, Kerr KL et al. Dev Psychopathol 2021;33:533-44.
9. Armans M, Addante S, Ciciolla et al. Advers Resil Sci 2020;1:295-305.

DOI:10.1002/wps.21042

Clearing the air: clarifying the causal role of smoking in mental illness

Decades of observational research have identified a vast range of risk factors which may contribute to the onset of various mental health conditions. A recent review published in this journal[1] brought together data from 380 meta-analyses on this topic, finding over 1,000 different associations for even just non-genetic factors which may influence the risk of mental disorders. Examples of well-established risk factors include adversity/abuse in childhood and stressful employment circumstances in adulthood[1]. Additionally, a more recent body of research has strongly implicated a range of physical health conditions/behaviors – such as diabetes, physical inactivity and obesity – as being associated with an increased risk of mental illness[1,2].

Within this framework, tobacco smoking has emerged as holding particularly strong associations with the onset of mental health conditions. Meta-analyses of longitudinal studies have found strong evidence for a prospective association between smoking and mental disorders, particularly major depression, psychotic disorders and opioid use disorder[1,2]. However, findings from these traditional observational studies may be subject to bias from reverse causation (for example, through unmeasured prodromal symptoms leading individuals to initiate smoking) and residual confounding (for example, through other unmeasured behaviors that influence both smoking and mental health).

Mendelian randomization (MR) is an increasingly applied epidemiological methodology which can address these biases, by using genetic variants known to predispose individuals to certain behaviors/outcomes (e.g., initiating smoking, or smoking heavily), and examining their associations with other outcomes (e.g., mental health diagnoses)[3,4]. In MR, the genetic variants act as instrumental variables, inherited at random and fixed at conception, thus reducing bias from confounding and reverse causation[3]. A number of MR studies on smoking and mental health have already been conducted to examine causal relations, and a recent systematic review of this literature identified high-quality evidence for an effect of smoking on depression, schizophrenia and bipolar disorder[4].

However, there are several limitations of these studies that must be considered[3]. First, although MR studies suggest that smoking behaviors are causal for some mental health outcomes, there is a high degree of bidirectionality, with strong evidence for reverse effects also apparent for depression and schizophrenia[2,4]. This presents the possibility of a vicious cycle, whereby symptoms of mental illness increase smoking and dependence, while smoking increases the risk and severity of mental health conditions. Second, we do not fully understand as yet the biological mechanisms underlying the majority of smoking genetic instruments used in MR analyses. Therefore, the strongest evidence for causal effects of smoking on mental illness will ultimately come from triangulating results across different research methodologies.

The gold standard approach to determine causality would be to conduct a randomized controlled trial (RCT), but it would be unethical to test the effects of tobacco smoking as an experimental exposure directly in this way (due to the addictive potential, and known effects on physical health). Nonetheless, the mental health outcomes of smoking cessation interventions in RCTs can instead be used to infer causal relations. Indeed, a 2021 Cochrane review of 102 studies on this topic consistently showed that people who quit smoking, on average, experienced an improvement in all mental health outcomes examined[5].

Importantly, the observed effects: a) were robust to multiple sensitivity analyses; b) persisted when adjusting for a broad range of socio-demographic, behavioral and clinical covariates; and c) were evident across the 56 RCTs, collectively showing improved mental health outcomes from smoking cessation among participants who had decided to quit smoking *before* being randomized to smoking cessation vs. control interventions (thus eliminating

the potential of reverse causality)[5].

Despite the growing causal evidence, the neurobiological pathways through which smoking adversely affects mental health have yet to be ascertained. One plausible mechanism is related to neuroadaptations in nicotinic pathways in the brain[6] which are associated with psychological withdrawal symptoms, such as depressed mood, agitation and anxiety. Withdrawal symptoms are alleviated by smoking but return when blood levels of nicotine decline at around 20 min after smoking, resulting in repeated changes in a smoker's psychological state throughout the day[6], and perhaps also supporting the "self-medication hypothesis" around smoking and mental health. The fluctuations in mood state experienced by smokers could worsen mental health over time, and the associated biological effects of withdrawal-induced psychological symptoms could increase the risk of developing mental illness[6].

Another potential biological pathway relates to inflammation and oxidative stress, which are both implicated in a range of mental health conditions. A large cohort study in 2021 confirmed that current smoking was associated with increased oxidative stress biomarkers, in a dose-response fashion[7]. Alongside this, the observation that those who had quit smoking for >10 years had similar oxidative stress biomarker levels as never smokers indicates that the biological effects relevant to mental health are reversible[7], which is also consistent with the aforementioned evidence from RCTs showing that cessation improves mental health status in smokers[5].

Continued research into the mechanistic pathways involved in the effects of smoking on mental health will serve to both confirm the nature of indicated causal relations, and increase our understanding of how cessation or other strategies can improve neurological and psychological outcomes in smokers (with or without diagnosed mental illness). Relatedly, the recent adoption of e-cigarettes across society calls for more research on how their use impacts mental health.

While studies in psychiatric settings have suggested that e-cigarettes may be a beneficial tool for helping people with mental illness to reduce tobacco use[8], and thus the adverse physical and mental health effects of smoking, other research in the general population has indicated that nicotine consumption in e-cigarette form may still impact adversely on psychological well-being[9]. Further research is needed to establish a clear evidence base and consensus around the use of e-cigarettes with regards to mental health, in the general population as well as in psychiatric settings.

Meanwhile, as the literature around the magnitude and mechanisms of the psychiatric effects of nicotine and tobacco smoking continues to evolve, promoting smoking cessation in populations with or at-risk for mental illness should be considered as an urgent priority anyway. In recent decades, public health initiatives in many Western societies have successfully reduced tobacco smoking across the general population. However, these initiatives have failed to reach some of most vulnerable members of society, resulting in disparities in tobacco smoking among mental health populations becoming even more apparent than ever. People with mental illness now smoke >40% of all cigarettes sold, and account for around half of all smoking-related deaths across the population, making this single health behavior a key driver of the premature mortality observed in people with severe mental illness[8].

In summary, there is an increasingly strong triangulation of evidence from various study designs and populations that smoking adversely impacts on mental health, in terms of both enhancing the risk of mental illness, and increasing psychiatric symptoms in those with and without diagnosed conditions. While the research priorities lie with elucidating the causal mechanisms for the effects, the clinical priorities pertain more immediately to establishing and disseminating effective smoking cessation interventions within mental health care, in order to protect both the physical and mental health of smokers treated for mental illness.

Joseph Firth[1,2], Robyn E. Wootton[3,4], Chelsea Sawyer[1], Gemma M. Taylor[5]

[1]Division of Psychology and Mental Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK; [2]Greater Manchester Mental Health NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK; [3]Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, UK; [4]Nic Waals Institute, Lovisenberg Diaconal Hospital, Oslo, Norway; [5]Addiction and Mental Health Group, Department of Psychology, University of Bath, Bath, UK

1. Arango C, Dragioti E, Solmi M et al. World Psychiatry 2021;20:417-36.
2. Firth J, Solmi M, Wootton RE et al. World Psychiatry 2020;19:360-80.
3. Wootton RE, Jones HJ, Sallis HM. Mol Psychiatry 2022;27:53-7.
4. Treur JL, Munafò MR, Logtenberg E et al. Psychol Med 2021;51:1593-624.
5. Taylor GM, Lindson N, Farley A et al. Cochrane Database Syst Rev 2021;3: CD013522.
6. Benowitz NL, Hukkanen J, Jacob P. Handb Exp Pharmacol 2009;192:29-60.
7. Salem AA, Trares K, Kohl M et al. Environ Res 2022;204:111923.
8. Firth J, Siddiqi N, Koyanagi AI et al. Lancet Psychiatry 2019;6:675-712.
9. Lechner WV, Janssen T, Kahler CW et al. Prev Med 2017;96:73-8.

DOI:10.1002/wps.21023

A clinically useful model of psychopathology must account for interpersonal dynamics

A useful taxonomy of psychopathology should not only describe variations in mental disorder, but also explain how they occur and point to therapeutic solutions. Contemporary diagnostic models based on a system of polythetic disorder categories do not validly capture the covariation of disorders and symptoms across people, introducing both disorder comorbidity and heterogeneity. As a result, significant advances in explaining discrete categories of psychopathology or deriving disorder-specific therapeutic solutions have not been forthcoming.

These failures have led to new approaches to psychiatric tax-

onomy, such as the Research Domain Criteria (RDoC)[1] and the Hierarchical Taxonomy of Psychopathology (HiTOP)[2]. Both of these systems conceptualize symptoms and disorders as dimensions that can be arranged hierarchically, with narrow symptoms being related to one another because of their mutual associations with broader domains. By reconfiguring mental health variables, HiTOP and RDoC provide evidence-based models of how people differ from one another on average and enable more reliable predictions about what kinds of dysfunctions people are likely to experience.

However, these models are missing two elements that are critical for explaining mental health problems and generating treatments. First, whereas HiTOP and RDoC account for psychopathology solely in terms of elevated levels of certain signs and symptoms within the person, psychopathology manifests as a pattern of dynamic transactions between people and their environments[3]. Persons with psychotic symptoms misperceive information about the world around them; persons with anxious symptoms experience benign situations as threatening; persons with antisocial features experience dangerous situations as exciting, often increasing risks to self and others.

Second, neither RDoC nor HiTOP conceptualize how people move dynamically through their lives. Mental health problems and associated dysfunction are generally not constant. They are more often evoked and manifest in certain situational contexts. The psychotic person misperceives *certain kinds of things*, the anxious person usually worries about *certain kinds of problems*, and the antisocial person seeks *certain kinds of thrills*. HiTOP and RDoC can make predictions about which people are more or less likely to experience mental health problems in the abstract, but not when, where, and how these problems will manifest in the situations people encounter in their lives.

Contemporary integrative interpersonal theory (CIIT) is a model of personality and psychopathology built on 70 years of evidence regarding how people differ from one another (what people are like) and how they function in environmental contexts (what people do)[4]. Like HiTOP and RDoC, it provides a taxonomic model and suite of well-validated tools for assessing individual differences in personality and psychopathology[5]. However, in contrast to HiTOP and RDoC, CIIT is fundamentally concerned with how people function in the context in which they live. The model has two key features that complement new approaches to diagnosis.

First, CIIT is essentially relational. The transition from understanding individuals in a vacuum to understanding people in context has been a stepping stone across scholarly pursuits. Philosophy became intersubjective when the existentialists understood that Descartes had to be thinking about something. The periodic table was derived from the principle that electrons serve the function of connecting elements with one another. Nuclear power was enabled by Einstein's insight about the connection between energy and time. Major models of psychopathology still operate on the assumption that mental disorder can be understood as something that occurs in a vacuum. In contrast, the first assumption of CIIT is that fundamentally important functional expressions of personality and psychopathology occur in interpersonal situations[6].

In CIIT, the interpersonal situation – encompassing direct in-person interactions with objects in the environment, most centrally other people, as well as mental representations of interactions, both recollected and imagined – is considered the basic unit of personality and psychopathology[5]. In the interpersonal situation, self and other interact through four interpenetrating systems that account for the important features of socio-affective function and dysfunction. Each system is represented by two-dimensional circular (circumplex) planes reflecting the major empirically supported dimensions of interpersonal functioning or emotion. The *self system* is structured by the individual's agentic and communal motives. The *affect system* is organized around the person's level of emotional arousal and valence. The *behavior system* includes each person's behavioral dominance and warmth. The *perception system* reflects each person's perceptions of agency and communion in themselves and the other. The interactions among these systems mark key components of dyadic processes that drive an interpersonal situation, as self and other dynamically cycle through continuous transaction.

Second, CIIT is fundamentally dynamic. It is assumed that the satisfaction of motives for agency (power, status) and communion (connection, love) drive interpersonal behavior. This leads to specific, probabilistic predictions about how people will tend to transact with others via affective, behavioral and perceptual processes, and how that can go wrong. Adaptive functioning is not defined by dispositional levels *per se*: rather, it is defined by the ability to stably yet flexibly coordinate and satisfy self and others' motives within the contexts of developmental, socio-cultural and situational demands. Accordingly, dysfunction reflects sustained breakdown in any of the processes that support and maintain the flexible, stable and effective regulation of self, affect and/or interpersonal behavior.

Circumplex measurement tools have been developed to capture variation in the self, affect and behavior system, and multiperspective assessments can be used to capture variations in perception. Such tools include self- and informant-report questionnaires and rating scales, experience sampling via mobile devices, and computer-facilitated continuous observational assessment methods[7]. The dimensions of CIIT and its associated assessment methods can be used to distinguish people from one another, on average, as in HiTOP or RDoC, but they can also be used to describe how people vary from themselves across time and situations. These methods allow for empirical tests of hypotheses about dynamics in group-based research and in individual clinical cases. Parameters from validated dynamic interpersonal assessment measures have been empirically related to dysfunction[8] and psychotherapeutic processes[9].

CIIT moves beyond models that describe how people differ from one another on average, and how those differences pose risk for symptoms, to also account for the context in which those symptoms manifest, and what kinds of environmental transactions can exacerbate or alleviate them. By marrying a structural model of individual differences with a functional model of person-environment transactions, CIIT supports a fuller understand-

ing of personality, psychopathology and intervention, and provides a relational and dynamic complement to individual-based models such as HiTOP and RDoC.

Christopher J. Hopwood[1], Aaron L. Pincus[2], Aidan G.C. Wright[3]
[1]University of Zurich, Zurich, Switzerland; [2]Pennsylvania State University, State College, PA, USA; [3]University of Pittsburgh, Pittsburgh, PA, USA

1. Cuthbert BN. World Psychiatry 2014;13:28-35.
2. Krueger RF, Kotov R, Watson D et al. World Psychiatry 2018;17:282-93.
3. Hopwood CJ, Wright AGC, Bleidorn W. Nat Rev Psychol (in press).
4. Pincus AL. In: Lenzenweger MF, Clarkin JF (eds). Major theories of personality disorder. New York: Guilford, 2005:282-331.
5. Pincus AL, Hopwood CJ, Wright AGC. In: Funder D, Rauthmann JF, Sherman R (eds). Oxford handbook of psychological situations. New York: Oxford University Press, 2020:124-42.
6. Wright AGC, Pincus AL, Hopwood CJ. PsyArXiv 2022; doi:10.31234/osf.io/fknc8.
7. Pincus AL, Sadler P, Woody E et al. In: Hopwood CJ, Bornstein RF (eds). Multimethod clinical assessment. New York: Guilford, 2014:51-91.
8. Wright AGC, Stepp SD, Scott LN et al. J Abnorm Psychol 2017;126:898-910.
9. Altenstein D, Krieger T, Grosse Holtforth M. J Couns Psychol 2013;60:445-52.

DOI:10.1002/wps.21044

Non-specific psychopathology: a once and future concept

A popular strategy for criticizing diagnostic categories in psychiatry is to point out that two people can meet criteria for the same disorder, yet have few or even no symptoms in common. For instance, two people can meet the diagnostic criteria for major depressive disorder and share only one symptom. For post-traumatic stress disorder (PTSD), two people can meet the diagnostic criteria and share no symptoms.

Some critics also enumerate the different ways to meet diagnostic criteria. To illustrate, there are 126 ways to combine the nine DSM depression criteria and meet the cut-off of five to be diagnosed. Considering all combinations, there are 227 ways to meet criteria for depression using the DSM. Does this amount to 227 kinds of depression?

When introducing the concept of operational definitions, Bridgman wrote: "If we have more than one set of operations, we have more than one concept, and strictly there should be a separate name to correspond to each different set of operations"[1, p.10]. Certainly, we should attempt to understand the implications of different operational definitions of the same diagnostic concept, but some philosophers of science believe that Bridgman took it too far.

Let us look at an example from psychological testing. On the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) depression scale, with 57 items scored 0 or 1, a person has to score 26 or greater to cross the depression threshold. Doing the math, there over 12 quadrillion ways for a person to score 26 alone. To claim that we should attempt to name over 12 quadrillion kinds of depression on the MMPI-2 is absurd, and at least potentially makes the claiming that there are 227 kinds of depression seem somewhat silly.

One reason why two people can meet the same diagnostic criteria and share only one symptom is that operationalized diagnostic concepts typically under-represent the symptom picture, i.e., they lack content validity. In part, this is because many nosologists adopt a convention regarding differential diagnosis which holds that, ideally, a diagnostic criterion set should indicate when a disorder is present, distinguish the disorder from non-disorders, and distinguish the disorder from other disorders. In technical language, diagnostic criteria should be both sensitive to the presence of a disorder and specific to that disorder.

When non-specific symptoms are de-emphasized, two people who share only one depression symptom may nevertheless be similar on other common features of the disorder that are not included in the over-specified criterion set. For depression, common but non-specific symptoms include, for instance, anger, anxiety, depersonalization, gastrointestinal distress, headaches, and rumination.

In addition to being under-representative, operational definitions are *open concepts*, meaning that new information and new uses for a concept can impel us revise the concept and extend it in different directions. According to the theory of open concepts, there is an inherent indeterminacy to the phenomena of psychiatry and, thus, psychiatric concepts cannot be closed off once and for all, because there are potentially further facts on the horizon that keep the process of defining and refining alive. This means that non-specific symptoms which have been relegated to the background can be brought into the foreground, and vice versa. The historical transitions from classic hysteria to somatic symptom disorders and PTSD might be considered an example of background-foreground shifts.

The mutable, protean nature of psychiatric disorders is not a new observation. Writing about hysteria in the 17th century, Sydenham noted that its symptoms varied so greatly and were so irregular that it was difficult to describe the disorder with any precision[2]. More recently, psychopathologists have re-recognized the relevance of non-specific psychopathology.

One example is the pluripotential risk syndrome described by McGorry and colleagues[3]. Phenotypically broad and difficult to subtype, it is named a "syndrome" because the symptoms are associated with a decline in functioning. These symptoms include an intensification of normal traits such as worry and anger, and the appearance of novel features such as hypervigilance and compulsivity. The symptoms also ebb and flow in a "heterotypic" fashion. Heterotypic can refer to both the same risk profile having a broad range of outcomes ("multifinality") and a single individual expressing shifting symptom pictures over time ("a divergent trajectory")[4]. Symptoms in the ebb and flow may be transient and remit. Alternatively, they may develop into more specific risk syn-

dromes for broad categories such as mood disorder or psychosis. This may be followed by a prodrome stage and eventually a specified category such as major depressive disorder, but such a linear trajectory is not the norm.

A second example is from factor analytic psychology. The general psychopathology factor "p" represents a common cause of and liability to all forms of psychopathology[5]. Higher scores on "p" are associated with varied and severe symptom pictures. One reason why it has been difficult to validate disorder-specific etiologies may be because many risk factors and causes are themselves associated with the general factor (i.e., are non-specific).

The "p" factor has been incorporated into the project to develop a hierarchical taxonomy of psychopathology and placed at the apex of the hierarchy. Underneath "p" are broad dimensions such as internalizing, externalizing, and thought disorder. Specified categories such as major depressive disorder and panic disorder are nested under the dimensions, but it is not foreordained that digging down to more specific constructs will be the most useful strategy. As an analogy, if someone is having an allergic reaction to pain medication, one might want to know if he/she took a non-steroidal anti-inflammatory drug, but whether it was specifically ibuprofen or aspirin is irrelevant.

Berrios[6] argues that the list of symptoms used to describe psychopathology was prematurely closed in the 19th century and it is unlikely to be extended unless psychiatrists attend less to diagnosing disorders and more to describing symptoms. Maj et al[7] argue that it would be helpful to have measures that assess the whole range of depression symptoms beyond what is contained in diagnostic criteria lists.

A potential barrier to a project of extension is that concepts such as depression have considerable face validity, due in part to their familiarity. This entrenchment may function as an *a priori* constraint if people assign more weight to symptoms that seem to fit with familiar concepts, and background those that do not.

One caveat to a shift toward the study of non-specific symptoms that cut across traditional diagnostic categories is in reference to what 19th century European thinkers called "disease forms". Parnas[8] and Thornton[9] argue that symptoms may seem non-specific because they often refer to decontextualized, abstract concepts such as obsessions and anhedonia. In their view, symptoms can have more specificity within the *gestalts* represented by constructs such as schizophrenia. For instance, obsessions and compulsions can appear transdiagnostic on the surface, but be qualitatively distinct in different diagnostic contexts. To illustrate, on the psychosis spectrum, the content of obsessions and compulsions tends to be more sexual and aggressive and the symptoms have a delusional character in which, unlike for anxiety disorders, the person does not view them as irrational.

An important scientific goal should be to explain why psychiatric problems often begin with an intensification of non-specific symptoms that ebb and flow, in some cases being mutable or protean and in others settling into specified syndromes. The theory of open concepts also suggests that constructs for psychiatric disorders have been and will be "imperfect" not only due to a lack of knowledge, or because they are operationalized, or because they are descriptive, not etiological. They are also imperfect because of the inherent and inevitable limits to conceptualizing complex, noisy phenomena.

Peter Zachar
Department of Psychology, Auburn University Montgomery, Montgomery, AL, USA

The author thanks M. Waugh, D. Stein and M. Maj for helping him better articulate his ideas.

1. Bridgman PW. The logic of modern physics. New York: Macmillan, 1927.
2. Sydenham T. In: Hunter R, Macalpine I (eds). Three hundred years of psychiatry (1535-1860). Oxford: Oxford University Press, 1963:221-4.
3. McGorry PD. Schizophr Res 2010;120:49-53.
4. Nolen-Hoeksema S, Watkins ER. Perspect Psychol Sci 2011;6:589-609.
5. Lahey BB, Moore TM, Kaczkurkin AN et al. World Psychiatry 2021;20:57-63.
6. Berrios GE. In: Hamanaka T, Berrios GE (eds). Two millennia of psychiatry in West and East. Tokyo: Gakuju Shoin, 2003:81-91.
7. Maj M, Stein DJ, Parker G et al. World Psychiatry 2020;19:269-93.
8. Parnas J. Schizophr Bull 2011;37:1121-30.
9. Thornton T. In: Keil G, Keuck L, Hauswald R (eds). Vagueness in psychiatry. Oxford: Oxford University Press, 2017:119-37.

DOI:10.1002/wps.21043

What's in the name "schizophrenia"? A clinical, research and lived experience perspective

"What's in a name? That which we call a rose by any other name would smell just as sweet". In response to the growing international momentum for renaming "schizophrenia", some people have invoked this famous Shakespearean line from *Romeo and Juliet* to suggest that changing a word is irrelevant for efforts addressing the inaccuracies and stigma associated with the term. As persons with research, lived, clinical and/or peer support experience, we respectfully disagree.

What *is* in a name is how it is used. A name should do no harm. However, since its conception over a century ago, the name "schizophrenia" has carried with it discrimination, stigma and misunderstanding. The term was first conceived by E. Bleuler in 1908 and derived from Greek to mean "split-mind", an idea that diverges from modern scientific and colloquial understandings of the experiences it describes, and from treatment advances[1]. As our colleague L. Larson from the Consumer Advisory Board of Massachusetts Mental Health Center stated, "The term schizophrenia hasn't evolved with the treatment"[2].

The term has also been used to oppress. In his book *The Protest Psychosis: How Schizophrenia Became a Black Disease*[3], J. Metzl suggests that the name was distorted to mean "racialized aggression", and was used to diagnose and institutionalize Black men who were incarcerated after participating in US Civil Rights demonstrations. The tensions within society may have transformed "schizophrenia" into an instrument of systemic racism to oppress Black Americans, at least during the 1960s.

Several initiatives around the world have attempted to address the problems associated with the term "schizophrenia". These include name changes in some Asian countries, with evidence of benefits such as decreased prejudice and stigma, more clinicians willing to disclose diagnosis to patients, and an increased number of patients willing to seek care[4]. Within the field, professional organizations, journals and the DSM-5 have revised their terminology to reflect the spectrum nature of the condition. Advocates of a new term also point to the successful name changes for other psychiatric conditions, such as from Multiple Personality Disorder to Dissociative Identity Disorder, and from Manic Depressive Illness to Bipolar Disorder. Furthermore, in a broader societal context, there is increasing attention to the importance of language and our choice of words.

Additionally, several survey studies strongly support renaming "schizophrenia", including two recent ones conducted in Italy[5] and the US[6]. The US survey[6] comprised the largest and most diverse sample, with multiple stakeholder groups including people with lived experience, families, mental health clinicians, researchers, government officials and the general public. This study uniquely partnered with people with lived experience of psychosis in all aspects of the project, thus gaining vital and under-represented expertise and perspectives. The most popular alternate name was Altered Perception Syndrome, followed by Psychosis Spectrum Syndrome and Neuro-Emotional Integration Disorder. Of note, Altered Perception Syndrome was the one alternate term from this survey coined by a person with lived experience of the condition and not used as an alternative name for "schizophrenia" in the literature or in other countries. The popularity of this term underscores how imperative it is to include the ideas and opinions of people living with the condition in all renaming initiatives.

However, far beyond beginning and ending with one word, the efforts to rename "schizophrenia" signal a call to action for the field and are part of a larger movement toward using person-centered, recovery-oriented, and experience-based language to support the well-being and aspirations of people with this and other mental health conditions. Language allows us to connect with others and to understand ourselves. It is not only based on definitions; it is intertwined with the actions we take and is affected by the world around us. The word "schizophrenia" is a particularly poignant example of the influence language bears on people, both in society's views and within identity. In a recent commentary, E. Saks writes of schizophrenia as a lifelong companion and of its name and construct becoming "too sclerotic"[7]. As she notes, "A name change may do more than anything to destabilize society's concepts".

Self- and public stigma, prejudice and discrimination are compounded by labels assigned to symptoms and experiences. Emphasizing advances in treatment and acceptance of experiences while removing the negative connotations of labels such as "schizophrenia" may encourage more people to seek support early and to advocate for their own mental wellness. Indeed, guidance has recently been published for clinicians when sharing psychosis diagnoses with individuals and their families, using the INSPIRES acronym: to use individualized, normalizing and non-stigmatizing, setting-specific, person-centered, informational, reassuring and inspiring, empathetic and empowering language, and then to express strategic next steps[8]. This approach helps "focus on instilling a sense of hope for recovery rather than simply informing individuals with illness of their symptoms and prognosis". Changing the name "schizophrenia" is one of several stepping stones on the path to improving support for the people we serve with language that illustrates the hope in recovery.

We appreciate that a name change is not easy and takes time. We also know that some people have argued that the time is not yet right for a name change; they note that a revised name should not be considered until new scientific findings emerge. But, we would ask, when exactly *is* the right time? It has been over a century since the term "schizophrenia" was coined. When will there be enough research and treatment advances to warrant a name change? We certainly still had (and have) a long way to go in our understanding and treatment of other mental health conditions whose names have already been changed.

A name change is not a panacea for the problems associated with the term "schizophrenia", and it would need to be accompa-

nied by other initiatives such as public education and legislation. As with most complex problems, the solution needs to be strategic, coordinated and multi-pronged. More research is also vital, as consensus on any new name should ideally be derived from a large, diverse sample of all relevant stakeholders and a rigorous scientific consensus. It is particularly critical to continue to include the voices of people who live with the condition, who are often marginalized and suffer inequities, a point cogently and eloquently illustrated by a recent paper in this journal[9] describing the lived experience of psychosis.

Words matter. If a name change can even be part of what leads to improved lives for people with the condition, then isn't it worth it? Why keep a name that the majority of people with the condition are not comfortable with, that they feel is stigmatizing and discriminates against them, and that dissuades them from seeking out care? Isn't that reason enough?

What's in a name? Names shift to reflect transformation, and new names catalyze change. As E. Dickinson wrote, "I know nothing in the world that has as much power as a word".

Raquelle I. Mesholam-Gately[1,2], Dan Johnston[3], Matcheri S. Keshavan[1,2]
[1]Department of Psychiatry, Harvard Medical School, Boston, MA, USA; [2]Massachusetts Mental Health Center, Public Psychiatry Division of the Beth Israel Deaconess Medical Center, Boston, MA, USA; [3]Consumer Advisory Board, Massachusetts Mental Health Center, Public Psychiatry Division, Beth Israel Deaconess Medical Center, Boston, MA, USA

The authors thank all members of the Consumer Advisory Board of Massachusetts Mental Health Center and other contributors to the Renaming Schizophrenia Survey Project for their ideas, inspiration, and review of this letter.

1. Keshavan MS, Nasrallah HA, Tandon R. Schizophr Res 2011;127:3-13.
2. Brown K. "Schizophrenia" still carries a stigma. Will changing the name help? New York Times, December 20, 2021.
3. Metzl JM. The protest psychosis: how schizophrenia became a black disease. Boston: Beacon Press, 2009.
4. Yamaguchi S, Mizuno M, Ojio Y et al. Psychiatry Clin Neurosci 2017;71:347-62.
5. Lasalvia A, Vita A, Bellomo A et al. Schizophr Res 2021;228:502-9.
6. Mesholam-Gately RI, Varca N, Spitzer C et al. Schizophr Res 2021;238:152-60.
7. Saks ER. Schizophr Res 2022;242:121-2.
8. Keshavan MS, Davis B, Friedman-Yakoobian M et al. Schizophr Res 2022;239:92-4.
9. Fusar-Poli P, Estradé A., Stanghellini G et al. World Psychiatry 2022;21:168-88.

DOI:10.1002/wps.21033

Are language features associated with psychosis risk universal? A study in Mandarin-speaking youths at clinical high risk for psychosis

Natural language processing (NLP) analyses have shown decreased coherence (tangentiality, derailment) and complexity (poverty of content) in schizophrenia and in clinical high risk (CHR) states for psychosis. We reported previously in this journal[1] that an NLP machine learning classifier, which included measures of coherence and complexity, predicted psychosis onset in two independent English-speaking CHR samples. Moreover, reduced complexity has been associated with increased pauses and negative symptoms in at-risk youths[2].

Multiple recent NLP studies in schizophrenia and CHR cohorts, using different methods, have largely found this same pattern of disturbance in the structure of language and speech[3]. Most of these studies have been conducted in English, with notable exceptions including Dutch, Portuguese and Spanish[4]. It remains unknown, however, whether NLP findings obtained in English or other Indo-European languages would generalize to less similar languages, such as Mandarin, which has very different grammatical and prosodic conventions.

This study included 20 help-seeking CHR youth and 25 healthy controls who were recruited as part of the US National Institute of Mental Health (NIMH)-funded Shanghai-At-Risk for Psychosis (SHARP) study at the Shanghai Mental Health Center, where institutional review board approval was obtained. Caseness and symptoms were determined using the Structured Interview for Psychosis-Risk Syndromes (SIPS)[5]. Subjects were Han Chinese and spoke Mandarin fluently, and they provided informed consent. Sex distribution was similar between CHR subjects and controls (55% vs. 48% female), but CHR subjects were younger (19.6±6.4 vs. 24.9±1.9 years) and had less education (11.4±4.0 vs. 16.7±1.4 years).

Interviews were approximately 30 min in length, and were based on qualitative methods previously described[6]. They were transcribed verbatim in Mandarin and translated into English using Google Translate, with verification by bilingual researchers. Audio recordings were diarized (segmented by speaker using time stamps from transcription) so that acoustic analyses could be done of subjects' speech.

NLP features analyzed for both English and Mandarin included coherence, complexity, and sentiment (i.e., emotional valence – positive, negative, neutral), as reported previously[1,7]. For English NLP only, sentiment also included anger, fear, sadness, joy and disgust; frequency of wh-words (e.g., "which') was also assessed. For Mandarin NLP only, frequency of measure words, possessives, and localizers (e.g., *gongzuo-shang*, "during work"; or *liangge-renzhijian*, "between two people") was also calculated[8]. Acoustic features analyzed in Mandarin included those characteristic of schizophrenia or CHR states among English-speaking subjects, including abnormal pauses, flat intonation, voice breaks, and

pitch variation[7].

All features were corrected for age and education by applying regression coefficients from healthy controls, and highly correlated features were removed from analysis. Machine learning classification was done using random forest and support vector machines (SVM) for Mandarin NLP, English NLP, and acoustics, with each experiment repeated 20 times, identifying the top five features of each model. Associations between linguistic features (cross-language analysis) and with symptoms (symptom inference) were also tested (see also supplementary information).

Each of the three SVM machine learning classifiers showed high accuracy in discriminating spoken language in CHR subjects from that of healthy controls: English-specific NLP (95%), Mandarin-specific NLP (94%), and acoustic analysis (88%), with similar results for random forest. Top features for the English-specific NLP machine learning were wh-word and noun use (greater in CHR), and coherence, adjective use and adverb use (all less in CHR). Top features for the Mandarin-specific NLP machine learning were localizer use (greater in CHR), and positive sentiment, two metrics of coherence, and adjective use (all decreased in CHR). Of note, features common to the NLP machine learning for both languages were highly correlated, specifically coherence (r=.70) and adjective use (r=.60).

For acoustics, the top features in the machine learning classifier were two pause metrics, and three indices of acoustic quality: chroma #11 (timbre/quality), bandwidth formant #1 (dysphonia/hoarseness), and spectral spread (energy – decibels/frequency). Of note, only acoustic features were significantly associated with symptoms (negative: r=0.69, p=8E–4, positive: r=0.49, p=3E–2) (see also supplementary information).

Several important findings emerge from this proof-of-principle study. First, in Mandarin, spoken language can differentiate CHR subjects from healthy controls with high accuracy, using either linguistic or acoustic features. Second, the application of English-specific NLP to transcripts translated from Mandarin has utility, as there was comparable accuracy for both the English-specific and Mandarin-specific NLP. Further, there was overlap in top features in the two NLP classifiers, specifically decreases in adjective use and coherence, with both of these features highly correlated across the two languages, suggesting that these key metrics survive translation. Nonetheless, the application of Mandarin-specific NLP allowed the identification of a key linguistic feature that would not be captured otherwise – the increased use of localizers among CHR subjects – which may reflect concreteness or increased use of idioms; this is a new finding that merits replication and further investigation. Finally, the acoustic classifier, in addition to having high accuracy, identified features similar to those found in English-speaking CHR and schizophrenia cohorts, including abnormal pause behavior, and indices of voice quality and energy. As in prior studies, acoustic features were associated with symptoms, in particular negative symptoms.

This study is the first to use natural language processing and acoustic analyses to characterize spoken language among native Mandarin speakers in China identified as at clinical high risk for psychosis. Our findings support the idea that there may be universal features of spoken language disturbance across psychosis and its risk states, particularly in respect to reduced coherence, but also word usage and pause behavior that may index reduced complexity. Yet, our study also shows that there are language-specific features characteristic of psychosis risk, suggesting that it is essential to also analyze spoken language using language-specific NLP methods.

This is a small proof-of-principle study with the potential confounds of age and education, and none of the classifiers generated were cross-validated in a second cohort. Therefore, these findings should be investigated and replicated in a larger cohort of Mandarin-speaking CHR subjects and healthy controls who are more similar in demographics.

More broadly, future studies should include a similar heuristic of using both English-based and language-specific NLP approaches, as well as acoustic analyses, to assess spoken language in CHR cohorts (e.g., English, Mandarin, Cantonese, Korean, Spanish, German, Portuguese, Danish, French, Italian) from around the world, as is planned for the Accelerating Medicines Partnership® Program – Schizophrenia, to determine both universal and language-specific features of language disturbance characteristic of clinical risk for psychosis.

Carla Agurto[1], Raquel Norel[1], Bo Wen[1], Yanyan Wei[2], Dan Zhang[2], Zarina Bilgrami[3], Xiaolu Hsi[4], Tianhong Zhang[2], Ofer Pasternak[5], Huijun Li[6], Matcheri Keshavan[4], Larry J. Seidman[4], Susan Whitfield-Gabrieli[7], Martha E. Shenton[5,8], Margaret A. Niznikiewicz[9], Jijun Wang[2], Guillermo Cecchi[1], Cheryl Corcoran[10], William S. Stone[4]

[1]IBM T.J. Watson Research Center, Yorktown Heights, NY, USA; [2]Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China; [3]Department of Psychology, Emory University, Atlanta, GA, USA; [4]Department of Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA; [5]Department of Psychiatry, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; [6]Department of Psychology, Florida A&M University, Tallahassee, FL, USA; [7]Department of Psychology, Northeastern University, Boston, MA, USA; [8]Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; [9]Department of Psychiatry, VA Boston Healthcare System, Brockton, MA, USA; [10]Icahn School of Medicine at Mount Sinai, New York, NY, USA

The authors would like to thank the study participants at the Shanghai Mental Health Center. C. Agurto and R. Norel contributed equally to the work. Supplementary information on this study can be found at https://wenboown.github.io/SHARP-NLP-Letter-to-WP/.

1. Corcoran CM, Carrillo F, Fernandez-Slezak D et al. World Psychiatry 2018; 17:67-75.
2. Stanislawski ER, Bilgrami ZR, Sarac C et al. NPJ Schizophr 2021;7:3.
3. Corcoran CM, Cecchi G. Biol Psychiatry Cogn Neurosci Neuroimaging 2020; 5:770-9.
4. Corcoran CM, Mittal VA, Bearden CE et al. Schizophr Res 2020;226:158-66.
5. McGlashan TH, Walsh BC, Woods SW. Structured Interview for Psychosis-Risk Syndromes. New Haven: PRIME Research Clinic, Yale School of Medicine, 2017.
6. Ben-David S, Birnbaum ML, Eilenberg ME et al. Psychiatr Serv 2014;65:1499-501.
7. Agurto C, Cecchi GA, Norel R et al. Neuropsychopharmacology 2020;45:823-32.
8. Chappell H, Peyraube A. In: Xu D (ed). Space in languages of China: cross-linguistic, synchronic and diachronic perspectives. Dordrecht: Springer, 2008: 15-37.

DOI:10.1002/wps.21045

Association between stressful life events and psychosis relapse: a 2-year prospective study in first-episode psychosis

Stressful life events occurring after the onset of psychosis have been associated with poorer long-term outcomes[1]. However, methodological issues with existing evidence, such as inadequate consideration of the confounding effect of illness stage and of socio-demographic and clinical variables, limit a clear understanding of the implications of this finding. Further, as most available evidence is based on retrospective studies, which are susceptible to recall bias, prospective evidence that life events preceded and were in reasonable temporal proximity to the psychosis relapse is needed to support the notion that these events might have a precipitating role.

About one in two patients will present with a relapse severe enough to require hospital readmission within the first two years of their first episode of psychosis[2]. Relapses not only cause considerable suffering to the individuals and their families, but also have implications for utilization of health care resources. Here, we employed a prospective cohort approach to investigate the effect of stressful life events on the risk of relapse, as indexed by hospital admission, over the first two years following psychosis onset.

First-episode non-organic psychosis patients (ICD-10: F20-F29 and F30-F33) aged 18-65 years, admitted to psychiatric services in the catchment area of South London, were prospectively recruited and followed up for at least two years. Stressful life events that occurred over the follow-up period were assessed using the Brief Life Events Questionnaire (BLEQ), a tool that allows the assessment of the time of occurrence of each event and its emotional impact, with high validity and reliability[3]. A full treatment history was recorded by the World Health Organization (WHO) Life Chart Schedule[4]. Relapse was defined as admission to a psychiatric inpatient unit because of exacerbation of psychotic symptoms within two years of first presentation to psychiatric services and receiving a diagnosis of psychosis.

Separate survival analyses were carried out to investigate the effect of any life events and of total number of life events (occurring within the two-year period following onset of psychosis) on time to first relapse, using Cox proportional hazards regression in a multivariable model controlling for the effect of potential confounders (gender, ethnicity, relationship status, age of psychosis onset, care intensity at onset, diagnosis at onset, medication adherence, alcohol use, cigarette use, other illicit drug use). As the proportional hazards assumption was violated at different levels of cannabis use, the model was stratified by that variable. Kaplan-Meier plots (created using the 'survminer' package in R) were used to depict unadjusted survival data.

Two hundred fifty-six patients with first-episode psychosis were recruited into the study. Most of them were men (61%), of non-White ethnic origin (66%), and not in a relationship (74%). The prevalence of cigarette use was 57%, that of problematic alcohol use 14%, that of cannabis use 39%, and that of other illicit drug use 18%. The mean age at psychosis onset was 28.06±8.03 years. Most patients presented with non-affective psychosis (82%), were admitted to hospital close to the psychosis onset (78%) and in the context of a compulsory admission (60%).

Within two years from the onset of the disorder, 36% of recruited patients experienced at least one relapse of psychosis requiring hospital admission. The highest number of relapses recorded in the study period was three, with the longest hospital stay lasting 14.8 months.

Patients who had experienced any stressful life event following their psychosis onset (42%) had a significantly higher risk of relapse (as indexed by hospitalization) compared to those who did not experience any stressful life event (hazard ratio, HR=1.71, 95% CI: 1.11-2.64, p=0.016), after controlling for the above-mentioned socio-demographic and clinical factors (see also supplementary information).

Including medication adherence into the model, while still controlling for the socio-demographic and clinical factors, did not substantially change the results (HR=1.77, 95% CI: 1.13-2.79, p=0.013). A higher risk of relapse was observed as a function of the number of experienced stressful life events, but this was statistically significant only after adjusting for medication adherence as well as for the above-mentioned socio-demographic and clinical factors (HR=1.23, 95% CI: 1.01-1.50, p=0.037).

Among the socio-demographic and clinical factors controlled for, African ethnic origin, not being in a relationship, being a cigarette user, receiving a higher care intensity at onset (i.e., being hospitalized) and having poor medication adherence were all significantly associated with increased risk of relapse in survival analyses (see supplementary information).

In this study, we attempted to address most limitations of previous research. In particular, we used a prospective longitudinal design to avoid the recall bias that is inherent to retrospective studies[5]. Our results, therefore, provide evidence to support a temporal relationship between exposure to stressful life events and subsequent psychosis relapse, in line with the "triggering" hypothesis of psychosis[6]. Further, by restricting recruitment to first-episode patients, we were able to mitigate the potentially confounding effect of a highly variable clinical course of psychosis, that is especially relevant to patients suffering from psychotic disorders of longer duration.

Higher clinical severity at onset[7] and poor medication adherence[2] have been found to be robust indicators of subsequent admissions and poor outcome in patients with psychosis. Also, converging evidence supports higher odds of poor outcome among psychosis patients of non-White ethnic origin[8], and in cigarette users[9]. By including these predictors in our model, we found that our results were consistent with previous work, but we were able to add stressful life events to the list of risk factors for psychosis relapse that are supported by robust evidence.

By lending support to the notion that stressful life events can have a significant role in psychotic relapse, the present results

may have clinical and public health implications for the prevention and treatment of psychosis. In particular, they call for approaches allowing for real-time measurement of life events in clinical settings, so that timely interventions can be implemented to pre-empt potential adverse consequences.

Marco Colizzi[1,2], Alexis E. Cullen[1,3], Natasha Martland[1], Marta Di Forti[1], Robin Murray[1], Tabea Schoeler[1,4], Sagnik Bhattacharyya[1]
[1]Department of Psychosis Studies, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; [2]Unit of Psychiatry, Department of Medicine, University of Udine, Udine, Italy; [3]Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden; [4]Department of Computational Biology, University of Lausanne, Lausanne, Switzerland

S. Bhattacharyya has been supported by a UK National Institute for Health Research (NIHR) Clinician Scientist Award (NIHR CS-11-001) and by the UK Medical Research Council (MR/J012149/1). The authors acknowledge infrastructure support from the NIHR Mental Health Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. Supplementary information on the study is available at https://sites.google.com/view/supp-inf-colizzi-22-world-psy.

1. Martland N, Martland R, Cullen AE. Psychol Med 2020;50:2302-16.
2. Alvarez-Jimenez M, Priede A, Hetrick SE et al. Schizophr Res 2012;139:116-28.
3. Brugha T, Bebbington P, Tennant C et al. Psychol Med 1985;15:189-94.
4. Susser E, Finnerty M, Mojtabai R et al. Schizophr Res 2000;42:67-77.
5. Howard RW. Behav Res Methods 2011;43:931-41.
6. Day R. Acta Psychiatr Scand 1981;64:97-122.
7. Colenda CC, Hamer RM. Psychiatr Q 1989;60:227-36.
8. Bhugra D, Leff J, Mallett R et al. Psychol Med 1997;27:791-8.
9. Colizzi M, Carra E, Fraietta S et al. Schizophr Res 2016;170:311-7.

DOI:10.1002/wps.21046

Social determinants of health and selection bias in genome-wide association studies

The rapid pace of research continues to shed light on the complex genetic architecture that contributes to psychiatric disorder risk. The pace has been accelerated in part by the collection and analysis of increasingly larger samples of participants for genome-wide association studies (GWAS).

Large-scale GWAS typically report very small, statistically significant associations with numerous common variants, and increasingly produce polygenic risk scores (PRS) by combining associations of these variants with disorders in weighted or otherwise transformed summation scores. However, there are still fundamental limitations in how GWAS findings in psychiatry are generated and interpreted.

Two central limitations to scientific progress are the insufficient consideration of the social determinants of health and the selection bias in GWAS samples. These issues are also relevant to associations discovered in next generation fine mapping. Our views in this respect largely derive from psychiatric epidemiology, a field in which these issues have been prominent over a long period.

The goals of GWAS include identifying novel risk loci, quantifying genetic risk, and improving our understanding of the underlying pathophysiological mechanisms of mental disorders. The results tend to be presented in ways that are forward-looking, with claims that findings from GWAS could eventually be translated into useful applications. Implications for individualized risk prediction or "precision medicine" in clinical settings are often highlighted.

The success of GWAS in identifying large numbers of genetic variants that could be markers for risk loci depends heavily on the size of discovery samples, exemplified by the strong correlation between discovery sample size and number of statistically significant associations with genetic variants that are identified[1]. Indeed, it has been shown for some traits that the increasing rate of identification of statistically significant genetic loci does not appear to plateau, even as sample sizes surpass a million observations[1].

Given that some unknown proportion of these loci may signal a valid association, the study design of many GWAS has placed a strong emphasis on accruing larger and larger samples, with some authors calling – somewhat controversially – for prioritization of increasing sample sizes over other important features of study design[2].

There is no doubt that aggregating larger samples could be helpful in achieving the goals of GWAS. At the same time, however, an undue focus on sample size as the overriding priority in study design could undermine progress. We propose that more attention to the two fundamental features we discuss here – social determinants of health and selection bias – is essential to advance the potential of GWAS for understanding pathophysiology, and ultimately for contributing to clinical care and population mental health.

The prevalence of social factors within and across samples directly affects validity of GWAS findings with respect to mental disorders. The value of GWAS lies in their potential to identify genetic alleles that may influence a given phenotypic outcome. In the search for such alleles, however, GWAS rarely consider how the patterning of social determinants of health in their sample may influence observed results. The sample composition with respect to social determinants will have a major impact on the magnitude of the effects detected for individual genetic loci, as well as for PRS.

The impact of sample composition on effect size follows directly from the basic logic of epidemiology, and the potential to influence results has been evident in empirical studies – for example, in one study that considered characteristics such as acculturation and age at immigration, and their relationships with outcomes such as body mass index and blood pressure[3].

Emerging study designs, such as genome-wide environment interaction studies (GWEIS), combined with movements towards large-scale measurement of social determinants of health via electronic health records and linkage to biobanks, represent potentially

important steps toward ameliorating bias as well as detecting environmental influences on disorders[4].

An ill-considered implementation of GWEIS, however, may have the unintended consequence of reducing the social environment to an uninformative measure. For example, decades of psychiatric epidemiological research make clear that basic checklists, or worse, single questions about "stressful life events" are of limited (if not zero) utility for characterizing the social environment.

Moreover, GWAS rarely consider in any depth the theories that could explain the social patterning of mental health in their samples. Taking a social relational approach to studying the environment is critical. Social determinants of health include not only individual characteristics (e.g., income, highest completed education) but also structural determinants and social arrangements (e.g., class location). Recent studies suggest that important relationships between genetic risk and social context may be present even at the neighborhood level (e.g., collective efficacy)[5], and the influence of sociocultural context is likely to be greater at higher levels (e.g., racialized minority vs. dominant majority; nations with plentiful vs. scarce resources). Consequently, GWEIS limited to the narrow range captured by standard biological measures of the "exposome" will also be rather uninformative in this regard. We hope that current efforts to include social concepts and their measures will prove successful[6]. Altogether, grounding our study designs in evidence-based social theories will accelerate progress toward meaningful gene-by-environment investigations.

With regards to selection bias, increasing sample size allows valid signals of causation to emerge only when models are not misspecified. Otherwise, the *meaning* of the signals obtained in GWAS, and what these associations fundamentally represent – whether a true association, false positive, interaction with other genes, or something else – remains unclear.

Selection bias occurs when individuals in a study population differ systematically from, and are not representative of, the target population (the population that you want to make inferences about). When selection bias is present, large sample sizes will amplify biased results, which is often the case in studies using "well" or "normal" controls as well as those with minimal phenotyping[7].

The importance of remaining attuned to this aspect of study design therefore increases – not decreases – as sample size increases. In other words, while increases in sample size reduce random error, they amplify systematic error. For example, it is now generally accepted that GWAS findings are not directly transportable between populations of different ancestry. This is not only because of allelic variation across populations, but also because socially constructed categories such as "race" may intersect and interact with genes or with other causal factors in the environment[8].

As a result, researchers must be highly attentive to the causal architecture that underlies the PRS or other genetic measurements. At a minimum, specifying target populations and reporting what is known about potential selection bias from those populations, as well as specifying hypotheses about effect measure modifiers of PRSs, is essential for valid inferences based on GWAS results[9].

In sum, the fundamental goal of psychiatric epidemiology is disease prevention. Genetic psychiatry has immense potential to contribute to this cause, but must match the enthusiasm for large sample sizes with an equal consideration for study design and interpretation. A paradigm shift – away from an overriding focus on sample size, and toward comprehensive assessments of social determinants of health in genetic discovery samples – will better advance our understanding of the genetic architecture of mental disorders and of its implications for prevention.

Kira E. Riehm[1], Katherine M. Keyes[1], Ezra S. Susser[1,2]
[1]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA; [2]New York State Psychiatric Institute, New York, NY, USA

K.E. Riehm is supported by the National Institute of Mental Health (NIMH) Psychiatric Epidemiology Training Program (5T32MH013043-50) at Columbia University.

1. Visscher PM, Wray NR, Zhang Q et al. Am J Hum Genet 2017;101:5-22.
2. Mitchell BL, Thorp JG, Wu Y et al. JAMA Psychiatry 2021;78:1152-60.
3. Hollister BM, Farber-Eger E, Aldrich MC et al. Front Genet 2019;10:428.
4. Aschard H, Lutz S, Maus B et al. Hum Genet 2012;131:1591-613.
5. McDaniel T, Wilson DK, Coulon MS et al. Ann Behav Med 2020;55:708-19.
6. Robinson O, Tamayo I, De Castro M et al. Environ Health Perspect 2018;126:077005.
7. Schwartz S, Susser E. Psychol Med 2011;41:1127-31.
8. Borrell LN, Elhawary JR, Fuentes-Afflick E et al. N Engl J Med 2021;384:474-80.
9. Westreich D, Edwards JK, Lesko CR et al. Am J Epidemiol 2019;188:438-43.

DOI:10.1002/wps.21047

A Minimum Service Package (MSP) to improve response to mental health and psychosocial needs in emergency situations

Emergency situations such as armed conflicts, natural disasters, epidemics and famines deeply affect people's mental health and psychosocial well-being. Globally, one in five (22.1%) people living in areas affected by a conflict during the previous ten years have mental disorders such as depression, anxiety, post-traumatic stress disorder, bipolar disorder, or schizophrenia[1]. For children, adverse experiences in emergencies can disrupt cognitive, emotional, social and physical development, with enduring consequences[2,3]. Emergencies affect the availability of already sparse mental health services, and can erode the ability of families, caregivers and communities to support each other.

Over the years, attention to mental health and psychosocial support in emergencies has grown remarkably[4]. Mental health has now become a routine part of primary health care interventions in humanitarian settings[5]. The accumulation of evidence around a suite of brief scalable psychological interventions tai-

lored to the needs of people affected by emergencies has fueled optimism that we can effectively treat common mental health conditions with relatively modest means[6,7].

But there is no reason for complacency. The grim reality is that, in humanitarian settings, mental health and psychosocial support remains insufficiently prioritized, and programming is often still fragmented, inconsistent and inequitable[8]. There is a recognized need for a single easy-to-use package that strengthens collective humanitarian action by facilitating a unified response, integrating mental health and psychosocial support into various sectors such as health, protection, education, gender-based violence, nutrition, shelter, and camp coordination and management.

The new Mental Health and Psychosocial Support Minimum Service Package (MHPSS MSP) has been spearheaded by the World Health Organization (WHO), the United Nations (UN) International Children's Emergency Fund (UNICEF), the UN High Commissioner for Refugees (UNHCR), and the UN Population Fund (UNFPA). It is a resource for organizations which plan, support, coordinate, implement, fund and evaluate humanitarian activities. These include governments, national and international non-governmental organizations, civil society, Red Cross and Red Crescent networks, UN agencies, and financial donors. It outlines a set of activities that have the highest priority in meeting the immediate critical mental health needs of emergency-affected populations, based on existing guidelines, evidence, research, and expert consensus.

Each MSP activity is presented with a brief introduction explaining why the activity is important, a checklist of actions required to implement the activity safely and effectively, a list of relevant guidelines to support implementation, and associated costs (e.g., staff salaries) for consideration. For example, the section "Provide mental health care as part of general health services" briefly explains why this is needed (e.g., better accessibility, less stigma), specifies the recommended actions (e.g., adapting training materials, ensuring supervision), and lists relevant and up-to-date guidance (e.g., the WHO mhGAP Humanitarian Intervention Guide, mhGAP-HIG).

Humanitarian actors writing programme proposals can easily see what each activity entails and what budget is needed. Financial donors can use the MSP when making decisions about resource allocation. Persons coordinating the humanitarian response can see where the gaps are in different sectors (e.g., health, education), and what additional activities may be needed to meet the mental health needs of affected populations.

The MSP has been developed over three years based on literature reviews, consultations, and peer review by key stakeholders in global, regional and frontline positions. The initial draft was field-tested globally, with demonstration sites in Colombia, Iraq, North-East Nigeria, South Sudan, and Ukraine. Feedback was collected from hundreds of humanitarian actors across sectors and regions. The final version will be launched by the primary coordination body for humanitarian assistance, the Inter-Agency Standing Committee (IASC).

We have already received some preliminary feedback on how the MSP is informing the emergency response to the war in Ukraine. A staff member of an organization writing a regional mental health response strategy noted that the MSP allowed them to do this much more quickly and efficiently. A donor, reviewing a proposal to support psychiatric hospitals in Ukraine, observed that the MSP was helpful in understanding and evaluating the proposal. The MSP is also informing the development of a strategic mental health framework supported by the First Lady of Ukraine.

The MSP is relevant to any humanitarian emergency that requires a coordinated international response. However, it can also be relevant for smaller emergencies, for disaster risk reduction (especially relevant because of the climate crisis), and for longer-term development programming. Humanitarian crises have a long-lasting impact on mental health, and it is therefore essential to work from the onset towards mental health and social care systems that can be sustained over time[9].

The use of the MSP is expected to lead to better coordinated and more predictable, timely, and evidence-informed responses that make effective use of limited resources and improve the scale and quality of programming. It has the potential to be transformative and to give a major boost by prioritizing activities, providing a shared language for advocacy and planning, and supporting coordinated implementation of activities. This should ultimately lead to better mental health outcomes for large numbers of emergency-affected people, including vulnerable groups, who often receive less attention and investment.

Inka Weissbecker[1], Caoimhe Nic A. Bhaird[2], Vania Alves[2], Peter Ventevogel[3], Ann Willhoite[2], Zeinab Hijazi[4], Fahmy Hanna[1], Prudence Atukunda Friberg[5], Henia Dakkak[6], Mark van Ommeren[1]
[1]Department of Mental Health and Substance Use, World Health Organization, Geneva, Switzerland; [2]Child Protection Programme, UNICEF, New York, NY, USA; [3]Public Health Section, UNHCR, Geneva, Switzerland; [4]Office of Director of Programmes, Programme Division, UNICEF, New York, NY, USA; [5]Humanitarian Unit, Act Church of Sweden, Uppsala, Sweden; [6]Humanitarian Office, UNFPA, New York, NY, USA

The MSP field test version is under review by the IASC for endorsement and is accessible at https://mhpssmsp.org/en. Funding for the project has been provided by the Netherlands Ministry of Foreign Affairs; Education Cannot Wait; the UK Foreign, Commonwealth & Development Office; and the Global Protection Cluster. The authors alone are responsible for the views expressed in this letter, that do not necessarily represent the views, decisions or policies of the institutions with which they are affiliated.

1. Charlson F, van Ommeren M, Flaxman A et al. Lancet 2019;394:192-4.
2. UNICEF. The state of the world's children 2021. New York: UNICEF, 2021.
3. World Health Organization. World mental health report: transforming mental health for all. Geneva: World Health Organization, 2022.
4. Jones L, Ventevogel P. World Psychiatry 2021;20:2-3.
5. Fine SL, Kane JC, Spiegel PB et al. BMC Med 2022;20:183.
6. Acarturk C, Uygun E, Ilkkursun Z et al. World Psychiatry 2022;21:88-95.
7. Bryant RA, Bawaneh A, Awwad M et al. PLoS Med 2021;19:e1003949.
8. MHPSS Collaborative. Follow the money. Copenhagen: MHPSS Collaborative, 2021.
9. Epping-Jordan JE, van Ommeren M, Ashour HN et al. Int J Ment Health Syst 2015;9:15.

DOI:10.1002/wps.21048

Moving away from the scarcity fallacy: three strategies to reduce the mental health treatment gap in LMICs

The mental health treatment gap – defined as the difference between the number of people who have mental disorders and those who can access appropriate treatment – is estimated to be as high as 85% in low- and middle-income countries (LMICs), compared to only 40% in high-income countries (HICs)[1]. This high treatment gap in LMICs is unacceptable and represents an urgent global health priority. Here, we argue that possible solutions to reduce this gap do exist within many LMICs.

Conventionally, we often blame resource scarcity for the higher mental health treatment gap in LMICs. This includes: a) human resource scarcity (shortage of specialized personnel), b) financial scarcity (income shortage at individual, family and national levels), and c) structural scarcity (e.g., concerning infrastructure, health systems, and policies)[1]. From our experience, we believe that this resource scarcity mindset is a fallacy. It is time to move away from this mindset if we are to minimize the mental health treatment gap in LMICs.

Of course, it is true that, when compared to HICs, LMICs have a scarcity of mental health experts. But such an argument is based on the idea that the HIC model – where care is provided by "expert" caregivers such as psychiatrists and psychologists – is the gold-standard approach that LMICs must first attain to reduce the treatment gap. This Western/HIC model, however, does not often translate into accessible and effective care in LMICs[2]. Rather than focusing mostly on *expert* caregivers, LMICs should find answers to two important questions: a) what kind of human resources do we have in our hands, and b) how can we innovatively use these resources alongside the rather more expensive and relatively unavailable specialists?

Indeed, there is a huge pool of utilizable human resources that can be trained to recognize symptoms of mental disorders, offer first aid psychosocial support, and refer upwards and accept referrals downwards for continued support. These include: a) families, that are traditionally the primary caregivers; b) individuals with lived experience, who can be supportive of people with mental disorders; c) an inexhaustible pool of clergy and traditional healers, who are often the first contacts for care even when specialists are available; d) community health volunteers, who are the backbone of community health services and the link between families, communities and community facilities; e) school teachers and counselors available in every school; f) peer counselors in schools, colleges and universities, who are trusted within their circles more than others outside those circles, including specialists; g) the nurses and clinical officers at community health service centers; h) general physicians working in communities.

We have found that these different human resources can be expertly engaged to provide evidence-based interventions using the World Health Organization (WHO)'s mhGAP Intervention Guide (mhGAP-IG)[3], and that peers – as young as 18 to 22 years – can provide evidence-based intervention in schools[4]. We, therefore, aver that in a way LMICs are not human resource poor, but rather that they have abundant resources which can be coopted into delivery of mental health services.

As to financial scarcity, poverty at individual, familial and national levels often leads to inaccessibility of expensive psychotherapies and unavailability of psychotropic medications in LMICs. However, expensive psychotherapies can now be replaced by inexpensive ones delivered by trained lay providers[5], and less costly generic medications are increasingly becoming available. Furthermore, a dialogue with families and patients should be encouraged about the costs of medications vis-à-vis what they can afford within their means, and when and where to seek help.

As to structural scarcity, it is our experience that there is an oversupply of infrastructure that can be used, at almost no cost, for psychoeducation, treatment efforts awareness, prevention and rehabilitation in LMICs. These include: a) homesteads; b) community halls and squares, church and school halls, and open marketplaces; c) waiting places at community health facilities; d) the already existing social support systems, from family to community levels; e) the often used meeting places under trees.

Beyond moving away from the resource scarcity fallacy, efforts that prioritize fostering a team spirit can also be crucial in reducing the mental health treatment gap[6]. These may include bringing together different relevant stakeholders at the community level, including any available mental health experts[7], to engage in participatory dialogues on perceptions of mental illness; impact of mental health on individuals, their families and communities; and human rights and mental health. Dialogues can also identify perceived barriers to mental health care, such as stigma, and how these barriers can be overcome. Importantly, this approach promotes community ownership and responsibility for good mental health. Of course, the composition of dialogue will vary, but should – at the minimum – include patients, families, community opinion leaders, service providers and policy makers.

How we think about recovery is also important. On the one hand, recovery can be defined to mean a complete disappearance of symptoms. On the other, it can mean a reduction of symptoms that allows the patient to engage in other equally pressing life priorities. Consider a mother who suffers from depression. She often must make an informed decision on whether to attend a clinical appointment or if she is feeling well enough to prioritize getting food for and taking care of her children[8]. Whereas a clinician may not consider her "recovered", she may consider herself "well enough" and "recovered" to make the informed decision to prioritize caring for her children. Thus, a contextual determination of recovery is important, because our conceptualization of the treatment gap is affected by how we define recovery.

We believe that solutions to reduce the mental health treatment gap already exist in many LMICs. We have listed three possible strategies here. What gives us hope is that across our work we have demonstrated that these three strategies can feasibly allow us to deliver affordable, available, accessible and evidence-based

mental health services, and to perhaps reduce the treatment gap to levels seen in HICs using the currently available resources[3,4].

Of course, there are other strategies, such as promoting liaisons between different disciplines to provide a one-shop holistic and integrated approach to management of physical and mental comorbidities and associated psychosocial determinants; maximizing the integration of technology to increase access to mental health[9]; and collaborative LMIC and HIC research on cost-effective treatments, risk and protective factors – including biomarkers – and priorities in global mental health.

If we rethink strategies and models and prioritize those that are innovative and context appropriate, we can reduce the treatment gap in LMICs with existing resources even as new resources continue to be developed.

David M. Ndetei[1-3], Victoria Mutiso[2,3], Tom Osborn[4]
[1]Department of Psychiatry, University of Nairobi, Nairobi, Kenya; [2]Africa Mental Health Research and Training Foundation, Nairobi, Kenya; [3]WPA Collaborating Centre for Research and Training, Nairobi, Kenya; [4]Shamiri Institute, Nairobi, Kenya

1. World Health Organization. mhGAP Intervention Guide for mental, neurological and substance use disorders in non-specialized health settings, version 2.0. Geneva: World Health Organization, 2016.
2. Murray LK, Dorsey S, Haroz E et al. Cogn Behav Pract 2014;21:111-23.
3. Mutiso VN, Musyimi CW, Rebello TJ et al. Soc Psychiatry Psychiatr Epidemiol 2018;53:1277-87.
4. Osborn TL, Venturo-Conerly KE, Arango S et al. JAMA Psychiatry 2021;78:829-37.
5. Sharan P. World Psychiatry 2022;21:466-7.
6. Mutiso VN, Gitonga I, Musau A et al. Int J Ment Health Syst 2018;12:1-13.
7. Kestel D. World Psychiatry 2022;21:333-4.
8. Mendenhall E, Kohrt BA, Norris SA et al. Lancet 2017;389:951-63.
9. Hall J, Jordan S, van Ommeren M et al. World Psychiatry 2022;21:156-7.

DOI:10.1002/wps.21054

Update on implementation of WPA Action Plan 2020-2023

Despite the ongoing impacts of the COVID-19 pandemic and the difficulties in getting connected[1,2], the WPA has remained active in its professional work and in meeting its objectives during the last two years. As the situation is getting a bit better now, our work to implement the current triennium Action Plan also gets a noticeable momentum[3]. The WPA Executive Committee and its different components as well as the Secretariat staff are committed to fulfill their responsibilities[4-8].

The WPA has strengthened its virtual work and e-learning programmes among colleagues and trainees around the world, which has led to offering more online educational activities. Our educational portal is providing an excellent educational resource to our membership, and we were delighted to produce several educational modules, courses, teaching sessions and online training programmes[8].

The enhanced and quicker development of the WPA education learning management system has also promoted the development of new education and training modules. Like many other professional organizations, the WPA has encouraged and supported its membership to use e-mental health tools and e-learning techniques. The portal also gives ready access to WPA's existing training materials available in several languages.

The available programmes on our educational portal and learning management system include new modules such as those on comorbidities of mental disorders, mood disorders and dementias. Recently held courses have attracted a lot of contacts, especially those on ICD-11, tele-psychiatry and yoga. Similarly, webinars on early intervention in psychosis, updates in psychopharmacology, psychotherapy,and child and adolescent psychiatry have attracted many participants.

Various programmes outlined in the Action Plan 2020-2023 are also ongoing. Working Groups are implementing several initiatives in areas of training and research, and clinical updates. Activities by the Working Groups on Managing Comorbidity of Mental and Physical Health, Early Intervention in Psychosis, Public Mental Health, and Promotion of Psychiatry Among Medical Students are drawing additional attention to the current needs and opportunities in these areas of work[9-11].

The support of the WPA Scientific Sections is playing a vital role in our activities. In a highly stimulating way, the Sections' work is providing a great motivation to young psychiatrists to benefit from experts' contributions. Similarly, WPA's network of Collaborating Centres[12] has been involved in various scientific initiatives, including joint educational seminars and support to young psychiatrists in research and other related activities.

The WPA Collaborating Centre Group and the WPA Working Group on Medical Students offered medical students and psychiatric trainees the opportunity to obtain travel fellowships to attend the 22nd World Congress of Psychiatry in Bangkok. Psychiatric trainees were invited to submit an essay on the topic of "Forced displacement and mental health: challenges and resilience", while medical students were invited to submit an essay on "Breaking the silence: how is stigma a barrier to mental health?". Over 40 submissions were received from 15 countries from the trainees, and over 150 entries were submitted by the medical students from 39 different countries. The quality of entries was outstanding, and the judges were pleased to review and assess so many good essays from around the world, which is an encouraging reflection of the talent amongst future psychiatrists.

Unfortunately, in addition to the COVID-19 pandemic, several other adversities affected us in many parts of the world during this triennium. The WPA established an Advisory Committee for Responses to Emergencies (ACRE), that brought together the leaders of the larger Member Societies to facilitate practical and concrete support to Member Societies in need. This work continued mobilizing and fostering collaboration, information collection, and development of local, national and international strategies to cope with the mental health consequences of emergencies throughout 2021 and 2022.

In response to the war, and due to the grave concerns about the well-being of the Ukrainian people, particularly psychiatric patients and psychiatric staff, the ACRE planned and initiated its support to Ukraine. Reflecting our long-standing opposition to non-defensive military activities and mindful of the recent statements of various health and welfare organizations, as well as the vote of the United Nations General Assembly condemning the invasion of Ukraine, the WPA also expressed grave concern at reports of attacks on civilian facilities such as private residences, schools and hospitals, and of civilian casualties, including children, women, older persons, and persons with disabilities.

The WPA also established an educational trauma resource centre on its website for mental health professionals, in Ukrainian, Russian and other languages, to help with the mental health challenges that people from Ukraine are currently facing. Support from our Member Societies and other components was very encouraging[13].

Looking at Afghanistan's deteriorating conditions, that are not only causing a humanitarian crisis, but also adding concerns about provisions and delivery of health care for the general population, the WPA, as a part of its ACRE project, worked with its fellow Afghan mental health professionals to offer ongoing support through the provision of medicines, patient assessments and training. Similarly, the WPA offered support for buying psychotropic medicines to Sri Lanka, as the country was going through the worst economic crisis that it has faced in its history.

With the start of the WPA eNewsletter in 2021, we are facilitating sharing of activities and reports from our membership. The Newsletter has emerged as a strong medium for our visibility on the social media platform and a better communication among different components of the Association[14].

World Psychiatry, the WPA official journal, has achieved the unprecedented impact factor of 79.683 and continues ranking as the number one in the list of psychiatric journals and in the Social Sciences Citation Index. The journal is published regularly in

three languages (English, Spanish and Russian), with individual issues or articles also available on the WPA website in other languages (Chinese, French, Arabic, Turkish, Japanese, Romanian and Polish). More than 60,000 mental health professionals regularly receive the electronic or the print version of the journal. All the back issues can be freely downloaded from the PubMed system and the WPA website.

We very much enjoyed our successful hybrid World Congress of Psychiatry that took place in Bangkok in August 2022. While adapting and innovating new resources, we were able to redesign the scientific programme and ensured coverage of the most timely clinical, academic and research topics to our membership[15]. I am also pleased that we are actively working for our next World Congress to be held in Vienna, Austria from September 28 to October 1, 2023.

As we kick off for the last year of this triennium, we are positive that the challenges that will undoubtedly come, as the long-term impact on mental health following this pandemic becomes more evident, will be addressed effectively. We are enthusiastic and learning fast with the changes and look forward with confidence to the future, remaining fully committed and confident to fulfilling our triennium's goals, and to continuing with our efforts to build up the future of psychiatry and mental health together.

Afzal Javed
WPA President

1. World Health Organization. The impact of COVID-19 on mental, neurological and substance use services: results of a rapid assessment. Geneva: World Health Organization, 2020.
2. Konstantinos N, Fountoulakis KN, Grigorios N et al. Psychiatry Res 2022;315:114702.
3. Javed A. World Psychiatry 2022;21:325-6.
4. Wasserman D. World Psychiatry 2021;20:309-10
5. Morozov PV. World Psychiatry 2021;20:310-1.
6. Pi E. World Psychiatry 2022;21:477-8.
7. Botbol M. World Psychiatry 2022;21:479-80.
8. Ng RMK. World Psychiatry 2022;21:478-9.
9. Thomson S, Chumakov E, Van Hoof J et al. World Psychiatry 2022;21:161-2.
10. Azeem MW, Liu HY, Imran N et al. World Psychiatry 2022:21:328-9.
11. Campion J, Javed A. World Psychiatry 2022;21:330-1.
12. Bhui KS, Fiorillo A, Stein D et al. World Psychiatry 2016;15:300.
13. Wasserman D, Ng R. WPA ACRE sub-committee created online resource centre to support Ukrainians experiencing trauma. www.wpanet.org.
14. Javed A. Re-launch of the WPA Newsletter: WPA Review! www.wpanet.org.
15. Javed A. WCP22 closed the doors on it's first in-person World Annual Congress in 2 years. www.wpanet.org.

DOI:10.1002/wps.21049

WPA's humanitarian actions for Ukrainian psychiatrists and psychiatric patients

In May 2020, the WPA established an Advisory Committee for Responses to Emergencies (ACRE), by bringing together the leaders of the larger Member Societies to provide practical and concrete aid to Member Societies in need. The group aimed at fostering education, information collection, and the development of local, national and international strategies to cope with the mental health consequences of emergencies. The aid was given in many parts of the world, including Asia, America and other regions[1-4].

Following Russia's invasion of Ukraine in February 2022, and due to the grave concerns about the well-being of Ukrainian people, particularly psychiatric patients and staff, the ACRE set up a sub-committee, chaired by the WPA President-Elect, to plan and implement support to Ukrainian psychiatrists and the Ukrainian Psychiatric Associations by actively providing humanitarian and medical aid through WPA Member Societies[5]. Presently, there are approximately 6.6 million Ukrainian people displaced within the country, and almost equal numbers of refugees in Europe, some of them reaching other continents[6].

A close collaboration was established between the WPA and the leadership of the European Psychiatric Association (EPA) (President: P. Falkai; President-Elect: G. Dom; Secretary General: J. Beezhold) as well as the EPA Council of National Psychiatric Associations (Chair: J. Samochowiec)[7].

The war in Ukraine affects both the physical and mental health of Ukrainian people. Supporting mental health of the population as well as providing support for persons with mental ill health is key. Therefore, one of the main goals of the ACRE subcommittee was to establish a WPA online trauma resource center, under the leadership of R. Ng (Interim WPA Secretary General and WPA Secretary for Education).

The EPA, whose national psychiatric associations are also Member Societies of the WPA, made available a repository of literature on treatment of people with trauma. Furthermore, a series of webinars, *Help for Helpers*, specially designed for people working in war conditions, was created by mental health professionals. The goal was to provide knowledge to the public on how to help traumatized family members, friends and neighbors.

The WPA online trauma resource center was established as a central point to collect and provide evidence-based materials and resources in Ukrainian, Russian and other languages, to help psychiatrists and other individuals to respond to the mental health challenges that people from Ukraine are currently facing.

In creating this resource center, the WPA and its Scientific Sections established collaborations with a number of professional organizations in addition to the EPA, including the European College of Neuropsychopharmacology (ECNP)[8] and Mental Health First Aid (MHFA)[9,10], to bring together relevant self-help materials to support those in need.

The WPA educational trauma resource center can be visited at www.wpanet.org/ukraine-resources. In this center, readers are provided with a one-stop station where they can find a quick overview of the existing educational materials offered by the various organizations. They can then click on the relevant links and be re-directed to the educational materials in the webpages

of the organizations. The materials include written guidelines, resource packages, videos, and webinars for mental health professionals on delivering psychological support and crisis intervention to refugees and displaced people. There are also self-help online materials for war victims, refugees, and their caregivers.

Supplementing the online resources in the trauma resource center, the WPA website also hosts an educational portal in which there are over 20 free webinars and learning modules covering a diverse range of mental health topics, that can be readily accessed by mental health professionals supporting war victims, refugees, and displaced persons. Finally, there is a list of volunteer organizations in Europe that provide free online consultations and support to Ukrainian people in need.

The WPA educational trauma resource center is updated as we receive more and new information. If you wish to contribute any relevant resources developed by your own organization, please contact the WPA Secretariat (wpasecretariat@wpanet.org).

WPA Member Societies have also provided direct help to Ukraine. Moreover, they have helped refugees in the receiving countries with psychiatric aid. Specialized psychiatric services for women and children with a focus on Ukrainian families have been established. Many Member Societies have appealed to numerous governmental and non-governmental organizations, as well as to pharmaceutical companies, to increase their awareness of psychiatric patients' needs in Ukraine, including their demand for psychotropic drugs. Almost all European national psychiatric societies have undertaken numerous relevant activities[11,12].

The EPA and the Polish Psychiatric Association regularly invite the WPA to attend their meetings with the two Ukrainian Psychiatric Associations, as well as with the neighboring European countries' psychiatric associations, to continually discuss the needs regarding humanitarian and medical aid in Ukraine. The transfer to Ukraine of several medical supplies, including psychotropic drugs, was provided by the central office of Lundbeck in Europe, stimulated by the WPA. A series of medication transports to Ukrainian hospitals, based on the lists provided by the Ukrainian Psychiatric Associations, were organized by the local Lundbeck subsidiary. The primary needs are for antipsychotic medications in the form of short-acting intramuscular injections and long-acting injections.

The Polish Psychiatric Association purchased electric generators and delivered them to Lviv, to be then transferred to other psychiatric hospitals in Ukraine, including Odessa, Chernihiv, Mykolaiv, Zaporizzhia, and Ivano-Frankivsk regions. The transfer of ambulances for community psychiatry in Lviv region is on the way. Moreover, the Association sent basic equipment, sleeping mats, bedding, mattresses, backpacks, cleaning products, personal hygiene products, tools for renovation and construction, as well as psychiatric medications through the Polish Agency for Materials and Strategic Reserves.

The other main goal of the WPA ACRE sub-committee is to offer economic support through donations from WPA Member Societies. The EPA's *Fund for Ukraine* supports Ukrainian psychiatric units and patients for the purchase of medications, equipment and other needed materials. The WPA plans to use its own fund to help psychiatrists in Ukraine in reconstructing their services during and after the war.

Generous donations have been made by the American Psychiatric Association, the Royal Australian and New Zealand College of Psychiatrists, the Japanese Society of Psychiatry and Neurology, the Mexican Psychiatric Association; the German Association for Psychiatry, Psychotherapy and Psychosomatics; the Croatian Psychiatric Association, the Hungarian Psychiatric Association, the Finnish Psychiatric Association, the Italian Psychiatric Association, the French Psychiatric Associations, and the Polish Psychiatric Association. Donations have also come from individual psychiatrists around the world.

If you wish to donate to the WPA fund for this purpose, you can do this using the link www.wpanet.org/post/call-for-donations-to-supply-medications-for-mentally-ill-patients-in-ukraine.

Danuta Wasserman[1,2], Helen Herrman[3,4], Afzal Javed[5,6], Roger M.K. Ng[7,8], Peter Falkai[9,10], Jerzy Samochowiec[11,12], Jonas Eberhard[13], Norman Sartorius[14]

[1]WPA President Elect; [2]Karolinska Universitet, Stockholm, Sweden; [3]WPA Past President; [4]University of Melbourne, Melbourne, VIC, Australia; [5]WPA President; [6]Pakistan Psychiatry Research Centre, Fountain House, Lahore, Pakistan; [7]WPA Interim Secretary General; [8]Department of Psychiatry, Central South University, Hunan, China; [9]EPA President; [10]Max Planck Institute of Psychiatry, Munich, Germany; [11]Chair, EPA Council of National Psychiatric Associations; [12]Pomeranian Medical University, Szczecin, Poland; [13]Division of Psychiatry, Department of Clinical Sciences, Lund University, Lund, Sweden; [14]Association for the Improvement of Mental Health Programmes, Geneva, Switzerland

1. Javed A. World Psychiatry 2021;20:146.
2. Javed A. World Psychiatry 2021;20:451-2.
3. Javed A. World Psychiatry 2022;21:325-6.
4. Herrman H, Chkonia E, Pinchuk I et al. World Psychiatry 2021;20:147-8.
5. World Psychiatric Association. www.wpanet.org.
6. United Nations. Ukraine emergency. www.unhcr.org/ukraine-emergency.html.
7. European Psychiatric Association. www.europsy.net.
8. European College of Neuropsychopharmacology. www.ecnp.eu.
9. Mental Health First Aid Australia. mhfa.com.au.
10. Mental Health First Aid. www.mhfa-ersthelfer.de.
11. European Psychiatric Association. War in Ukraine: actions from the EPA and its Member Associations. mailchi.mp/europsy.net/interact-n12018-newsletter-of-epa-council-of-npas.
12. European Psychiatric Association. War in Ukraine: reactions from the EPA, the Council of NPAs and Member Associations. mailchi.mp/europsy.net/interact-n12018-newsletter-of-epa-council-of-npas.

DOI:10.1002/wps.21050

WPA Working Group on Defining and Managing Autism Spectrum Disorder: spreading knowledge for the next generations of psychiatrists

Autism spectrum disorder (ASD) is believed to affect approximately one in every 100 individuals on the planet, across ethnicities and socioeconomic categories[1], although information from low- and middle-income countries remains very scarce. Actually, 2.3% of American youngsters have been diagnosed with autism[2], a number that has nearly tripled over the previous 20 years.

Individuals with ASD show a high co-occurrence of other neurodevelopmental disorders and adjunctive mental health problems. Intellectual disability is reported in 35.2% of cases, and borderline intellectual functioning in 23.1%, while only 41.7% have IQ scores in the average or higher range[2]. Attention-deficit/hyperactivity disorder (ADHD) is also quite common, with peaks of 50%[3]. Adjunctive psychiatric problems have been observed to be up to six times more prevalent than in the general population, and to present differently, particularly in those with poor or no verbal communication abilities who only convey their distress through behavioral manifestations[4].

Many of these patients lack access to the essential diagnosis and care, which leads to overmedication. Approximately half of people with ASD, particularly those in the low-functioning portion, receive psychotropic drugs[4]. In about one-third of situations, medications are administered in the absence of a psychiatric diagnosis with the aim to treat problem behaviors, such as aggressive or self-injurious behaviors, and/or without proper follow-up and tapering when feasible[5].

Significant training gaps for psychiatrists and other mental health professionals have been identified in this area at all levels of the clinical education system, in addition to knowledge, planning and service delivery challenges[4].

The scientific community's disregard for the mental health of those with low-functioning ASD and/or intellectual disability has not only been unfair, but has also been inappropriate, because the advancement of scientific knowledge in this area may have important implications for the entire neuroscientific field: for instance, the ability to recognize psychiatric symptoms in patients with cognitive and communicative limitations based on observable and behavioral changes from baseline; the understanding of the relationship between early specific cognitive deficits and psychopathological vulnerability; the definition of the grade of adjunctive functional impairment and clinical distress associated with the co-occurrence of psychopathological conditions. Even care models that were initially developed in the field of ASD/intellectual disability, as well as models to address social health issues (such as stigma and labeling) may be now relevant to general psychiatry and other neuroscientific disciplines.

The professional training gap and other unmet mental health needs related to ASD are paid prominent attention in the Action Plan 2020-2023 of the WPA Working Group on Defining and Managing ASD. A major objective of this Working Group, which has just been achieved, is the publication of a comprehensive textbook on psychiatric disorders in people with low-functioning ASD and/or intellectual disability, including the most recent research knowledge on the prevalence, risk and etiological factors, clinical features, assessment procedures and tools, diagnostic criteria, treatment and prognosis[6].

This volume, entitled *Textbook of Psychiatry for Intellectual Disability and Autism Spectrum Disorder*, has been produced under the aegis of the WPA, and includes 43 chapters written by 116 of the most authoritative experts in the area. It has been edited by M. Bertelli, S. Deb and K. Munir, from the WPA Section on Psychiatry of Intellectual and Developmental Disorders, and A. Hassiotis and L. Salvador-Carulla, outstanding contributors to the WPA activities related to ASD and intellectual disability.

The book has been inspired by the will of sharing knowledge and transmitting passion to colleagues, especially young and future colleagues. In fact, it is intended for use by graduate students and trainees of university faculties, practitioners in clinical disciplines or people having management roles in developmental disability services and education, and to a lesser extent by undergraduate students, parents, attorneys and advocacy groups.

This textbook helps clinicians to overcome diagnostic challenges and provide more effective care that is tailored to the specific needs of individuals with ASD and/or intellectual disability. Researchers will find in the volume a summary of current knowledge about an area of psychiatry that is new to them or that intersects their own specialty in the wider field of neurodevelopmental disorders.

Beside the production of the textbook, the WPA Working Group on Defining and Managing ASD has started the development of educational materials on key diagnostic features of ASD and co-occurring mental health issues for the WPA Educational Portal[7], in connection with a similar initiative of the WPA Working Group on Intellectual Developmental Disorders[8]. A prominent attention has been paid to the provision of strategies for interdisciplinary approaches, according to the overall WPA Action Plan 2020-2023[9,10], and to the promotion of partnerships for joint collaborative work in capacity building among medical students[11], young psychiatrists and allied professionals.

A further contribution to the improvement of training and practice in the field has been provided through the participation in the development of the World Health Organization (WHO)'s package of rehabilitation interventions for persons with ASD.

The importance of spreading knowledge to the next generation of psychiatrists and other mental health professionals has also been the focus of numerous presentations in presidential and special symposia organ-

ized by the Working Group at the last years' WPA World Congresses, and will continue to be highlighted at upcoming WPA events.

Marco O. Bertelli[1,2], Annamaria Bianco[2], Luis Salvador-Carulla[1-3], Afzal Javed[4,5]
[1]WPA Working Group on Defining and Managing Autism Spectrum Disorder; [2]CREA (Research and Clinical Centre), Fondazione San Sebastiano Misericordia, Florence, Italy; [3]Health Research Institute, Faculty of Health, University of Canberra, Canberra, NSW, Australia; [4]WPA President; [5]Pakistan Psychiatric Research Centre, Fountain House, Lahore, Pakistan

1. Zeidan J, Fombonne E, Scorah J et al. Autism Res 2022;15:778-90.
2. Maenner MJ, Shaw KA, Bakian AV et al. MMWR Surveill Summ 2021;70:1-16.
3. Hours C, Recasens C, Baleyte J. Front Psychiatry 2022;13:837424.
4. Bertelli MO, Salvador-Carulla L, Munir KM et al. World Psychiatry 2020;19:260.
5. Deb S, Kwok H, Bertelli M et al. World Psychiatry 2009;8:181-6.
6. Bertelli MO, Deb S, Munir K et al (eds). Textbook of psychiatry for intellectual disability and autism spectrum disorder. Cham: Springer, 2022.
7. Ng RMK. World Psychiatry 2022;21:478-9.
8. Munir KM, Roy A, Javed A. World Psychiatry 2022; 21:327-8.
9. Javed A. World Psychiatry 2021;20:451-2.
10. Javed A. World Psychiatry 2022;21:325-6.
11. Azeem MW, Liu HY, Imran N et al. World Psychiatry 2022;21:328-30.

DOI:10.1002/wps.21051

The relationship between physical and mental health: an update from the WPA Working Group on Managing Comorbidity of Mental and Physical Health

Compared with the general population, patients suffering from severe mental disorders have a 10 to 25-years shorter life expectancy, which requires urgent action from health care professionals and governments worldwide[1,2]. The factors associated with this high mortality rate can be grouped into those related to patients themselves, to psychiatrists, to other medical professionals, and to the health care system at large.

Among factors related to the persons with mental disorders, a significant role is played by the presence of comorbid physical illnesses – cardiovascular, respiratory, metabolic, infectious diseases, cancer and others – all of which are frequently given little attention in ordinary psychiatric practice[3,4].

Among the reasons for the high rates of physical comorbidity and its contribution to mortality of people with mental disorders is the long-standing separation of psychiatry from other branches of medicine, as well as the lack of attention of several psychiatrists to the physical health of their patients[5-8]. In addition, the collaboration of psychiatrists with primary care physicians and other clinicians is often poor, and other health care professionals often have negative attitudes towards people with mental disorders, underestimating the seriousness of their physical complaints.

Recently, several international bodies and associations, such as the World Health Organization (WHO), the WPA, the European Psychiatric Association, the UK Royal College of Psychiatrists and the UK Royal College of Practitioners, have taken action to improve the management of physical health of people with severe mental disorders. Among these activities, the revision of educational curricula for health care professionals has been proposed[9]. In 2017, the WPA created a Scientific Section on Comorbidity, and in January 2021 it established a Working Group on Managing Comorbidity of Mental and Physical Disorders chaired by N. Sartorius. The group includes experts in the field with different backgrounds from high-, medium- and low-income countries[10-12].

This Working Group has been requested: a) to identify areas of promising work related to comorbidity of mental and physical disorders, and to develop recommendations for WPA's involvement in research, education and service development concerning problems related to that comorbidity; b) to identify individuals and centres interested and willing to participate in WPA's program of research and education related to the comorbidity of mental and physical disorders; c) to liaise with other WPA Working Groups, with a view to ensure that problems of comorbidity are considered in the work of those groups; d) to propose the organization of symposia, workshops and other types of meetings addressing problems related to comorbidity of mental and physical disorders; e) to prepare reviews of evidence and drafts of position papers; f) to build up training programs (see https://www.wpanet.org/wg-on-comorbidity).

These tasks are being addressed by: a) the organization of collaborative and intersectional symposia and workshops during the World Congresses of Psychiatry, as well as during WPA Thematic and Regional Meetings; b) the development of a range of recorded lectures, live and recorded webinars, and resource documents; c) support to the development of in-country capacity in low-resource settings through the facilitation of high-impact activities and regional collaborations; d) support to the publication of articles in scientific journals as well as chapters in leading textbooks; e) partnership with national and international agencies such as the WHO, the United Nations International Children's Emergency Fund (UNICEF), the US National Institutes of Health (NIH), the Wellcome Trust, and the International Initiative for Disability Leadership, among others, in order to obtain funding in support of good clinical practice, research and training with relevance to low-resource countries; f) support to government initiatives, plans and policies as they intersect with the Working Group's remit; g) development of joint initiatives with other WPA Working Groups and Scientific Sections, in the salient areas of public mental health, and child and adolescent mental health; h) providing a selection of evidence-based interventions appropriate for service delivery platforms in low-resource regions;

i) creating a list of training and resources available to implement relevant interventions.

In March 2022, the Working Group organized a webinar on "Physical illnesses in patients with severe mental disorders: current challenges and practical implications for professionals", attended by more than 500 health care professionals, trainees in psychiatry and medical students, focusing on the complex interplay between physical and mental disorders. During the 22nd World Congress of Psychiatry, the Working Group organized a course on the same topic, which was very well attended. The topic of comorbidity was also discussed in the main plenary session and in a state-of-the-art symposium of the World Congress.

The Group has developed and made available on the WPA website educational materials on the comorbidity between depressive disorders and diabetes, depression and cancer, and depression and cardiovascular diseases (www.wpanet.org).

The Group is currently engaged in the organization of a series of free WPA webinars on comorbidity between mental disorders and infectious diseases (i.e., HIV, tuberculosis, COVID-19), and has started a collaboration with the International Society of Addiction Medicine, in order to organize educational activities related to the management of addictions and comorbid physical illness in people with severe mental disorders.

Andrea Fiorillo[1], Giovanni de Girolamo[2], Ivona Filipcic Simunovic[3], Oye Gureje[4], Mohan Isaac[5], Cathy Lloyd[6], Jair Mari[7], Vikram Patel[8], Andreas Reif[9], Elena Starostina[10], Paul Summergrad[11], Norman Sartorius[12]

[1]Department of Psychiatry, University of Campania "L. Vanvitelli", Naples, Italy; [2]Unit of Epidemiological Psychiatry, IRCCS S. Giovanni di Dio Fatebenefratelli, Brescia, Italy; [3]Department of Psychological Medicine, University Hospital Center, Zagreb, Croatia; [4]Department of Psychiatry, University College Hospital, Ibadan, Nigeria; [5]Division of Psychiatry, University of Western Australia and Fremantle Hospital, Fremantle, Australia; [6]School of Health, Wellbeing and Social Care, Open University, Milton Keynes, UK; [7]Department of Psychiatry, Universidade Federal de São Paulo, São Paulo, Brazil; [8]Department of Global Health and Social Medicine, Harvard Medical School, Boston, MA, USA; [9]Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, Goethe University Hospital, Frankfurt, Germany; [10]Vladimirsky Moscow Regional Research and Clinical Institute, Moscow, Russia; [11]Department of Psychiatry, Psychiatry and Inflammation Program, Tufts University School of Medicine, Boston, MA, USA; [12]Association for the Improvement of Mental Health Programmes, Geneva, Switzerland

1. Fiorillo A, Sartorius N. Ann Gen Psychiatry 2021; 20:52.
2. Correll CU, Solmi M, Croatto G et al. World Psychiatry 2022;21:248-71.
3. Firth J, Solmi M, Wootton RE et al. World Psychiatry 2020;19:360-80.
4. Taipale H, Tanskanen A, Mehtälä J et al. World Psychiatry 2020;19:61-8.
5. Sartorius N. Acta Psychiatr Scand 2018;137:369-70.
6. Maj M, van Os J, De Hert M et al. World Psychiatry 2021;20:4-33.
7. Stein DJ, Craske MJ, Rothbaum BO et al. World Psychiatry 2021;20:336-56.
8. McIntyre RS, Alda M, Baldessarini RJ et al. World Psychiatry 2022;21:364-87.
9. Fiorillo A, Sampogna G, Elkholy H et al. World Psychiatry 2021;20:149-50.
10. Javed A. World Psychiatry 2021;20:146.
11. Javed A. World Psychiatry 2021;20:451-2.
12. Javed A. World Psychiatry 2022;21:325-6.

DOI:10.1002/wps.21055

Pushing forward public mental health agenda and promotion of mental health

Major activities are needed to transform psychiatric and mental health care as well as public mental health to deliver on the United Nations (UN) Sustainable Development Goals (SDGs)[1]. We must orient our services towards sustainable and inclusive prevention, early intervention, treatment, care and rehabilitation, as well as manage social changes and threats while fostering transparency and continuity.

As the largest and most important psychiatric organization, with 145 national psychiatric associations from 121 countries around the entire globe and more than 250,000 members, the WPA has a decisive role to play in this process[2].

The WPA Planning Committee has identified key priorities for the incoming presidency starting in October 2023 during the World Congress of Psychiatry in Vienna. We are committed to focusing on educational/informational activities directed to psychiatrists, the public, patients and their families, other mental health professionals, and undergraduate and postgraduate students. More importantly, the issue of equal access to mental health care for all should be paramount.

To enhance public mental health and well-being as highlighted in Goal 3 (Good health and well-being) of the SDGs, we have prioritized focusing on equal access to psychiatric, mental health, and public mental health services according to the following ranking of the SDGs: Goal 10 (Reduce inequality), Goal 5 (Gender equality), Goal 4 (Quality education), and Goal 17 (Partnerships to achieve the goal).

The gaps uncovered by the WPA survey on educational activities[3,4] will be prioritized. The ambition is to produce and deliver, in the future, educational/informational materials in the six official languages of the World Health Organization (WHO) (Arabic, Chinese, English, French, Spanish and Russian) and hopefully even more.

Moreover, to improve the mental health of citizens, and of psychiatric in- and outpatients as well as the psychiatric team that treats them, we need to focus on healthy lifestyles such as physical activity[5], eating habits[6], behavioural changes[7], intellectual stimuli[8], workplace satisfaction[9], and sleep hygiene[10], all of which are critical for the improvement of mental health and prevention of poor mental health.

Psychiatry has many excellent evidence-based methods for pharmacological and psychotherapeutic treatments. Several universal and selective preventive interventions are feasible and cost-effective, and have shown to prevent poor mental health[11]. However, the role of healthy lifestyles and behavioural changes to improve mental health is under-prioritized. The Planning Committee believes that pedagogically tailored lifestyle activities will add value to the existing biological and psychological therapies when used daily in psychiatric care.

Many individuals suffering from mental health problems have never been exposed or provided with good examples at home or in school on how to choose healthy lifestyles: for example, how to purchase healthy foods, how to plan shopping lists, how to cook and what to do with leftovers. Many also lack information on how to achieve and maintain good sleep and how to plan and use their time depending on the different seasons. This type of important information should align with the traditions and social and economic realities of the individuals involved.

Physical activities, even when carried out in small amounts but done for some minutes daily in the morning, noon and evening, have a positive impact on health[5]. For psychiatric patients, physical activities performed in groups with psychiatric staff or family or community will not only influence well-being and health, but also the feelings of equality, cohesion, collaboration and mutual understanding, and the sense of belongingness, hopefully diminishing the stigma of mental disorders.

The lifestyle activities performed together with patients will most likely assist the physical fitness and healthy lifestyles of psychiatrists and other staff. Psychiatrists also need to take care of their own somatic and mental health and working conditions[12]. There are plans to produce short videos on each of the different lifestyle activities intended to be used in daily psychiatric practice, in conjunction with collection of good examples from the WPA Member Societies.

Awareness of the influence of environment and art on mental health should also be increased and incorporated into patients' activities[13,14]. The aforesaid good examples – such as having patients take care of flowers and plants in the wards; involving them in gardening and choosing decorative art – will be collected from the WPA Member Societies and disseminated through the WPA channels.

All interventions for improving healthy lifestyles should be scientifically evaluated. There are plans for the WPA to provide scientific guidance on methodology for developing cross-sectional, cohort, and case-control studies as well as randomized controlled trials measuring treatment-related satisfaction and improvement of general and mental health, including psychological, social and biological outcomes[11].

To achieve these goals, we will need advice and increased collaboration with national psychiatric and other medical societies, such as the World Medical Association, the World Pediatric Society, the International Federation of Medical Students Associations, the International Society for Physical Activity and Health, the World Organization of Family Doctors (WONCA), and other mental health associations. This also includes continuous collaboration with the WHO, and patient and family associations.

We acknowledge that ongoing activities introduced by former presidents, executive committees, and other committees of the WPA deserve continuous strong support. Of special interest are the activities pursued by the Advisory Committee for Responses to Emergencies (ACRE)[15] at a time when there are multiple devastating and life-threatening armed conflicts and wars as well as serious environmentally induced humanitarian and health crises that are impacting the mental health of the entire world population.

Danuta Wasserman[1,2], Celso Arango[3], Andrea Fiorillo[4], Saul Levin[5], Andrew C. Peters[6], G. Prasad Rao[7], Aida Sylla[8], Thelma Sanchez-Villanueva[9]

[1]WPA President Elect; [2]National Centre for Suicide Research and Prevention of Mental Ill-Health, Karolinska Institutet, Stockholm, Sweden; [3]Department of Child and Adolescent Psychiatry, Universidad Complutense, Madrid, Spain; [4]Department of Psychiatry, University of Campania, Naples, Italy; [5]Medical Director, American Psychiatric Association; [6]Chief Executive Officer, Royal Australian and New Zealand College of Psychiatrists; [7]Asha Hospital, Hyderabad, India; [8]University Cheikh Anta Diop, Dakar, Senegal; [9]Universidad de Guadalajara, Guadalajara, Mexico

1. United Nations. Transforming our world: the 2030 agenda for sustainable development. New York: United Nations, 2015.
2. Wasserman D. World Psychiatry 2021;20:309-10.
3. Ng RK. World Psychiatry 2022;21:478-9.
4. Sampogna G, Fiorillo A, Wasserman D. Asia-Pacific Psychiatry 2021;13:e12497.
5. Wasserman D. Acta Paediatr 2019;108:984-5.
6. Zahedi H, Djalalinia S, Sadeghi O et al. Nutr Neurosci 2022;25:1250-64.
7. Durkee T, Carli V, Floderus B et al. Int J Environ Res Public Health 2016;13:294.
8. Kasahara-Kiritani M, Hadlaczky G, Westerlund M et al. Int J Environ Res Public Health 2015;12:15937-42.
9. Sisask M, Värnik P, Värnik A et al. Health Educ J 2014;73:382-93.
10. Sarchiapone M, Mandelli L, Carli V et al. Sleep Med 2014;15:248-54.
11. Fusar-Poli P, Correll CU, Arango C. World Psychiatry 2021;20:200-21.
12. Gerada C. Beneath the white coat: doctors, their minds and mental health. London: Routledge, 2020.
13. Berry HL. World Psychiatry 2021;20:152-3.
14. Theorell T. Front Psychol 2021;12:742032.
15. Javed A. World Psychiatry 2022;21:325-6.

DOI:10.1002/wps.21052

New research on validity and clinical utility of ICD-11 vs. ICD-10 and DSM-5 diagnostic categories

A recent paper in *World Psychiatry*[1] summarized the recent literature on the validity and clinical utility of four new categories introduced in the ICD-11 chapter on mental disorders: complex post-traumatic stress disorder, prolonged grief disorder, gaming disorder, and compulsive sexual behaviour disorder. The reviewed evidence suggests that the new categories describe populations with clinically important features that were previously not recognized in the ICD classification, and that these populations have specific treatment needs that would go unmet if the new disorders are not included in the classification. Moreover, the addition of the new categories has had a positive impact in terms of health reporting as well as development and testing of new interventions[e.g., 2-4].

In the past two years, there have been further studies focused on other ICD-11 categories, testing their validity, clinical utility and/or interrater reliability in comparison with the corresponding categories in the

ICD-10 and/or the DSM-5[5]. Of special interest are four of these studies, dealing respectively with: a) the accuracy in diagnosing mood disorders depicted in case vignettes using ICD-11 vs. ICD-10 clinical descriptions and diagnostic guidelines[6]; b) the interrater reliability, concurrent validity, and clinical utility of the behavioural indicators introduced in the ICD-11 in order to improve the identification and treatment of individuals with disorders of intellectual development[7]; c) the sensitivity, specificity, and ability to predict the use of gender-affirming medical procedures of categories related to gender identity in the ICD-11 vs. DSM-5[8]; d) the clinical utility of the formulation of irritability and oppositionality in youth which has been proposed by the ICD-11 compared with the corresponding ICD-10 and DSM-5 models[9].

The first of the above-mentioned studies[6] reported that the use of ICD-11 guidelines, as compared with ICD-10 ones, allowed a more accurate detection of depressive episodes within the context of recurrent depressive disorder; led to lower rates of applying mood disorder diagnoses when none was warranted; and was associated with a less frequent misdiagnosis of depressive episodes as mixed depressive and anxiety disorder, or as prolonged grief disorder. However, some difficulties were found when differentiating between the ICD-11 categories of bipolar type I vs. type II disorder (a distinction not present in the ICD-10), and a poorer accuracy was observed when applying specifiers of severity of depression using the ICD-11 compared with the ICD-10 (a finding which has led to a revision of the ICD-11 severity specifiers for depressive episode).

The study focusing on behavioural indicators for disorders of intellectual development[7] found that these indicators had excellent interrater reliability (intra-class correlations between 0.91 and 0.97) and good to excellent concurrent validity (intra-class correlations between 0.66 and 0.82) across the four sites where the study was conducted. Furthermore, these indicators were rated as quick and easy to use and applicable across levels of severity; and as useful for treatment selection, prognosis assessment, communication between health care professionals, and education efforts. Finally, the indicators showed more diagnostic overlap between intellectual and adaptive functioning compared to standardized measures.

The study on the validity of categories related to gender identity[8] found that the sensitivity of the diagnostic requirements was equivalent in the ICD-11 (where these categories are not included in the chapter on mental disorders) and the DSM-5, but that the inclusion of the diagnostic requirements for distress and/or dysfunction in the DSM-5 is associated with a lower predictive power with respect to the use of gender-affirming medical procedures (i.e., history of hormone use and/or surgery). Furthermore, the ICD-11 diagnostic formulation was found to be more parsimonious and to contain more information about caseness than the DSM-5 model.

The Internet-based field study on diagnostic classification of irritability and oppositionality in youth[9], conducted with 196 clinicians from 48 countries, found that the formulation proposed in the ICD-11 (using chronic irritability as a qualifier for the diagnosis of oppositional defiant disorder) led to a more accurate identification of severe irritability and a better differentiation from boundary presentations compared to both the DSM-5 model (introducing the new category of disruptive mood dysregulation disorder) and the ICD-10 classification (listing oppositional defiant disorder as one of several conduct disorders without attention to irritability). Participants using the DSM-5 often failed to apply the diagnosis of disruptive mood dysregulation disorder when it was appropriate, and more frequently applied psychopathological diagnoses to irritability that was developmentally normative.

Further studies based on the use of case vignettes in samples recruited from the World Health Organization (WHO) Global Clinical Practice Network – now including more than 17,800 clinicians from more than 160 countries (https://gcp.network) – are now ongoing. These studies, along with other investigations conducted in clinical settings and with the experience in the use of the ICD-11 worldwide[10-16], will guide in the next few years possible refinements of the ICD-11 guidelines.

Matteo Di Vincenzo
WHO Collaborating Centre for Research and Training in Mental Health, Naples, Italy

1. Reed GM, First MB, Billieux J et al. World Psychiatry 2022;21:189-213.
2. Prigerson HG, Boelen PA, Xu J et al. World Psychiatry 2021;20:96-106.
3. Mauro C, Tumasian RA 3rd, Skritskaya N et al. World Psychiatry 2022;21:318-9.
4. Lew-Starowicz M, Draps M, Kowalewska E et al. World Psychiatry 2022;21:468-9.
5. First MB, Gaebel W, Maj M et al. World Psychiatry 2021;20:34-51.
6. Kogan CS, Maj M, Rebello TJ et al. J Affect Disord 2021;295:1138-50.
7. Lemay KR, Kogan CS, Rebello TJ et al. J Intellect Disabil Res 2022;66:376-91.
8. Robles R, Keeley JW, Vega-Ramirez H et al. Int J Clin Health Psychol 2022;22:100281.
9. Evans SC, Roberts MC, Keeley JW et al. J Child Psychol Psychiatry 2021;62:303-12.
10. Pinto da Costa M, Ng RMK, Reed GM. World Psychiatry 2021;20:148-9.
11. Cloitre M, Brewin CR, Bisson H et al. Eur J Psychotraumatol 2020;11:1739873.
12. Zisook S, Mohamad S, Johnson G et al. World Psychiatry 2021;20:303-4.
13. Evans SC, Weisz JR, Carvalho AC et al. J Consult Clin Psychol 2020;88:255-68.
14. Khoury B, El Khoury J, Fresan A et al. Cult Health Sex 2020;23:131-42.
15. Fabrazzo M. World Psychiatry 2022;21:163-4.
16. Pezzella P. World Psychiatry 2022;21:331-2.

DOI:10.1002/wps.21053

Acknowledgement
This publication has been partially supported by an unrestricted educational grant from Lundbeck Italia SpA, which is hereby gratefully acknowledged.