World Psychiatry

OFFICIAL JOURNAL OF THE WORLD PSYCHIATRIC ASSOCIATION (WPA)

Volume 17, Number 3 *October 2018*

WILEY

NEW IMPACT FACTOR: 30.000

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 140, spanning 120 different countries and representing more than 200,000 psychiatrists.

The WPA organizes the World Congress of Psychiatry every three years. It also organizes international and regional congresses and meetings, and thematic conferences. It has 72 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 Executive Committee
President – H. Herrman (Australia)
President-Elect – A. Javed (UK/Pakistan)
Secretary General – R.A. Kallivayalil (India)
Secretary for Finances – A. Soghoyan (Armenia)
Secretary for Meetings – M. Takeda (Japan)
Secretary for Education – R. Ng (Hong Kong-China)
Secretary for Publications – M. Botbol (France)
Secretary for Sections – T.G. Schulze (Germany)

WPA Secretariat
Geneva University Psychiatric Hospital, 2 Chemin du Petit Bel-Air, 1225 Chêne-Bourg, 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Research Reports containing unpublished data are welcome for submission to the journal.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.

Editor – M. Maj (Italy).
Editorial Board – H. Herrman (Australia), A. Javed (UK/Pakistan), R.A. Kallivayalil (India), A. Soghoyan (Armenia), M. Takeda (Japan), R. Ng (Hong Kong-China), M. Botbol (France), T.G. Schulze (Germany).
Advisory Board – H.S. Akiskal (USA), R.D. Alarcón (USA), D. Bhugra (UK), J.A. Costa e Silva (Brazil), J. Cox (UK), M. Jorge (Brazil), H. Katschnig (Austria), F. Lieh-Mak (Hong Kong-China), F. Lolas (Chile), J.E. Mezzich (USA), D. Moussaoui (Morocco), P. Munk-Jorgensen (Denmark), F. Njenga (Kenya), A. Okasha (Egypt), J. Parnas (Denmark), V. Patel (India), P. Ruiz (USA), N. Sartorius (Switzerland), A. Tasman (USA), S. Tyano (Israel), J. Zohar (Israel).

Office of the Editor – Department of Psychiatry, University of Naples SUN, Largo Madonna delle Grazie, 80138 Naples, Italy. Phone: +390815666502; Fax: +390815666523; E-mail: majmario@tin.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.pubmedcentral.nih.gov/tocrender.fcgi?journal=297&action=archive).

Classification of psychopathology: conceptual and historical background

In their paper in this issue of the journal, Krueger et al[1] consider two different approaches to the classification of psychopathology. Here I would like to focus on the conceptual and historical background of these approaches.

What the authors call "authoritative" nosology – represented currently in the US by the DSM-5 system[2] – evolved from classificatory efforts starting in the late 17th century, when large numbers of patients began to be collected in asylums in Central and Western Europe. These efforts were based on earlier attempts to classify general medical conditions, which were in turn heavily influenced by systems that classified animal and plant species as part of the beginning of zoology and botany as descriptive sciences[3]. So, I agree with the authors that the DSM is an historically influenced document, but I see this more positively than they do.

Medicine has worked by a gradual evolutionary process of the articulation of broad syndromes, many of which, with advancing knowledge, become divided up into more homogenous entities that then develop into what we might call "disorders" and eventually "diseases". In psychiatry, this process has been slower and more difficult than in most areas of medicine, but still represents an accumulated wisdom that typically works pretty well in the real world of patient care. How well it serves the goals of research is another matter.

As this brief history suggests, categories are inextricably intertwined with the world of clinical medicine. Individuals in care need to be given diagnoses because of the key dichotomies that exist in this world – to treat or not to treat, to discharge (from an emergency room) or hospitalize, to qualify for a particular treatment algorithm or not, to bill or not and, if to bill, with what specific code. This does not, of course, preclude quantitative measures, the focus of the nosologic approach advocated by the authors. These too are woven into the fabric of medicine. Think of temperature, pulse rates, fasting blood sugar, white blood cell counts and bone densities. These measures happily co-exist with the diagnostic world and are used nearly universally to monitor health and illness and guide therapy.

I worry that underneath this debate about continua versus categories there is a confusion between the "levels" of underlying physiology/etiology and clinical manifestation. Let me illustrate this by a "thought experiment":

> A steep south-facing slope in the high mountains received a heavy snowfall. The next morning dawns warm with a clear sky and strong sun. The temperature – a classical quantitative variable – at the lower levels of the snow pack starts to rise and melting increases gradually throughout the morning. Suddenly, in mid-afternoon, the snow pack starts to slide, ending in a dramatic avalanche.

This example illustrates a natural quantitative process – snow melting with increasing temperature – and a dramatic threshold effect. If you work for the ski patrol to prevent avalanches, you need to understand both processes.

Turning to medical applications, consider a femur with increasing levels of strain – a quantitative trait. At some point, the bone breaks with dramatic health consequences. Think of a coronary artery with increasing occlusions as cholesterol plaques increase. At some point, the blood flow and associated delivery of oxygen slips below a critical level. Heart tissue starts to die and a myocardial infarct occurs.

I agree that taxometric methods provide at most modest evidence for discrete diagnostic categories in psychiatry. But I want to add to this discussion a different and informative perspective – within individual analyses. Like when seeing an avalanche, when seeing an acute patient presenting in the emergency room with a broken femur or an active myocardial infarction, it is difficult to conclude that one should only be concerned with the underlying quantitative process. Something clinically dramatic and "categorical" has occurred that calls for immediate intervention. Consider the following brief psychiatric vignettes:

> A vulnerable individual, who stopped his antipsychotic medication four weeks ago, over 48 hours transitions from a non-psychotic state to a full-blown psychosis characterized by active auditory hallucinations and persecutory delusions about which he is quite preoccupied.

> An individual with prior bipolar illness in good remission, after traveling across five time zones and experiencing several nights of poor sleep, the next day, "flips" into a fully syndromal mania.

> You observe a friend with panic disorder in a crowded restaurant go from a calm, collected state in less than a minute to one of acute distress with sweating, panting, shaking and fear of dying.

While not all psychiatric disorders have such dramatic "avalanche-like" transitions, they are fairly common in clinical psychiatry and challenge the authors' conclusions that there is little viable evidence that psychiatric disorders need to be understood from a categorical perspective.

Let me turn to a quite different issue. I was concerned by the manner in which the authors characterize the DSM pro-

cess: "group discussions and associated political processes", manifesting "sociopolitical dynamics", issuing *ex cathedra* decisions with the final diagnoses resulting from "presumed authority and fiat". This tone will not aid interdisciplinary discourse. The authors imply that they are the objective scientists while those who worked on DSM are, by comparison, bogged down in political discourse and constrained by old-fashioned historical dictates. While this is not the place to discuss this in detail, any organized effort in science to develop classifications involves "sociopolitical dynamics". Readers who think otherwise might consult a history of the decision of the International Astronomical Union to remove Pluto from the official list of planets[4].

I want to conclude by talking about standards of diagnostic validation. At the risk of over-simplification, the Hierarchical Taxonomy of Psychopathology (HiTOP) program emphasizes psychometric methods in its typological proposals. Such methods have been key in the history of psychology, for example in the development of personality typologies and measures of various cognitive skills. So, it is sensible that they should be applied in the area of psychopathology. However, this approach differs considerably from the medical tradition emphasized by DSM. Put simply, the medical tradition wants diagnoses that tell us a lot about the patient – the course, the likely etiologic process, the best treatment, etc.. We organize our literature around our diagnoses, from cohort studies to randomized controlled trials.

The specific articulation of this viewpoint in psychiatry was first given by Robins and Guze[5] with their list of validators, substantially expanded since then. Since DSM-III, the role of the evaluation of validators in diagnostic change has, albeit somewhat unevenly, gradually increased. The main approach has been the use of literature reviews trying to summarize available information on validators. These questions were the specific focus of the Scientific Review Committee that evaluated every proposed diagnostic change in DSM-5[6]. The procedures developed for change in DSM-5 by the American Psychiatric Association's Steering Committee are empirically rigorous and data driven[7].

It is not surprising that the scientific disciplines of psychiatry and clinical psychology have developed different approaches to the creation and evaluation of diagnostic entities/dimensions. Optimal communication between these two disciplines, however, requires an understanding of the similarities and differences in these approaches, the relative strengths and limitations of each approach, and the acceptance by both sides that each is likely to be able to contribute meaningfully to the difficult challenge of designing an optimal psychiatric classification.

Kenneth S. Kendler
Virginia Institute for Psychiatric and Behavioral Genetics and Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA

1. Krueger RF, Kotov R, Watson D et al. World Psychiatry 2018;17:282-93.
2. American Psychiatric Association. Diagnostic and statistical manual of mental disorders, 5th ed. Arlington: American Psychiatric Association, 2013.
3. Kendler KS. Psychol Med 2009;39:1935-41.
4. Zachar P, Kendler KS. Philos Ethics Humanit Med 2012;7:4.
5. Robins E, Guze SB. Am J Psychiatry 1970;126:983-7.
6. Kendler KS. Psychol Med 2013;43:1793-800.
7. First MB, Kendler KS, Leibenluft E. JAMA Psychiatry 2017;74:115-6.

DOI:10.1002/wps.20549

Intergenerational transmission of trauma effects: putative role of epigenetic mechanisms

Rachel Yehuda, Amy Lehrner

James J. Peters Bronx Veterans Affairs Hospital, Bronx, NY, USA; Departments of Psychiatry and Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA

This paper reviews the research evidence concerning the intergenerational transmission of trauma effects and the possible role of epigenetic mechanisms in this transmission. Two broad categories of epigenetically mediated effects are highlighted. The first involves developmentally programmed effects. These can result from the influence of the offspring's early environmental exposures, including postnatal maternal care as well as in utero *exposure reflecting maternal stress during pregnancy. The second includes epigenetic changes associated with a preconception trauma in parents that may affect the germline, and impact fetoplacental interactions. Several factors, such as sex-specific epigenetic effects following trauma exposure and parental developmental stage at the time of exposure, explain different effects of maternal and paternal trauma. The most compelling work to date has been done in animal models, where the opportunity for controlled designs enables clear interpretations of transmissible effects. Given the paucity of human studies and the methodological challenges in conducting such studies, it is not possible to attribute intergenerational effects in humans to a single set of biological or other determinants at this time. Elucidating the role of epigenetic mechanisms in intergenerational effects through prospective, multi-generational studies may ultimately yield a cogent understanding of how individual, cultural and societal experiences permeate our biology.*

Key words: Intergenerational transmission, epigenetic mechanisms, trauma, offspring of trauma survivors, childhood adversity, post-traumatic stress disorder, developmental programming, fetoplacental interaction

(World Psychiatry 2018;17:243–257)

There is now converging evidence supporting the idea that offspring are affected by parental trauma exposures occurring before their birth, and possibly even prior to their conception.

On the simplest level, the concept of intergenerational trauma acknowledges that exposure to extremely adverse events impacts individuals to such a great extent that their offspring find themselves grappling with their parents' post-traumatic state. A more recent and provocative claim is that the experience of trauma – or more accurately the effect of that experience – is "passed" somehow from one generation to the next through non-genomic, possibly epigenetic mechanisms affecting DNA function or gene transcription[1-6].

Although both intergenerational (from F0 to F1) and transgenerational (from F0 to F3 or F4) transmission of environmental adversity effects have been established in animal models, studies in humans have not yet demonstrated that the effects of trauma are heritable through non-genomic (i.e., epigenetic) mechanisms. Nonetheless, there has been much excitement about, and even premature promulgation of, the idea that those effects are transmitted through DNA modifications, explaining the impact of familial experience[7].

The inclination to attribute offspring effects to epigenetic mechanisms in part reflects the inexact and varied use of the term "transmission". The original use was descriptive, and without mechanistic inferences. Now that animal research has defined a molecular pathway through which transmission of trauma effects might occur, more precise language is warranted to distinguish between clinical observation and biological mechanism. At the current time, the idea that epigenetic mechanisms underlie clinical observations in offspring of trauma survivors represents a hypothesis to be tested.

This review delineates potential epigenetic mechanisms that might be examined in relation to offspring effects, and provides insight into the type of studies that might be most informative.

THE ORIGIN OF STUDIES OF INTERGENERATIONAL TRAUMA EFFECTS

The concept of intergenerational trauma was introduced in the psychiatric literature through descriptions of behavioral and clinical problems in offspring of Holocaust survivors[8].

In a pivotal paper describing three patients who presented for psychiatric treatment, Rakoff[8] wrote: "The parents are not broken conspicuously, yet their children, all of whom were born after the Holocaust, display severe psychiatric symptomatology. It would almost be easier to believe that they, rather than their parents, had suffered the corrupting, searing hell".

This initial report generated mostly negative reactions, including caution about generalizing from what might have been idiosyncratic observations in a small number of extreme cases[9]. Some stakeholders may have felt that the suggestion that surviving the trauma of genocide had deleterious implications for progeny was stigmatizing in the face of the emerging cultural narrative regarding the Holocaust, which was one of survival against all odds, resilience, and defiance in the hope of preventing such occurrences in the future[10].

The initial anecdotal report, and the reactions to it, generated much empirical research on the question of whether and how Holocaust offspring, conceived and born after World War II, were affected. Hundreds of articles appeared, beginning in

the 1970s and continuing for some decades thereafter. The studies described in these reports either failed to find effects in Holocaust offspring, corroborated earlier clinical descriptions, attempted to restrict the observations of damaging effects to a subgroup, or pointed to serious methodological challenges in attempting to address this question empirically[11-13].

A wide range of phenomena was described in studies reporting behavioral difficulties in Holocaust offspring. These included: feelings of over-identification and fused identity with parents, impaired self-esteem stemming from minimization of offspring's own life experiences in comparison to the parental trauma, tendency towards catastrophizing, worry that parental traumas would be repeated, and behavioral disturbances such as experiencing anxiety, traumatic nightmares, dysphoria, guilt, hypervigilance and difficulties in interpersonal functioning. Such studies often did not account for parental psychopathology, but assumed it on the basis of parental exposure.

Similar types of symptoms were later described in the children of Vietnam veterans[14,15], a phenomenon that was termed "secondary traumatization"[16]. This concept did not imply an intergenerational transmission, but rather referred to the stressful nature of living with a traumatized individual who may be expressing symptoms and recounting or reliving horrific experiences[17].

In the absence of biological mechanisms to explain the reported findings, explanations were almost exclusively psychodynamic or behavioral. For example, it was suggested that trauma survivors externalized their post-traumatic symptoms through their nonverbal behaviors and unconscious reenactments of fear and grief, such that the child became a container for the unwanted, troubling experiences of the parent[18,19].

Distinctions between "transmission" from parent to child in which the disturbance in the child was a direct consequence of a psychiatric condition in the parent versus an effect reflecting the child's reaction to symptoms in parents[11,20] were made carefully in order to avoid misattributing offspring effects to earlier parental trauma exposures. Other perspectives – including family dynamics, attachment theory, social psychology and learning theory – were also brought to bear[11,21-24].

One the most provocative observations regarding Holocaust offspring was the report that Yom Kippur war veterans were more likely to develop post-traumatic stress disorder (PTSD) in response to combat if they had a Holocaust survivor parent[25]. A higher prevalence of PTSD, mood and anxiety disorders was also observed in Holocaust offspring, largely selected from a convenience sample of people seeking treatment for Holocaust-related problems, compared with controls[26]. These findings were replicated in a study assessing the relationship between PTSD in offspring and their own parents, assessed directly by clinical interview of the parent (s)[27].

The increased prevalence of PTSD in Holocaust offspring in response to their own traumatic exposures was later found to be associated with maternal PTSD in Holocaust survivors[28]. Although PTSD was found to occur in association with paternal PTSD in a study of Australian Vietnam Veterans and their offspring[29], the contribution of potential maternal symptoms, even through secondary traumatization, was not assessed. It is rare to identify a cohort in whom both mothers and fathers had similar exposures to an extreme trauma, or even a cohort in whom the impact of lifetime trauma was evaluated in both parents, and even rarer to have the opportunity to evaluate psychiatric morbidity in both parents and adult children.

While some aspects of intergenerational trauma effects remain contested, discussions about whether there are clinically observable intergenerational effects in offspring have become less contentious in the last several years, with the increasing recognition of the universality of this phenomenon.

Presently, there are discussions about the impact of historical events such as colonization, slavery and displacement trauma in many cultures, including First Nations and native American communities[30,31], African Americans[32,33], Australian aboriginals and New Zealand Maori[34,35], as well as in societies exposed to genocide, ethnic cleansing or war, such as Cambodians[36,37], Armenians[38,39], Rwandans[40,41], Palestinians[42], and communities in the former Yugoslavia[43]. There is also a growing literature about offspring effects following early maternal childhood maltreatment[44-47].

The intense focus on intergenerational effects in these different groups suggests that this topic has broad resonance and global applicability, and provides a mandate for increased attention to this area, including prospective, longitudinal studies that can be designed in the future to determine the mechanisms underlying this phenomenon.

THE INTRODUCTION OF BIOLOGICAL RESEARCH INTO THE STUDY OF INTERGENERATIONAL EFFECTS OF TRAUMA

Research addressing putative biological correlates of intergenerational effects began in the late 1990s[48]. The findings of an increased prevalence of PTSD among offspring with parental PTSD[25,27] raised the possibility that Holocaust offspring might have specific biological risk factors for PTSD and/or other trauma-associated mood and anxiety disorders, particularly following their own traumatic exposures. The introduction of biology into the debate about intergenerational trauma was a natural outcome of developments in the emerging field of the neurobiology of PTSD, that was beginning to clarify similar issues about the nature and long-term impact of trauma exposure[49].

The initial focus of these studies was on the hypothalamic-pituitary-adrenal (HPA) axis, for several reasons. First, the HPA axis is vulnerable to environmental perturbations. The initial hypothesis with respect to Holocaust offspring was that parental experiences might alter the regulation of stress-related pathways early in development. This idea was plausible, since the HPA axis is subject to early developmental programming[50,51]. Furthermore, dysregulation of stress neurocircuitry is a fundamental feature of mood and anxiety disorders[52-54],

including PTSD, found to be prevalent in offspring. Finally, there had been directionally interesting findings of low cortisol and increased glucocorticoid receptor (GR) sensitivity in Holocaust survivors and other trauma exposed individuals with PTSD, suggesting that the experience of trauma might leave long-lasting biological signatures in stress-related biology that could be a catalyst for longer-term adaptations[55].

As this work developed, advances in molecular biology, including an understanding of gene-environment interactions and the contribution of environmentally-induced changes in epigenetic regulation of HPA-related genes, provided the tools for examining how salient events could result in enduring, transformative, and possibly even inherited change, laying the groundwork for future molecular studies[56-59].

Studies published over the next decade demonstrated that, in the absence of their own traumatic exposures, offspring of Holocaust survivors were more likely to show HPA axis alterations associated with PTSD, such as lower cortisol levels and enhanced GR responsiveness[60-64]. Observations in offspring whose parents were exposed to other traumatic experiences accorded with these findings. For example, lower cortisol levels were observed in the adult offspring of combat veterans with PTSD compared to offspring of combat veterans without PTSD[65].

Subsequent investigations documented that maternal and paternal PTSD were associated with different biological outcomes. A *post-hoc* analysis of cortisol circadian rhythm data indicated that lower cortisol levels in adult Holocaust offspring were associated with maternal, but not paternal, PTSD[61]. In another study, several measures of GR sensitivity were found to be directionally different in offspring of mothers vs. fathers with PTSD[63]. Specifically, maternal PTSD was associated with lower urinary cortisol levels as well as greater GR sensitivity as measured by the lysozyme inhibition test (an *in vitro* measure of that sensitivity in peripheral tissue) as well as the dexamethasone suppression test (DST). An interaction of maternal and paternal PTSD on urinary cortisol and the DST demonstrated a decreased glucocorticoid sensitivity in offspring with paternal, but not maternal, PTSD.

Initial theories posited that offspring biological effects were reflections of their own experiences as a result of having traumatized parents who may have been symptomatic, neglectful, or otherwise impaired in parenting[11,21-25]. Differences in offspring effects based on parental gender could similarly be viewed through the lens that mothers and fathers might be associated with different types of parenting roles and behaviors. Thus, in essence, having a traumatized mother, father, or both constituted an early environmental experience that impacted the offspring. Supporting this idea were findings that Holocaust offspring reported higher levels of childhood trauma exposure than demographically similar comparison subjects, particularly if one or more parent had PTSD[66]. In fact, the low cortisol in offspring was found to be associated with offspring reports of emotional abuse[66]. By then it had been established that early childhood maltreatment in itself could result in lower cortisol levels[67-71].

Investigations of younger offspring of mothers who had themselves experienced abuse as children also demonstrated effects on cortisol levels. In one study, cortisol levels were found to be lower in the offspring of mothers with childhood maltreatment as well as bipolar disorder[72]. Lower cortisol and blunted cortisol reactivity were present in preadolescent boys and girls with maternal PTSD, even after controlling for youth traumatic event history and mental health symptoms[73]. A blunted cortisol reactivity to stress was observed in even younger offspring, toddlers aged 12-48 months, in association with maternal PTSD occurring as a result of interpersonal violence[74]. Infants of women exposed to maternal child abuse also displayed lower baseline cortisol when examined at 6 months of age[44].

Investigators also examined markers other than HPA axis parameters. One study reported that children of mothers exposed to childhood trauma, particularly emotional abuse, had higher sympathetic nervous system activation, which might be a marker for vulnerability to anxiety, compared to children of mothers with low emotional abuse, an effect that remained significant after accounting for maternal PTSD and depression, and for child trauma exposure[45]. In another study, maternal exposure to child abuse was associated with smaller intracranial volume, due to differences in cortical gray matter, in newborns examined within two weeks of birth[75]. This effect was reported to be independent of some potential confounding variables, such as maternal socio-economic status, obstetric complications, obesity, recent interpersonal violence, pre- and early postpartum stress, gestational age at birth, infant sex, and postnatal age at magnetic resonance imaging scan.

As studies begin to examine offspring prospectively, starting in close proximity to their birth, it will be easier to identify the relative contributions of preconception, *in utero*, and postnatal influences on offspring[76]. Indeed, part of the difficulty in studying adult offspring of trauma survivors, particularly retrospectively, is that it is difficult to make attributions about the origin of any observed biological manifestation. Such explorations must also invariably include the contribution of genotype, as it is becoming increasingly recognized that at least some "programmed" epigenetic modifications may be established through gene x environment effects[5,7]. Indeed, such interactions may help explain diversity in offspring responses to parental trauma effects.

POTENTIAL MECHANISMS FOR OBSERVED BIOLOGICAL EFFECTS IN OFFSPRING OF TRAUMA SURVIVORS

The first basic science approach to understanding offspring effects was the work of Meaney et al[77,78], beginning in the late 1980s. This team of researchers initially focused on long-term

effects of early handling of rat pups, using a model in which mothers were separated from their neonatal pups for several minutes each day. In adulthood, the handled rats had altered basal and stress-induced corticosterone levels as well as higher GR sensitivity on the low-dose DST and greater GR number in the hippocampus[77-79].

However, it subsequently became clear that the observed effects in offspring were mediated not by the maternal separation or the early handling by humans, but rather by the behavior of the mother upon being reunited with her pups in the home cage, specifically the extent of licking and grooming of pups. The offspring of mothers that displayed lower vs. higher licking and grooming demonstrated distinct neuroendocrine and behavioral parameters, which persisted from F1 to F2[80,81].

This clear example of developmental programming, in which postnatal exposures in the pups (i.e., variations in maternal licking and grooming behavior) induced enduring changes in behavior and HPA axis responsiveness, seemed relevant to the offspring of trauma survivors[82]. Interestingly, the neuroendocrine phenotype of Holocaust offspring with maternal PTSD was more consistent with maternal overprotection than neglect, as low cortisol levels in offspring were found to be associated with overprotection[83]. Maternal overprotection subsequent to stress exposure was also reported in association with low cortisol/dehydroepiandrosterone (DHEA) ratio in offspring[84].

In 2002, a seminal paper demonstrated that the rat offspring effects of licking and grooming were associated with an epigenetic change, namely, DNA methylation at a GR (nr3c1) gene promoter in the hippocampus[85,86]. Later work expanded this finding from epigenetic marks at a single gene promoter on one gene to clustered epigenetic changes in promoters associated with transcriptional activity across broad genomic areas[87-89]. The effects in adulthood were determined to be directly related to the early postnatal environmental exposures to variations in maternal care, since they were prevented by cross-fostering neonatal rat pups to mothers displaying different behavioral characteristics[81,86,90,91]. The elimination of offspring effects through cross-fostering is a potent example of social transfer of information through parental behavior – not parental DNA or biological inheritance. Yet these findings constituted a powerful example of how early environmental inputs and parental behavior could affect offspring DNA methylation, behavior, and the function of neuroendocrine stress responsiveness for more than one generation.

It is hard to overstate the level of excitement generated by the findings demonstrating an epigenetic alteration in brain in response to variations in postnatal maternal care. Though epigenetic mechanisms and their central role in development had been known since the 1940s, following C. Waddington's initial descriptions of these molecular mechanisms[92], these concepts had not been previously applied as explanations for how environmental exposures – such as parental behaviors – could reprogram stress hormone biology, affecting brain and behavior of progeny[93,94].

This elegant series of studies provided a clear molecular link between maternal behavior and gene function in offspring, mediated by epigenetic mechanisms, and producing functional biological correlates in endocrine and behavioral measures related to stress reactivity[95,96]. Meaney et al's work also made clear the possibility that epigenetic effects could occur at various stages throughout life, potentially influencing risk and vulnerability for chronic responses to trauma, such as PTSD, across the lifespan[82,97-101].

RELEVANCE OF EPIGENETIC MECHANISMS TO INTERGENERATIONAL EFFECTS

The term "epigenetics" refers to a set of potentially heritable changes in the genome that can be induced by environmental events. These changes affect the function of genomic DNA, its associated histone proteins, and non-coding RNAs, collectively referred to as chromatin, but do not involve an alteration of DNA sequence[102-104].

Of the many mechanisms of epigenetic regulation that have been described, DNA methylation at the cytosine site has been the best characterized in the mammalian genome[105,106]. Other regulators of chromatin include post-translational modification of histones and accompanying RNA-signaling as well as higher order changes in nucleosome organization[102].

Epigenetic modifications impact gene function by altering gene regulatory elements that affect the action of gene transcription factors[91]. Generally, methylation within specific regions of the gene is an efficient way of gene silencing and provides a molecular mechanism for the occurrence of gene-environment interactions independent of a specific genetic marker or gene version[107]. However, the actual contribution of genetic influences on environmentally-induced events has been insufficiently studied.

The impact of an epigenetic change on gene function is determined by the specific nature and location of an epigenetic mark on the gene and its proximity to the transcription start site, and possibly other genomic regulatory regions of interest[107-112]. It is not a trivial matter to determine the location on a gene, or within the genome, that would activate the relevant transcription factors which result in phenotypic change. The work of Meaney et al established a molecular mechanism for postnatal glucocorticoid programming, and identified the regions within the GR gene promoter that result in long-lasting changes in the biological systems associated with stress response in offspring[91,113].

Subsequent studies have built on this information by examining the 1F exon promoter, a relatively small area of the GR gene[57,114-120]. In fact, there may be numerous other areas of interest on the GR and other genes that have yet to be identified.

TRANSLATIONAL STUDIES LINKING EPIGENETIC FINDINGS ASSOCIATED WITH MATERNAL CARE IN ANIMALS TO CHILDHOOD ADVERSITY AND OFFSPRING EFFECTS IN HUMANS

The first documented study of the GR promoter in humans showed higher methylation of the hippocampal GR 1F promoter in post-mortem tissue of adult suicide victims with a history of childhood abuse, similar to findings in rodent pups raised by mothers who provided low levels of licking and grooming[121,122]. The findings in human brains of abuse victims implied that early developmental traumas such as those perpetrated by primary caretakers might influence the same neurobiological developmental systems as those involved in early maternal care[121].

Following this observation in post-mortem brain tissue, higher GR promoter methylation in circulating leukocytes of healthy adults was also found to be associated with disrupted, inadequate, or abusive parenting[123-125].

The above work provided a strong rationale for the examination of the GR promoter methylation in peripheral blood mononuclear cells of Holocaust offspring. In parallel with the neuroendocrine observations, the results of these analyses indicated a significant interaction of maternal and paternal PTSD on GR gene methylation[126]. The interaction demonstrated that, in the absence of maternal PTSD, offspring with paternal PTSD showed higher GR promoter methylation, whereas offspring with both maternal and paternal PTSD showed lower methylation of this promoter region. Lower GR 1F promoter methylation was significantly associated with greater GR sensitivity, as indicated by greater post-dexamethasone cortisol suppression. Furthermore, a clustering analysis of clinical self-report scales revealed that maternal and paternal PTSD were associated with different clinical indicators as well. Together, the data suggested that there are likely to be different underlying mechanisms for the intergenerational effects on offspring biology and behavior depending on parental gender and PTSD status.

Some findings in offspring, however, have not been directly linked with parental gender and PTSD status, in some cases because the small sample size prohibited such analysis. For example, a preliminary study examining FKBP5 intron 7 methylation in Holocaust survivors and their children demonstrated alterations at the same site within intron 7 in both parents and their children, with no specific consideration of parental gender or PTSD[127]. The FKBP5 gene encodes a protein that functions as a co-chaperone of the bound cortisol-glucocorticoid complex in the cell nucleus[128]. FKBP5 methylation in parents and their children were positively correlated. However, interestingly, they were directionally distinct (when compared to their respective control groups), with Holocaust offspring showing lower methylation at this site compared to demographically-matched controls, and Holocaust survivors demonstrating higher methylation compared to respective controls.

It is important to note that effects of parental behavior should not be conflated with directly "inherited" effects resulting from biological transmission from parent to child, even though both may be associated with epigenetic findings. Epigenetic mechanisms are operational throughout life and are highly responsive to environmental perturbations. It has now been shown that stressful experiences such as adult trauma change methylation of the GR gene in blood cells, whether primed by early experience or not[118,119,129,130].

PRENATAL MATERNAL CONTRIBUTIONS TO OFFSPRING VIA FETOPLACENTAL INTERACTIONS

An emerging body of literature has raised the possibility that maternal effects of trauma exposure might contribute to offspring effects through fetoplacental interactions[131-135]. This possibility is consistent with clinical, neuroendocrine and epigenetic findings, in which maternal and paternal PTSD predicted different psychiatric and biological outcomes in offspring[28,126].

The intrauterine environment presents a developmentally potent context[95], mechanistically distinct from postnatal parenting or family environment, through which maternal trauma or stressful experiences may influence fetal epigenetic programming of the HPA axis[136]. By 22 weeks of gestation, the fetal HPA axis is developed and functioning, although it continues to be sensitive to environmental influence[137,138]. The placenta nourishes and protects the fetus, buffering the effects of maternal glucocorticoids through the expression of placental 11B-hydroxysteroid dehydrogenase type 2 (11β-HSD2), an enzyme that converts cortisol to inactive cortisone[139].

In animal models, prenatal stress has been shown to lower expression of 11β-HSD2 mRNA and 11B-HSD2 activity, both of which are associated with increased 11β-HSD2 methylation in the placenta[140]. Such effects of prenatal stress would have profound consequences on fetal exposure to glucocorticoids and the development of glucocorticoid sensitive systems, such as the HPA axis.

The potential for maternal trauma or stress to program fetal development through placental alterations has been explored in animal and human studies, historically with an emphasis on HPA axis markers, but more recently using epigenetic measures[141-145].

The gestational stage of the fetus is an important determinant of the impact of prenatal exposures on offspring, indicative of developmentally sensitive windows of fetal development[146,147]. The relevance of gestational stage during maternal trauma exposure was highlighted in a prospective study of infants born to mothers who had been pregnant and had to evacuate the World Trade Center during the terrorist attacks on September 11, 2001[146]. Infant offspring demonstrated lower cortisol levels in association with maternal PTSD, particularly if the mother had been exposed to trauma in the third tri-

mester. At 9 months, maternal morning cortisol levels were inversely related to maternal ratings of infant distress and response to novelty. Mothers who had PTSD rated their infants as having greater distress to novelty than did mothers without PTSD[148], and the offspring of mothers with PTSD showed evidence of anxiety and behavioral disturbances.

The relevance of prenatal stage at exposure was also demonstrated by two important epidemiological studies of Swedish and Dutch famines, that identified transgenerational health and disease outcomes in children and grandchildren[149]. Phenotypic and epigenetic changes were observed in adults who were exposed *in utero* to the Dutch famine of 1944-45, but only among those exposed at the time of conception and during the first half of gestation, compared to those exposed in the third trimester or early postnatal period[150,151].

More recently, a relatively large epigenetic study of the Dutch hunger cohort (422 exposed and 463 sibling controls) identified alterations in DNA methylation specifically associated with *in utero* exposure to maternal famine[152]. Among those exposed early in gestation, additional CpG mediators were identified. Interestingly, exposure to famine during pregnancy had biological and behavioral effects on grandchildren, such as on adiposity[153]. This transgenerational effect has been attributed to the fact that prenatal exposure directly impacts both the fetus and the fetal germ cells, thus directly exposing the third generation. In a recent study, grandmaternal stress during pregnancy was associated with genome-wide methylation changes in offspring and grandchildren[154].

Studies of preconception exposure to trauma without specific consideration of gestational age of exposure have also been published. A number of smaller studies have identified an association of prenatal maternal trauma with methylation of the NR3C1 gene in offspring. Higher levels of NR3C1 methylation were observed in offspring aged 10-19 of mothers who experienced intimate partner violence during, but not prior to or following, pregnancy[155].

Higher methylation in the promoter of the NR3C1 gene was also observed in newborns of mothers in the Democratic Republic of Congo exposed to severe prenatal stress, with the strongest effect for maternal warzone stress experiences[156], and in the children of women exposed to the Tutsi genocide during pregnancy compared with non-genocide exposed women of the same ethnicity and pregnant at the same time, and their children[41]. Among offspring of women pregnant during the 1998 Quebec ice storm, those whose mothers experienced objective hardship, but not subjective distress, had methylation alterations in genes related to immune function[157]. These findings suggest enduring epigenetic alterations in offspring associated with maternal trauma during gestation.

Given the directionality of these findings, which is consistent with elevated cortisol levels, it may be that exposures in mothers which originate during pregnancy result in directionally different epigenetic alterations from those observed in offspring where the maternal (or paternal) trauma occurred prior to conception. There may also be different effects on offspring, as well mechanisms underlying these effects, depending on the history of trauma exposure and/or psychiatric symptoms prior to pregnancy.

One question that arises from studies of women traumatized prior to or during pregnancy is the extent to which effects on offspring are mediated by psychological symptoms or subjective reactions to adversity. It may be that intrauterine signals which affect fetal biology are driven by maternal symptoms such as anxiety, depression, or hyperarousal. It is certainly plausible that women with early childhood trauma or prenatal trauma exposure might experience pregnancy with ambivalence or distress[76]. Thus, any alteration in offspring may be mediated by mental health symptoms during gestation, and certainly extend to the postnatal environment. In studies of Holocaust offspring, perhaps the most salient observation has been that most differences in offspring phenotype were associated with persistent psychological effects of parents.

This question can also be partially addressed by considering studies of the effects of mood and anxiety disorders during pregnancy in the absence of trauma exposure. In one study, the effects of prenatal maternal depression on methylation levels in the promoter and exon 1F region of the NR3C1 gene in newborn cord blood identified a trimester effect, with third trimester maternal depression/anxiety associated with higher methylation of NR3C1 at a predicted NGF1-A binding site[141]. Functionally, methylation levels were associated with salivary cortisol stress responses in the newborns at 3 months, indicating that maternal mood and offspring HPA axis reactivity may be linked through epigenetic processes and sensitive to fetal developmental stage. In contrast, a study of pregnancy-related anxiety found that methylation at the 1F exon of the NR3C1 in offspring was influenced by maternal anxiety only during the first two trimesters[158].

FETOPLACENTAL INTERACTIONS: REGULATION BY SEX OF OFFSPRING

One of the most fascinating observations from studies examining the effects on offspring of maternal stress during pregnancy is that prenatal stress has different effects in male vs. female offspring[143,159-161]. In animal models of prenatal stress, exposure to chronic stress *in utero* increased male, but not female, HPA stress reactivity (e.g., behavioral response to the tail suspension test)[159,162]. These behaviors were transmitted to the next generation through the male germ line. Among mice exposed to stress during early, mid and late gestation, male F1 with early gestation prenatal stress exposure demonstrated behavioral indicators of stress responsivity and anhedonia, as well as alterations in GR and corticotrophin-releasing factor (CRF) expression and increased HPA axis responsivity, with corresponding alteration in CRF and nr3c1 gene methylation[159].

The importance of fetal sex, or more specifically, trophoblast cells from the embryo reflecting fetal sex, is that it may differentially regulate epigenetic signals in the placenta, leading to differential signaling that feeds back to the offspring[140]. These sex-related placental differences may confer protection or vulnerability to the fetus through differential exposure to maternal stress hormones. For example, early gestational stress exposure led to sex differences in expression and methylation of genes in the placenta associated with growth and nutrient transport[159].

A recent review of sex differences in HPA axis programming concluded that female offspring exposed to prenatal stressors had higher HPA axis reactivity than did similarly exposed males, with differences in placental expression of 11β-HSD enzymes, but that prenatal stress in humans was associated with alterations in diurnal cortisol secretion in males that were not apparent in females[163]. Thus there may be slightly different effects according to species and sex, depending on the parameter being measured.

While there is a strong suggestion that prenatal maternal effects produce a wide range of behavioral and biological outcomes in offspring, there is still an important need to provide clarification on the different contributions of maternal exposure, including the nature of the exposure, the timing of exposure in pregnancy, the sex of the fetus, the nature of maternal symptoms, or other potentially significant contributions such as nutrition, exposure to toxins, delivery factors, medication effects, socio-demographic variables, and other potential mediators.

In studies where offspring are also examined, it is difficult to break down effects of prenatal exposures from postnatal maternal factors, but studies examining offspring in close proximity to birth may be particularly informative regarding offspring biology. They will be less informative with respect to offspring phenotype as it is expressed later in life.

Studies of prenatal maternal exposures provide incomplete data regarding several other factors that may be relevant to offspring effects. Of particular interest are the potential contributions of preconception trauma in mothers (or fathers) to prenatal influences *in utero*. Preconception trauma exposure, prenatal stress, and postnatal parenting are unlikely to be independent in humans, adding to the complexity of drawing conclusions about specific influences on offspring.

INTERGENERATIONAL EFFECTS OF PRECONCEPTION MATERNAL TRAUMA

It is tempting to assume that findings of preconception trauma, particularly occurring prior to puberty, represent trauma-induced epigenetic changes to the oocyte that are maintained throughout embryogenesis and/or reestablished post-conception, thereby influencing the placental environment findings[164]. There are no studies to date examining this possibility in either animal or human samples. The complexities of examining this issue are obvious, because it is methodologically extremely difficult to separate out effects in an oocyte from effects of the fetoplacental environment. Although all of a female's oocytes are present at birth, they can be affected by environmental exposures, particularly during childhood[165]. Oocytes remain in a haploid de-methylated state until puberty, and are therefore vulnerable to environmental perturbations[166].

The notion that oocytes may be affected by preconception trauma is consistent with findings in Holocaust offspring in association with maternal age of exposure during the Holocaust. However, this explanation would decidedly be an inference. Maternal age at Holocaust exposure and maternal PTSD were found to independently influence urinary cortisol levels and cortisol metabolism in adult offspring, with the strongest effects in offspring of mothers who were children during World War II[167]. In an unpublished study, earlier age of maternal Holocaust exposure was also associated with lower FKBP5 methylation at intron 7 in offspring.

Such data must be interpreted with caution. Regarding exposures during World War II, including the studies of the Dutch famine, it is difficult, if not impossible, to ascertain exactly when the traumatic period began. The unknown variance associated with unmeasured stress in prior generations and its relevance to any maternal exposures is simply not known and creates a difficulty in ascertaining mechanisms. However, the limited data suggesting an association of an epigenetic alteration with maternal age at trauma exposure imply potential contributions of both *in utero* effects and possibly preconception epigenetic changes to gametes.

The difficulty in parsing different maternal contributors to offspring outcome does not mean that epigenetic changes to oocytes are not potential contributors to offspring phenotype – just that this has not yet been determined, and will require innovative methods of investigation. However, the possibility that trauma-related epigenetic changes in germ cells contribute to offspring phenotype has been demonstrated in association with sperm.

Offspring effects may be mediated, in part, by epigenetic changes in parental germ cells resulting from acquired parental stress exposures throughout life[3,168-170]. Germ cells in both females and males can be affected by trauma exposure, but the critical periods for affecting oocytes and sperm may differ. Accordingly, the nature of the effects may differ in oocytes and sperm in relation to trauma exposure. The extent to which exposure-related changes in germ cells are similar to epigenetic alterations in brain is an area for continued investigation[171,172].

PRECONCEPTION PATERNAL EFFECTS AND OFFSPRING PHENOTYPE: PROOF OF CONCEPT FOR THE ROLE OF SPERM

A rapidly growing literature has focused on paternal transmission through sperm[3,173]. Unlike oocytes, which are formed in females prior to birth, spermatogenesis in males is initiated

in the testes at puberty and continues throughout the lifespan[174]. Studying transmission through sperm eliminates confounds created by influences of fetoplacental environment, delivery factors, and maternal care as described above. Furthermore, paternal exposure to preconception stress at any stage of development might impact gametes but, as with females, there may be critical periods of vulnerability to insult.

Among the epigenetic mechanisms that have been implicated in paternal transmission of stress effects via sperm are DNA methylation, oxidative damage to sperm DNA, histone modifications, and changes in small noncoding RNA[175-179], including microRNA[180,181]. Changes in any of these properties in sperm could affect gene expression and other biological processes in the developing embryo and fetus, setting the stage for phenotypic change in offspring[182]. It is important to note that in cases where such processes then result in modifications of DNA methylation, the process of transmission would remain indirect, despite germ cell mediation. It is the event-related change in germ cell biology that produces the methylation mark, not the original "trauma".

To date, there are no known studies that have directly examined transgenerational effects mediated through sperm in humans. Thus, there is no information about epigenetic changes in sperm of fathers exposed to adversity with examination of potential corresponding changes in the sperm of their sons. However, there have been several observational studies demonstrating that environmental exposures in males – such as famine, obesity, smoking, alcohol consumption, exposure to toxins, and exposure to stress – result in subsequent behavioral and biologic effects in offspring[183-187]. Some of these exposures have also been associated with alterations in sperm of the exposed father. Still, the compelling data demonstrating heritable epigenetic alterations come from animal models[179,181], supported by an increasing understanding of the intricate details of epigenetic mechanisms associated with mammalian embryology and fetal development.

Contrary to initial understanding, it is now believed that some epigenetic changes in germ cells may survive the nearly global erasure of DNA methylation that occurs before implantation of the embryo, or associate with other epigenetic mechanisms[188,189]. DNA methylation marks are re-established following their erasure, allowing developmental processes, including cell differentiation, to occur[190]. Some embryonic cells will become germ cells (sperm and oocytes). In primordial germ cells, DNA methylation is again erased and re-established based on the sex of the transmitting parent[190]. Because of a phenomenon called imprinting, maternal and paternal genomes are differentially marked and re-programmed, and a small number of regions from the DNA of the parent of origin may remain with DNA methylation intact[173,189,191].

Genomic imprinting patterns can have major effects on the embryonic phenotype[192,193]. This provides at least one putative mechanism in addition to parent of origin effects for the transfer of an environmentally-induced epigenetic mark from one generation to another. It should be stated, however, that the exact nature of the mechanisms involved in transmission through gametes continues to remain obscure, and knowledge in this area is greatly expanding, even as such effects are demonstrated in mammalian studies[194-196].

It is of interest to compare effects of fathers who conceived during the Dutch hunger with effects of mothers who may have further influenced the development of the offspring *in utero*. Offspring of F1 fathers, but not F1 mothers, who were prenatally exposed to famine had higher body mass index and obesity rates as adults[197]. In Sweden, limited food supply affected mortality rates of grandchildren in a sex-specific manner through the paternal line. Restricted nutritional intake in paternal grandfathers affected mortality rates in grandsons only, whereas paternal grandmother access to food was associated with mortality of granddaughters. These effects were observed only when limited food access occurred prior to puberty, supporting the hypothesis that the transmission occurred through epigenetic programming of gametes and may be mediated by the X and Y sex chromosomes[181,195].

There are several observations that exposures of fathers or even grandfathers affect offspring through non-genomic mechanisms of transmission. A three-generational study of obesity in males and females demonstrated different risk and protective factors associated with grandparental and parental food availability during puberty[194]. Overeating in paternal grandfathers was associated with increased risk for diabetes in grandchildren, whereas limited food availability in fathers was associated with protection from cardiovascular death in sons. It was hypothesized that these changes were mediated by nutrition-related transgenerational effects down the male line, involving modifications of the DNA and/or histones in sperm. Interestingly, a reanalysis of these data showed that the child's early life circumstances were also relevant to findings from father to son but, when childhood factors in sons were controlled in statistical analyses, the transmission effects through the male line were strengthened[194].

The extent of paternal alcoholism has also been associated with neurological and behavioral deficits in offspring[198]. Changes in DNA methylation were observed in sperm from men with alcohol or opioid dependence[199,200], but effects in offspring were not measured. Smoking was reported to increase risk of childhood cancer in the offspring of male smokers[187], and was later found to be associated with reduced sperm count, motility and morphology, and altered sperm microRNA, mitochondria and protein in the smoker parent[201,202]. Data from the UK Avon Longitudinal Study of Parents and Children study identified effects of paternal smoking on offspring, but only when smoking occurred before puberty[195].

In these cases, it was hypothesized that environmental perturbations within the testes/epididymides led to epigenetic changes in the development or maturation of sperm that were then transferred to the oocyte at fertilization, affecting gene expression of the early embryo or modulating DNA methyltransferases or histone regulators.

In the absence of studies examining the effects of trauma through the male germ line in humans, the above findings demonstrate that a wide range of environmental exposures, not only exposure to extreme trauma, can have biological and behavioral effects that persist in one or more generations. Future studies examining behavioral and epigenetic effects in sperm in relation to pre- and post-pubertal trauma exposure in males and their offspring will greatly shed light on this topic.

STUDIES OF INTER- AND TRANSGENERATIONAL STRESS IN MALE RODENTS

Research on possible intergenerational transmission of stress effects through epigenetic marks in sperm has been conducted in rodents, and includes preconception exposures at various developmental stages to stressful and adverse social experiences[149,175,176,179,181,203,204]. Such studies have produced very compelling data suggesting that exposure to extreme stress in males can affect brain, behavior and sperm in the next generation[176,179].

In one study, male mice were fear conditioned with an odorant at two months of age (post-puberty but not yet adults)[175]. The odorant acetophenone paired with an electric shock resulted in behavioral sensitivity in the fear conditioned mice, with an accompanying change in DNA methylation in brain and sperm of the M71 receptor, which is involved in sensing acetophenone. An increased size in the M71 specific glomeruli in the olfactory epithelium and bulb was also observed[175]. The offspring (F1) of odor conditioned F0 males mated with naïve females also showed similar changes in brain and sperm. When the F1 males were themselves mated, changes in brain persisted in the F2 male offspring, demonstrating conservation of the effect through two generations.

In vitro fertilization was also used to implant the F0 sperm into a naïve female. This produced similar behavioral and biological findings in the F1, further pointing to biological inheritance through sperm. The *in vitro* fertilization study permitted changes to be attributed to sperm and not, for example, maternal reactions to behavior in the conditioned father during mating, or other potential confounds. To even more carefully eliminate any maternal contributions to offspring effects, a cross-fostering study was performed, which confirmed the absence of maternal effects on the observed offspring phenotype.

This series of studies provides a clear demonstration of an epigenetically mediated transgenerational biological inheritance through sperm of a behavioral trait and corresponding neuroanatomical brain changes that persist for two generations.

A similar observation of transgenerational paternal effects emerged from a different paradigm, in which two groups of male mice were exposed to a wide range of stressors over 42 days at puberty or adulthood[179]. These mice (F0) demonstrated behavioral changes in response to the stressor, and also changes in several specific sperm microRNA. Males were bred with naïve females and produced offspring with blunted HPA axis responsivity as well as changes in transcription of GR genes in the paraventricular nucleus[179].

These findings confirmed that early or later life exposures in the male mouse can affect germ cell microRNA, and are sufficient to result in a similar phenotype in the subsequent generation, again confirming the transmission through sperm in an independent animal model. This study is noteworthy for examining both male and female F1. Although significant sex differences were noted in endocrine and behavioral measures, there was no interaction between sex and paternal stress in the offspring of those exposed at puberty or adulthood.

An independent research team also demonstrated that small non-coding RNAs (sncRNAs), common in sperm, can mediate inheritance of environmentally acquired traits or phenotypes in mice[176]. Specifically, early life stress, modeled by unpredictable maternal separation and maternal stress, led to depressive-like behavioral patterns upon exposure to novel environments and changes in sncRNAs in F1 sperm. F0 exposed to several unpredictable maternal stressors and separation demonstrated changes that could be observed across two generations[176]. When altered microRNAs from sperm of the stressed males were injected into fertilized wild-type oocytes, comparable behavioral, metabolic and molecular outcomes were observed in the F2 offspring, indicating transmission of epigenetic marks. Furthermore, F3 offspring of these animals continued to show phenotypic differences, indicating conservation of stress effects through sperm.

Importantly, another study demonstrated that environmental enrichment following stress exposure in the F0 could reverse and prevent some of the effects[205]. Early maternal separation resulted in decreased nr3c1 DNA methylation in the hippocampus and sperm cells, as well as poor coping behavior. When environmental enrichment was applied at weaning until adulthood, the behavioral and methylation effects were no longer observed in the F0 or F1. These findings indicate that stress-induced changes to germ cells are not immutable and can be reversed by alternative environmental perturbations that are directed at stimulating plasticity. It is for this reason that environmental effects which cross the generations do not necessarily predict negative generational consequences – posing challenges for interpretation of such effects.

Furthermore, not all stressors impact sperm in an intergenerational manner. For example, in a social defeat model of stress, male and female F1 mice exhibited altered behaviors, and male F1 had a broader range of affected behaviors[204]. However, these results were not observed when offspring were generated by *in vitro* fertilization, implicating behavioral, rather than germ cell epigenetic, influence.

Thus, evidence is beginning to converge around the role of epigenetic mechanisms. However, there is much diversity in effects, and opportunities for modifying even strong effects of non-coding RNAs, chromatin, and DNA methylation. Future

research can delineate the exact nature of the stressors and their sensitivity to reversal through targeted environmental influences designed to enhance resilience[175,206,207].

CONCLUSIONS AND FUTURE DIRECTIONS

Scientific studies are rapidly identifying epigenetic mechanisms to explain how an environmental exposure may lead to an enduring change in the function of DNA that can be passed to future generations. This review emphasized two broad categories of offspring effects that may be underpinned by epigenetic mechanisms. The first involve accommodations made by offspring in response to their own environmental exposures in early life, or even *in utero*. These changes are likely to be mediated primarily by maternal trauma-related symptoms, but may be affected by multiple inputs, including paternal trauma-related effects. The second are the effects of a preconception parental trauma that remain in the germ cell and following conception, affecting the offspring's development *in utero* and subsequent postnatal phenotype.

In both cases, the transmission is a result of parental exposure effects. In the context of offspring born to two trauma survivors, these two modes of epigenetic influences are likely to interact, and it is indeed very difficult to parse out the many potential contributions to offspring phenotype, not to mention those related to the offspring's own experiences through childhood, adolescence and adulthood.

Epigenetic mechanisms have been favored over genetic explanations (or gene-environment interactions) of intergenerational effects in part because of their potential to explain the phenotypic differences in offspring associated with maternal vs. paternal trauma exposure. The state of the science in relation to human offspring at present is that, whereas some neuroendocrine and epigenetic alterations have been documented in connection with maternal and paternal trauma exposure and PTSD, studies have not yet conclusively demonstrated epigenetic transmission of trauma effects in humans.

Nonetheless, the findings in animal models implicating epigenetic mechanisms in the transmission of stress effects through germ cells have created much excitement for the possibility that similar mechanisms might be operating in humans. Identifying evidence for these mechanisms will require prospective, longitudinal, and multi-generational studies. Parallel studies in animals will permit a more rigorous elucidation of the effects of specific experiences and mechanisms through cross-fostering and *in vitro* fertilization studies.

Research on epigenetic inheritance of effects of trauma faces many scientific and methodological complexities, not to mention conceptual issues regarding interpretation of transmitted effects. This review has not examined the contribution of genetic factors to trauma-related epigenetic alterations, but future studies should incorporate an understanding of both the genetic and environmental factors that augment or mitigate offspring effects.

Other areas for future studies concern the relevance of age or developmental stage of the parental trauma exposure to offspring effects, as well as the notion that male and female offspring may be differentially affected by maternal and paternal trauma. Moreover, there is a very small, but emerging, literature regarding potential reversal of intergenerational effects and their implications for resilience[205].

At the present time, the field has not sufficiently grappled with the meaning of the intergenerational transmission of trauma effects for the offspring. It could be argued that this transmission is indicative of increased vulnerability. On the other hand, this transmission may extend the adaptive capacities of offspring through a biological preparation for adverse circumstances similar to those encountered by the parent. Ultimately, the potential utility, and possible stability, of an environmentally induced trait transmitted to an offspring will depend on the offspring's environmental context.

This review highlights some of the complexities involved in making inferences about the mechanisms that underlie intergenerational and transgenerational transmission. It is inarguable that people feel affected by the consequences of trauma exposure in previous generations. The assertion that an effect is truly transgenerational requires ruling out direct exposure of offspring as a causal mechanism. Thus, for females, traits must be observed in F3 females to be considered transgenerational, because the F1 female offspring is exposed to the stressor during gestation through the intrauterine environment. This may, in turn, affect programming of the F1 fetus' germline, which would be observed in her F2 offspring. Only the originally exposed mother's F3 offspring would not have been directly exposed to the stressor. In males, F1 may be influenced via the germ line of the exposed F0 father, but since sperm is not generated in the fetus (as ova are in females), transmission of trauma-associated traits to F2 would be considered transgenerational transmission.

These guidelines should be kept in mind as studies on effects of trauma on offspring in the next and subsequent generations are pursued. The concept of intergenerational transmission has resonated among offspring who feel affected by their parents' experience. The concept has also been embraced by communities that are affected by significant traumatic experiences through several generations. That there may be a biological or molecular representation of an intergenerational effect appears to validate the experience of offspring who may feel that they bear effects of their parents' hardship, even if the concept may also carry an implication that they are damaged, impaired, or permanently disadvantaged. It is also important to underscore the lack of permanence of effects once environmental conditions are altered.

Continued research in this field will likely reveal that epigenetically induced changes are a reflection of environmental exposure, and therefore by definition malleable. Even potentially heritable changes can be modified, because environments change. The role of genetics in mediating environmentally induced epigenetic effects remains an important frontier.

Regardless, the principle of epigenetic plasticity implies that changes to the epigenome might reset when the environmental insults are no longer present, or when we have changed sufficiently to address environmental challenges in a new way. It is the ability to flexibly respond to environmental stimuli that is fundamentally adaptive and the basis of human resilience.

ACKNOWLEDGEMENTS

The authors would like to thank L.M. Bierer and M.J. Meaney for their comments and very careful review of this paper, and A. Ropes for assistance with manuscript preparation.

REFERENCES

1. Chan JC, Nugent BM, Bale TL. Parental advisory: maternal and paternal stress can impact offspring neurodevelopment. Biol Psychiatry 2018;83: 886-94.
2. Clarke HJ, Vieux KF. Epigenetic inheritance through the female germ-line: the known, the unknown, and the possible. Semin Cell Dev Biol 2015;43: 106-16.
3. Bohacek J, Mansuy IM. Molecular insights into transgenerational non-genetic inheritance of acquired behaviours. Nat Rev Genet 2015;16: 641-52.
4. Ly L, Chan D, Trasler JM. Developmental windows of susceptibility for epigenetic inheritance through the male germline. Semin Cell Dev Biol 2015;43:96-105.
5. Ferguson-Smith AC. Genomic imprinting: the emergence of an epigenetic paradigm. Nat Rev Genet 2011;12:565-75.
6. Rando OJ. Intergenerational transfer of epigenetic information in sperm. Cold Spring Harbor Perspect Med 2016;6.
7. Yehuda R, Lehrner A, Bierer LM. The public reception of putative epigenetic mechanisms in the transgenerational effects of trauma. Submitted for publication.
8. Rakoff V. A long term effect of the concentration camp experience. Viewpoints 1966;1:17-22.
9. Sigal JJ, Weinfeld M. Trauma and rebirth: intergenerational effects of the Holocaust. New York: Praeger, 1989.
10. Yehuda R, Giller E. Comments on the lack of integration between the Holocaust and PTSD literatures. PTSD Research Quarterly 1994;5:5-7.
11. Kellerman NP. Psychopathology in children of Holocaust survivors: a review of the research literature. Isr J Psychiatry Relat Sci 2001;38:36.
12. Solkoff N. Children of survivors of the Nazi Holocaust: a critical review of the literature. Am J Orthopsychiatry 1992;62:342.
13. Van Ijzendoorn MH, Bakermans-Kranenburg MJ, Sagi-Schwartz A. Are children of Holocaust survivors less well-adapted? A meta-analytic investigation of secondary traumatization. J Trauma Stress 2003;16:459-69.
14. Ancharoff MRM, Munroe J, Fisher L. The legacy of combat trauma: clinical implications of intergenerational transmission. In: Danieli Y (ed). International handbook of multigenerational legacies of trauma. Boston: Springer, 1998:257-76.
15. Harkness L. Transgenerational transmission of war-related trauma. In: Wilson JP, Raphael B (eds). International handbook of traumatic stress syndromes. Boston: Springer, 1993:635-43.
16. Rosenheck R, Nathan P. Secondary traumatization in children of Vietnam veterans. Hosp Community Psychiatry 1985;36:538-9.
17. Figley CR. Catastrophes: an overview of family reactions. In: Figley CR, McCubbin HI (eds). New York: Brunner/Mazel, 1983:3-20.
18. Barocas HA, Barocas CB. Separation-individuation conflicts in children of holocaust survivors. J Contemp Psychother 1980;11:6-14.
19. Kestenberg JS. Psychoanalyses of children of survivors from the holocaust: case presentations and assessment. J Am Psychoanal Assoc 1980; 28:775-804.
20. Schwartz S, Dohrenwend BP, Levav I. Nongenetic familial transmission of psychiatric disorders? Evidence from children of Holocaust survivors. J Health Soc Behav 1994;35:385-402.
21. Danieli Y. International handbook of multigenerational legacies of trauma. New York: Plenum, 1998.
22. Bar-On D, Eland J, Kleber RJ et al. Multigenerational perspectives on coping with the Holocaust experience: an attachment perspective for understanding the developmental sequelae of trauma across generations. Int J Behav Dev 1998;22:315-38.
23. Abrams MS. Intergenerational transmission of trauma: recent contributions from the literature of family systems approaches to treatment. Am J Psychother 1999;53:225.
24. Schechter DS, Zygmunt A, Coates SW et al. Caregiver traumatization adversely impacts young children's mental representations on the MacArthur Story Stem Battery. Attach Hum Dev 2007;9:187-205.
25. Solomon Z, Kotler M, Mikulincer M. Combat-related posttraumatic stress disorder among second-generation Holocaust survivors: preliminary findings. Am J Psychiatry 1988;145:865-8.
26. Yehuda R, Schmeidler J, Wainberg M et al. Vulnerability to posttraumatic stress disorder in adult offspring of Holocaust survivors. Am J Psychiatry 1998;155:1163-71.
27. Yehuda R, Schmeidler J, Giller EL et al. Relationship between posttraumatic stress disorder characteristics of Holocaust survivors and their adult offspring. Am J Psychiatry 1998;155:841-3.
28. Yehuda R, Bell A, Bierer LM et al. Maternal, not paternal, PTSD is related to increased risk for PTSD in offspring of Holocaust survivors. J Psychiatr Res 2008;42:1104-11.
29. O'Toole BI, Burton MJ, Rothwell A et al. Intergenerational transmission of post-traumatic stress disorder in Australian Vietnam veterans' families. Acta Psychiatr Scand 2017;135:363-72.
30. Evans-Campbell T. Historical trauma in American Indian/Native Alaska communities: a multilevel framework for exploring impacts on individuals, families, and communities. J Interpers Violence 2008;23:316-38.
31. Gone JP. Redressing First Nations historical trauma: theorizing mechanisms for indigenous culture as mental health treatment. Transcult Psychiatry 2013;50:683-706.
32. Eyerman R. Cultural trauma: slavery and the formation of African American identity. Cambridge: Cambridge University Press, 2001.
33. DeGruy J. Post traumatic slave syndrome: America's legacy of enduring injury and healing. Baltimore: Uptone, 2005.
34. Raphael B, Swan P, Martinek N. Intergenerational aspects of trauma for Australian Aboriginal people. In: Danieli Y (ed). International handbook of multigenerational legacies of trauma. Boston: Springer, 1998: 327-39.
35. Pihama L, Reynolds P, Smith C et al. Positioning historical trauma theory within Aotearoa New Zealand. AlterNative 2014;10:248-62.
36. Field NP, Muong S, Sochanvimean V. Parental styles in the intergenerational transmission of trauma stemming from the Khmer Rouge regime in Cambodia. Am J Orthopsychiatry 2013;83:483-94.
37. Münyas B. Genocide in the minds of Cambodian youth: transmitting (hi)-stories of genocide to second and third generations in Cambodia. J Genocide Res 2008;10:413-39.
38. Azarian-Ceccato N. Reverberations of the Armenian genocide: narrative's intergenerational transmission and the task of not forgetting. Narrat Inq 2010;20:106-23.
39. Karenian H, Livaditis M, Karenian S et al. Collective trauma transmission and traumatic reactions among descendants of Armenian refugees. Int J Soc Psychiatry 2011;57:327-37.
40. Roth M, Neuner F, Elbert T. Transgenerational consequences of PTSD: risk factors for the mental health of children whose mothers have been exposed to the Rwandan genocide. Int J Ment Health Syst 2014;8:12.
41. Perroud N, Rutembesa E, Paoloni-Giacobino A et al. The Tutsi genocide and transgenerational transmission of maternal stress: epigenetics and biology of the HPA axis. World J Biol Psychiatry 2014;15:334-45.
42. Barron IG, Abdallah G. Intergenerational trauma in the occupied Palestinian territories: effect on children and promotion of healing. J Child Adolesc Trauma 2015;8:103-10.
43. Svob C, Brown NR, Takšić V et al. Intergenerational transmission of historical memories and social-distance attitudes in post-war second-generation Croatians. Mem Cognit 2016;44:846-55.
44. Brand SR, Brennan PA, Newport DJ et al. The impact of maternal childhood abuse on maternal and infant HPA axis function in the postpartum period. Psychoneuroendocrinology 2010;35:686-93.
45. Jovanovic T, Smith A, Kamkwalala A et al. Physiological markers of anxiety are increased in children of abused mothers. J Child Psychol Psychiatry 2011;52:844-52.

46. Juul SH, Hendrix C, Robinson B et al. Maternal early-life trauma and affective parenting style: the mediating role of HPA-axis function. Arch Womens Ment Health 2016;19:17-23.
47. Myhre MC, Dyb GA, Wentzel-Larsen T et al. Maternal childhood abuse predicts externalizing behaviour in toddlers: a prospective cohort study. Scand J Publ Health 2014;42:263-9.
48. Yehuda R. Biological factors associated with susceptibility to posttraumatic stress disorder. Can J Psychiatry 1999;44:34-9.
49. Yehuda R, McFarlane AC. Conflict between current knowledge about posttraumatic stress disorder and its original conceptual basis. Am J Psychiatry 1995;152:1705-13.
50. Seckl JR. Glucocorticoids, developmental 'programming' and the risk of affective dysfunction. Prog Brain Res 2007;167:17-34.
51. Matthews SG. Early programming of the hypothalamo-pituitary-adrenal axis. Trends Endocrinol Metab 2002;13:373-80.
52. Price JL, Drevets WC. Neurocircuitry of mood disorders. Neuropsychopharmacology 2010;35:192.
53. Shin LM, Liberzon I. The neurocircuitry of fear, stress, and anxiety disorders. Neuropsychopharmacology 2010;35:169-91.
54. Nestler EJ, Barrot M, DiLeone RJ et al. Neurobiology of depression. Neuron 2002;34:13-25.
55. Yehuda R, Kahana B, Binder-Brynes K et al. Low urinary cortisol excretion in Holocaust survivors with posttraumatic stress disorder. Am J Psychiatry 1995;152:982-6.
56. McGowan PO. Epigenomic mechanisms of early adversity and HPA dysfunction: considerations for PTSD research. Front Psychiatry 2013;4:110.
57. Palma-Gudiel H, Cordova-Palomera A, Leza JC et al. Glucocorticoid receptor gene (NR3C1) methylation processes as mediators of early adversity in stress-related disorders causality: a critical review. Neurosci Biobehav Rev 2015;55:520-35.
58. Yehuda R, Koenen KC, Galea S et al. The role of genes in defining a molecular biology of PTSD. Dis Markers 2011;30:67-76.
59. Zannas AS, Provencal N, Binder EB. Epigenetics of posttraumatic stress disorder: current evidence, challenges, and future directions. Biol Psychiatry 2015;78:327-35.
60. Yehuda R, Halligan SL, Bierer LM. Cortisol levels in adult offspring of Holocaust survivors: relation to PTSD symptom severity in the parent and child. Psychoneuroendocrinology 2002;27:171-80.
61. Yehuda R, Teicher MH, Seckl JR et al. Parental posttraumatic stress disorder as a vulnerability factor for low cortisol trait in offspring of holocaust survivors. Arch Gen Psychiatry 2007;64:1040-8.
62. Yehuda R, Blair W, Labinsky E et al. Effects of parental PTSD on the cortisol response to dexamethasone administration in their adult offspring. Am J Psychiatry 2007;164:163-6.
63. Lehrner A, Bierer LM, Passarelli V et al. Maternal PTSD associates with greater glucocorticoid sensitivity in offspring of Holocaust survivors. Psychoneuroendocrinology 2014;40:213-20.
64. Yehuda R, Bierer LM, Schmeidler J et al. Low cortisol and risk for PTSD in adult offspring of holocaust survivors. Am J Psychiatry 2000;157:1252-9.
65. Yahyavi ST, Zarghami M, Naghshvar F et al. Relationship of cortisol, norepinephrine, and epinephrine levels with war-induced posttraumatic stress disorder in fathers and their offspring. Rev Bras Psiquiatr 2015;37:93-8.
66. Yehuda R, Halligan SL, Grossman R. Childhood trauma and risk for PTSD: relationship to intergenerational effects of trauma, parental PTSD, and cortisol excretion. Dev Psychopathol 2001;13:733-53.
67. Heim C, Newport DJ, Bonsall R et al. Altered pituitary-adrenal axis responses to provocative challenge tests in adult survivors of childhood abuse. Am J Psychiatry 2001;158:575-81.
68. Carpenter LL, Carvalho JP, Tyrka AR et al. Decreased adrenocorticotropic hormone and cortisol responses to stress in healthy adults reporting significant childhood maltreatment. Biol Psychiatry 2007;62:1080-7.
69. Heim C, Ehlert U, Hanker JP et al. Abuse-related posttraumatic stress disorder and alterations of the hypothalamic-pituitary-adrenal axis in women with chronic pelvic pain. Psychosom Med 1998;60:309-18.
70. King JA, Mandansky D, King S et al. Early sexual abuse and low cortisol. Psychiatry Clin Neurosci 2001;55:71-4.
71. Trickett PK, Noll JG, Susman EJ et al. Attenuation of cortisol across development for victims of sexual abuse. Dev Psychopathol 2010;22:165-75.
72. Schreuder MM, Vinkers CH, Mesman E et al. Childhood trauma and HPA axis functionality in offspring of bipolar parents. Psychoneuroendocrinology 2016;74:316-23.
73. Danielson CK, Hankin BL, Badanes LS. Youth offspring of mothers with posttraumatic stress disorder have altered stress reactivity in response to a laboratory stressor. Psychoneuroendocrinology 2015;53:170-8.
74. Cordero MI, Moser DA, Manini A et al. Effects of interpersonal violence-related post-traumatic stress disorder (PTSD) on mother and child diurnal cortisol rhythm and cortisol reactivity to a laboratory stressor involving separation. Horm Behav 2017;90:15-24.
75. Moog NK, Entringer S, Rasmussen JM et al. Intergenerational effect of maternal exposure to childhood maltreatment on newborn brain anatomy. Biol Psychiatry 2018;83:120-7.
76. Yehuda R, Meaney MJ. Relevance of psychological symptoms in pregnancy to intergenerational effects of preconception trauma. Biol Psychiatry 2018;83:94-6.
77. Meaney MJ, Aitken DH, Bodnoff SR et al. The effects of postnatal handling on the development of the glucocorticoid receptor systems and stress recovery in the rat. Prog Neuropsychopharmacol Biol Psychiatry 1985;9:731-4.
78. Meaney MJ, Aitken DH, Bodnoff SR et al. Early postnatal handling alters glucocorticoid receptor concentrations in selected brain regions. Behav Neurosci 1985;99:765-70.
79. Meaney MJ, Aitken DH, van Berkel C et al. Effect of neonatal handling on age-related impairments associated with the hippocampus. Science 1988;239:766-8.
80. Francis D, Diorio J, Liu D et al. Nongenomic transmission across generations of maternal behavior and stress responses in the rat. Science 1999;286:1155-8.
81. Liu D, Diorio J, Tannenbaum B et al. Maternal care, hippocampal glucocorticoid receptors, and hypothalamic-pituitary-adrenal responses to stress. Science 1997;277:1659-62.
82. Yehuda R, Bierer LM. The relevance of epigenetics to PTSD: implications for the DSM-V. J Trauma Stress 2009;22:427-34.
83. Yehuda R, Bierer LM. Transgenerational transmission of cortisol and PTSD risk. Prog Brain Res 2008;167:121-35.
84. Ullmann E, Licinio J, Barthel A et al. Persistent LHPA activation in German individuals raised in an overprotective parental behavior. Sci Rep 2017;7:2778.
85. Weaver IC, Szyf M, Meaney MJ. From maternal care to gene expression: DNA methylation and the maternal programming of stress responses. Endocr Res 2002;28:699.
86. Weaver IC, Cervoni N, Champagne FA et al. Epigenetic programming by maternal behavior. Nat Neurosci 2004;7:847-54.
87. McGowan PO, Suderman M, Sasaki A et al. Broad epigenetic signature of maternal care in the brain of adult rats. PLoS One 2011;6:e14739.
88. Weaver IC, Hellstrom IC, Brown SE et al. The methylated-DNA binding protein MBD2 enhances NGFI-A (egr-1)-mediated transcriptional activation of the glucocorticoid receptor. Philos Trans R Soc Lond B Biol Sci 2014;369:1652.
89. Suderman M, McGowan PO, Sasaki A et al. Conserved epigenetic sensitivity to early life experience in the rat and human hippocampus. Proc Natl Acad Sci USA 2012;109:17266-72.
90. Champagne F, Meaney MJ. Like mother, like daughter: evidence for non-genomic transmission of parental behavior and stress responsivity. Prog Brain Res 2001;133:287-302.
91. Meaney MJ, Szyf M. Environmental programming of stress responses through DNA methylation: life at the interface between a dynamic environment and a fixed genome. Dialogues Clin Neurosci 2005;7:103-23.
92. Waddington CH. The epigenotype. Int J Epidemiol 2012;41:10-3.
93. Kappeler L, Meaney MJ. Epigenetics and parental effects. BioEssays 2010;32:818-27.
94. Meaney MJ, Szyf M. Maternal care as a model for experience-dependent chromatin plasticity? Trends Neurosci 2005;28:456-63.
95. Seckl JR, Meaney MJ. Glucocorticoid "programming" and PTSD risk. Ann N Y Acad Sci 2006;1071:351-78.
96. Weaver IC. Epigenetic programming by maternal behavior and pharmacological intervention. Nature versus nurture: let's call the whole thing off. Epigenetics 2007;2:22-8.
97. Bagot RC, Meaney MJ. Epigenetics and the biological basis of gene x environment interactions. J Am Acad Child Adolesc Psychiatry 2010;49:752-71.
98. Dudley KJ, Li X, Kobor MS et al. Epigenetic mechanisms mediating vulnerability and resilience to psychiatric disorders. Neurosci Biobehav Rev 2011;35:1544-51.

99. Klengel T, Binder EB. Epigenetics of stress-related psychiatric disorders and gene x environment interactions. Neuron 2015;86:1343-57.
100. Kofink D, Boks MP, Timmers HT et al. Epigenetic dynamics in psychiatric disorders: environmental programming of neurodevelopmental processes. Neurosci Biobehav Rev 2013;37:831-45.
101. Matosin N, Halldorsdottir T, Binder EB. Understanding the molecular mechanisms underpinning gene by environment interactions in psychiatric disorders: the FKBP5 model. Biol Psychiatry 2018;83:821-30.
102. Bonasio R, Tu S, Reinberg D. Molecular signals of epigenetic states. Science 2010;330:612-6.
103. D'Urso A, Brickner JH. Epigenetic transcriptional memory. Curr Genet 2017;63:435-9.
104. Zannas AS, Chrousos GP. Epigenetic programming by stress and glucocorticoids along the human lifespan. Mol Psychiatry 2017;22:640-6.
105. Novik KL, Nimmrich I, Genc B et al. Epigenomics: genome-wide study of methylation phenomena. Curr Issues Mol Biol 2002;4:111-28.
106. Du J, Johnson LM, Jacobsen SE et al. DNA methylation pathways and their crosstalk with histone methylation. Nat Rev Mol Cell Biol 2015;16:519-32.
107. Meaney MJ, Ferguson-Smith A. Epigenomic regulation of the neural transcriptome: the meaning of the marks. Nat Neurosci 2010;13:1313-8.
108. Deaton AM, Bird A. CpG islands and the regulation of transcription. Genes Dev 2011;25:1010-22.
109. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet 2012;13:484-92.
110. Zhang Z, Liu J, Kaur M et al. Characterization of DNA methylation and its association with other biological systems in lymphoblastoid cell lines. Genomics 2012;99:209-19.
111. Jones PA. The DNA methylation paradox. Trends Genet 1999;15:34-7.
112. Provencal N, Binder EB. The neurobiological effects of stress as contributors to psychiatric disorders: focus on epigenetics. Curr Opin Neurobiol 2015;30:31-7.
113. Roth TL, Sweatt JD. Regulation of chromatin structure in memory formation. Curr Opin Neurobiol 2009;19:336-42.
114. Argentieri MA, Nagarajan S, Seddighzadeh B et al. Epigenetic pathways in human disease: the impact of DNA methylation on stress-related pathogenesis and current challenges in biomarker development. EBioMedicine 2017;18:327-50.
115. Daskalakis NP, Yehuda R. Site-specific methylation changes in the glucocorticoid receptor exon 1F promoter in relation to life adversity: systematic review of contributing factors. Front Neurosci 2014;8:369.
116. Turecki G, Meaney MJ. Effects of the social environment and stress on glucocorticoid receptor gene methylation: a systematic review. Biol Psychiatry 2016;79:87-96.
117. Castro-Vale I, van Rossum EF, Machado JC et al. Genetics of glucocorticoid regulation and posttraumatic stress disorder – What do we know? Neurosci Biobehav Rev 2016;63:143-57.
118. Labonte B, Azoulay N, Yerko V et al. Epigenetic modulation of glucocorticoid receptors in posttraumatic stress disorder. Transl Psychiatry 2014;4:e368.
119. McNerney MW, Sheng T, Nechvatal JM et al. Integration of neural and epigenetic contributions to posttraumatic stress symptoms: the role of hippocampal volume and glucocorticoid receptor gene methylation. PLoS One 2018;13:e0192222.
120. van Zuiden M, Geuze E, Willemen HL et al. Glucocorticoid receptor pathway components predict posttraumatic stress disorder symptom development: a prospective study. Biol Psychiatry 2012;71:309-16.
121. McGowan PO, Sasaki A, D'alessio AC et al. Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. Nat Neurosci 2009;12:342-8.
122. Labonte B, Yerko V, Gross J et al. Differential glucocorticoid receptor exon 1(B), 1(C), and 1(H) expression and methylation in suicide completers with a history of childhood abuse. Biol Psychiatry 2012;72:41-8.
123. Tyrka AR, Price LH, Marsit C et al. Childhood adversity and epigenetic modulation of the leukocyte glucocorticoid receptor: preliminary findings in healthy adults. PLoS One 2012;7:e30148.
124. Tyrka AR, Parade SH, Eslinger NM et al. Methylation of exons 1D, 1F, and 1H of the glucocorticoid receptor gene promoter and exposure to adversity in preschool-aged children. Dev Psychopathol 2015;27:577-85.
125. Tyrka AR, Parade SH, Welch ES et al. Methylation of the leukocyte glucocorticoid receptor gene promoter in adults: associations with early adversity and depressive, anxiety and substance-use disorders. Transl Psychiatry 2016;6:e848.
126. Yehuda R, Daskalakis NP, Lehrner A et al. Influences of maternal and paternal PTSD on epigenetic regulation of the glucocorticoid receptor gene in Holocaust survivor offspring. Am J Psychiatry 2014;171:872-80.
127. Yehuda R, Daskalakis NP, Bierer LM et al. Holocaust exposure induced intergenerational effects on FKBP5 methylation. Biol Psychiatry 2016;80:372-80.
128. Klengel T, Binder EB. FKBP5 allele-specific epigenetic modification in gene by environment interaction. Neuropsychopharmacology 2015;40:244-6.
129. Schur RR, Boks MP, Rutten BPF et al. Longitudinal changes in glucocorticoid receptor exon 1F methylation and psychopathology after military deployment. Transl Psychiatry 2017;7:e1181.
130. Yehuda R, Flory JD, Bierer LM et al. Lower methylation of glucocorticoid receptor gene promoter 1F in peripheral blood of veterans with posttraumatic stress disorder. Biol Psychiatry 2015;77:356-64.
131. Buss C, Entringer S, Moog NK et al. Intergenerational transmission of maternal childhood maltreatment exposure: implications for fetal brain development. J Am Acad Child Adolesc Psychiatry 2017;56:373-82.
132. Choi KW, Sikkema KJ. Childhood maltreatment and perinatal mood and anxiety disorders: a systematic review. Trauma Violence Abuse 2016;17:427-53.
133. Kim DR, Bale TL, Epperson CN. Prenatal programming of mental illness: current understanding of relationship and mechanisms. Curr Psychiatry Rep 2015;17:5.
134. Moog NK, Buss C, Entringer S et al. Maternal exposure to childhood trauma is associated during pregnancy with placental-fetal stress physiology. Biol Psychiatry 2016;79:831-9.
135. Nemoda Z, Szyf M. Epigenetic alterations and prenatal maternal depression. Birth Defects Res 2017;109:888-97.
136. Sosnowski DW, Booth C, York TP et al. Maternal prenatal stress and infant DNA methylation: a systematic review. Dev Psychobiol 2018;60:127-39.
137. Challis JR, Sloboda D, Matthews SG et al. The fetal placental hypothalamic-pituitary-adrenal (HPA) axis, parturition and post natal health. Mol Cell Endocrinol 2001;185:135-44.
138. Mastorakos G, Ilias I. Maternal and fetal hypothalamic-pituitary-adrenal axes during pregnancy and postpartum. Ann N Y Acad Sci 2003;997:136-49.
139. Reynolds RM. Glucocorticoid excess and the developmental origins of disease: two decades of testing the hypothesis. Psychoneuroendocrinology 2013;38:1-11.
140. Nugent BM, Bale TL. The omniscient placenta: metabolic and epigenetic regulation of fetal programming. Front Neuroendocrinol 2015;39:28-37.
141. Oberlander TF, Weinberg J, Papsdorf M et al. Prenatal exposure to maternal depression, neonatal methylation of human glucocorticoid receptor gene (NR3C1) and infant cortisol stress responses. Epigenetics 2008;3:97-106.
142. Babenko O, Kovalchuk I, Metz GA. Stress-induced perinatal and transgenerational epigenetic programming of brain development and mental health. Neurosci Biobehav Rev 2015;48:70-91.
143. O'Donnell KJ, Meaney MJ. Fetal origins of mental health: the developmental origins of health and disease hypothesis. Am J Psychiatry 2017;174:319-28.
144. Provencal N, Binder EB. The effects of early life stress on the epigenome: from the womb to adulthood and even before. Exp Neurol 2015;268:10-20.
145. Reynolds RM, Jacobsen GH, Drake AJ. What is the evidence in humans that DNA methylation changes link events in utero and later life disease? Clin Endocrinol 2013;78:814-22.
146. Yehuda R, Engel SM, Brand SR et al. Transgenerational effects of posttraumatic stress disorder in babies of mothers exposed to the World Trade Center attacks during pregnancy. J Clin Endocrinol Metab 2005;90:4115-8.
147. Glover V, O'connor T, O'Donnell K. Prenatal stress and the programming of the HPA axis. Neurosci Biobehav Rev 2010;35:17-22.
148. Brand SR, Engel SM, Canfield RL et al. The effect of maternal PTSD following in utero trauma exposure on behavior and temperament in the 9-month-old infant. Ann N Y Acad Sci 2006;1071:454-8.
149. Bale TL. Lifetime stress experience: transgenerational epigenetics and germ cell programming. Dialogues Clin Neurosci 2014;16:297.
150. Heijmans BT, Tobi EW, Stein AD et al. Persistent epigenetic differences associated with prenatal exposure to famine in humans. Proc Natl Acad Sci USA 2008;105:17046-9.
151. Ravelli GP, Stein ZA, Susser MW. Obesity in young men after famine exposure in utero and early infancy. N Engl J Med 1976;295:349-53.
152. Tobi EW, Slieker RC, Luijk R et al. DNA methylation as a mediator of the association between prenatal adversity and risk factors for metabolic disease in adulthood. Sci Adv 2018;4.

153. Painter RC, Osmond C, Gluckman P et al. Transgenerational effects of prenatal exposure to the Dutch famine on neonatal adiposity and health in later life. BJOG 2008;115:1243-9.
154. Serpeloni F, Radtke K, de Assis SG et al. Grandmaternal stress during pregnancy and DNA methylation of the third generation: an epigenome-wide association study. Transl Psychiatry 2017;7:e1202.
155. Radtke KM, Ruf M, Gunter HM et al. Transgenerational impact of intimate partner violence on methylation in the promoter of the glucocorticoid receptor. Transl Psychiatry 2011;1:e21.
156. Mulligan C, D'Errico N, Stees J et al. Methylation changes at NR3C1 in newborns associate with maternal prenatal stress exposure and newborn birth weight. Epigenetics 2012;7:853-7.
157. Cao-Lei L, Massart R, Suderman MJ et al. DNA methylation signatures triggered by prenatal maternal stress exposure to a natural disaster: project ice storm. PLoS One 2014;9:e107653.
158. Hompes T, Izzi B, Gellens E et al. Investigating the influence of maternal cortisol and emotional state during pregnancy on the DNA methylation status of the glucocorticoid receptor gene (NR3C1) promoter region in cord blood. J Psychiatr Res 2013;47:880-91.
159. Mueller BR, Bale TL. Sex-specific programming of offspring emotionality after stress early in pregnancy. J Neurosci 2008;28:9055-65.
160. McCormick CM, Smythe JW, Sharma S et al. Sex-specific effects of prenatal stress on hypothalamic-pituitary-adrenal responses to stress and brain glucocorticoid receptor density in adult rats. Brain Res 1995;84: 55-61.
161. Weinstock M. Gender differences in the effects of prenatal stress on brain development and behaviour. Neurochem Res 2007;32:1730-40.
162. Morgan CP, Bale TL. Early prenatal stress epigenetically programs dysmasculinization in second-generation offspring via the paternal lineage. J Neurosci 2011;31:11748-55.
163. Carpenter T, Grecian S, Reynolds R. Sex differences in early-life programming of the hypothalamic-pituitary-adrenal axis in humans suggest increased vulnerability in females: a systematic review. J Dev Orig Health Dis 2017;8:244-55.
164. Miska EA, Ferguson-Smith AC. Transgenerational inheritance: models and mechanisms of non-DNA sequence-based inheritance. Science 2016; 354:59-63.
165. Cortessis VK, Thomas DC, Levine AJ et al. Environmental epigenetics: prospects for studying epigenetic mediation of exposure-response relationships. Hum Genet 2012;131:1565-89.
166. Faulk C, Dolinoy DC. Timing is everything: the when and how of environmentally induced changes in the epigenome of animals. Epigenetics 2011;6:791-7.
167. Bader HN, Bierer LM, Lehrner A et al. Maternal age at holocaust exposure and maternal PTSD independently influence urinary cortisol levels in adult offspring. Front Endocrinol 2014;5:103.
168. Jirtle RL, Skinner MK. Environmental epigenomics and disease susceptibility. Nat Rev Genet 2007;8:253-62.
169. Guerrero-Bosagna C, Skinner MK. Environmentally induced epigenetic transgenerational inheritance of phenotype and disease. Mol Cell Endocrinol 2012;354:3-8.
170. Rodgers AB, Bale TL. Germ cell origins of posttraumatic stress disorder risk: the transgenerational impact of parental stress experience. Biol Psychiatry 2015;78:307-14.
171. McCarthy MM, Auger AP, Bale TL et al. The epigenetics of sex differences in the brain. J Neurosci 2009;29:12815-23.
172. Mirbahai L, Chipman JK. Epigenetic memory of environmental organisms: a reflection of lifetime stressor exposures. Mutat Res 2014;764-765:10-7.
173. Rando OJ. Daddy issues: paternal effects on phenotype. Cell 2012;151: 702-8.
174. Desai N, Ludgin J, Sharma R et al. Female and male gametogenesis. In: Falcone T, Hurd WW (eds). Clinical reproductive medicine and surgery. New York: Springer, 2017:43-62.
175. Dias BG, Ressler KJ. Parental olfactory experience influences behavior and neural structure in subsequent generations. Nat Neurosci 2014;17: 89-96.
176. Gapp K, Jawaid A, Sarkies P et al. Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. Nat Neurosci 2014;17:667-9.
177. Guerrero-Bosagna C, Weeks S, Skinner MK. Identification of genomic features in environmentally induced epigenetic transgenerational inherited sperm epimutations. PLoS One 2014;9:e100194.
178. Milekic MH, Xin Y, O'Donnell A et al. Age-related sperm DNA methylation changes are transmitted to offspring and associated with abnormal behavior and dysregulated gene expression. Mol Psychiatry 2015;20:995-1001.
179. Rodgers AB, Morgan CP, Bronson SL et al. Paternal stress exposure alters sperm microRNA content and reprograms offspring HPA stress axis regulation. J Neurosci 2013;33:9003-12.
180. Jenkins TG, Carrell DT. The sperm epigenome and potential implications for the developing embryo. Reproduction 2012;143:727-34.
181. Szyf M. Nongenetic inheritance and transgenerational epigenetics. Trends Mol Med 2015;21:134-44.
182. Schagdarsurengin U, Steger K. Epigenetics in male reproduction: effect of paternal diet on sperm quality and offspring health. Nat Rev Urol 2016;13: 584-95.
183. Fullston T, Teague EMCO, Palmer NO et al. Paternal obesity initiates metabolic disturbances in two generations of mice with incomplete penetrance to the F2 generation and alters the transcriptional profile of testis and sperm microRNA content. FASEB Journal 2013;27:4226-43.
184. Anway MD, Cupp AS, Uzumcu M. Epigenetic transgenerational actions of endocrine disruptors and male fertility. Science 2005;308:1466-9.
185. Friedler G. Paternal exposures: impact on reproductive and developmental outcome. An overview. Pharmacol Biochem Behav 1996;55:691-700.
186. Cicero TJ. Effects of paternal exposure to alcohol on offspring development. Alcohol Res Health 1994;18:37.
187. Ji BS, Shu XO, Zheng W et al. Paternal cigarette smoking and the risk of childhood cancer among offspring of nonsmoking mothers. J Natl Cancer Inst 1997;89:238-43.
188. Chong S, Whitelaw E. Epigenetic germline inheritance. Curr Opin Genet Dev 2004;14:692-6.
189. Lane N, Dean W, Erhardt S et al. Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. Genesis 2003;35:88-93.
190. Reik W, Dean W, Walter J. Epigenetic reprogramming in mammalian development. Science 2001;293:1089-93.
191. Buiting K, Groß S, Lich C et al. Epimutations in Prader-Willi and Angelman syndromes: a molecular study of 136 patients with an imprinting defect. Am J Hum Genet 2003;72:571-7.
192. Lawson HA, Cheverud JM, Wolf JB. Genomic imprinting and parent-of-origin effects on complex traits. Nat Rev Genet 2013;14:609-17.
193. Wolf JB, Hager R, Cheverud JM. Genomic imprinting effects on complex traits: a phenotype-based perspective. Epigenetics 2008;3:295-9.
194. Kaati G, Bygren LO, Pembrey M et al. Transgenerational response to nutrition, early life circumstances and longevity. Eur J Hum Genet 2007; 15:784.
195. Pembrey ME, Bygren LO, Kaati G et al. Sex-specific, male-line transgenerational responses in humans. Eur J Hum Genet 2006;14:159.
196. Kaati G, Bygren LO, Edvinsson S. Cardiovascular and diabetes mortality determined by nutrition during parents' and grandparents' slow growth period. Eur J Hum Genet 2002;10:682.
197. Veenendaal MV, Painter RC, Rooij S et al. Transgenerational effects of prenatal exposure to the 1944-45 Dutch famine. BJOG 2013;120:548-54.
198. Boschen KE, Keller SM, Roth TL et al. Epigenetic mechanisms in alcohol- and adversity-induced developmental origins of neurobehavioral functioning. Neurotoxicol Teratol 2018;66:63-79.
199. Ouko LA, Shantikumar K, Knezovich J et al. Effect of alcohol consumption on CpG methylation in the differentially methylated regions of H19 and IG-DMR in male gametes – Implications for fetal alcohol spectrum disorders. Alcohol Clin Exp Res 2009;33:1615-27.
200. Chorbov VM, Todorov AA, Lynskey MT et al. Elevated levels of DNA methylation at the OPRM1 promoter in blood and sperm from male opioid addicts. J Opioid Manag 2011;7:258.
201. Hamad MF, Shelko N, Kartarius S et al. Impact of cigarette smoking on histone (H2B) to protamine ratio in human spermatozoa and its relation to sperm parameters. Andrology 2014;2:666-77.
202. Marczylo EL, Amoako AA, Konje JC et al. Smoking induces differential miRNA expression in human spermatozoa: a potential transgenerational epigenetic concern? Epigenetics 2012;7:432-9.
203. Braun K, Champagne FA. Paternal influences on offspring development: behavioural and epigenetic pathways. J Neuroendocrinol 2014;26:697-706.
204. Dietz DM, Laplant Q, Watts EL et al. Paternal transmission of stress-induced pathologies. Biol Psychiatry 2011;70:408-14.
205. Gapp K, Bohacek J, Grossmann J et al. Potential of environmental enrich-

ment to prevent transgenerational effects of paternal trauma. Neuropsychopharmacology 2016;41:2749-58.
206. Mychasiuk R, Harker A, Ilnytskyy S et al. Paternal stress prior to conception alters DNA methylation and behaviour of developing rat offspring. Neuroscience 2013;241:100-5.
207. Lim JP, Brunet A. Bridging the transgenerational gap with epigenetic memory. Trends Genet 2013;29:176-86.

DOI:10.1002/wps.20568

The severity of psychiatric disorders

Mark Zimmerman, Theresa A. Morgan, Kasey Stanton

Department of Psychiatry and Human Behavior, Brown University School of Medicine, Rhode Island Hospital, Providence, RI, USA

The issue of the severity of psychiatric disorders has great clinical importance. For example, severity influences decisions about level of care, and affects decisions to seek government assistance due to psychiatric disability. Controversy exists as to the efficacy of antidepressants across the spectrum of depression severity, and whether patients with severe depression should be preferentially treated with medication rather than psychotherapy. Measures of severity are used to evaluate outcome in treatment studies and may be used as meaningful endpoints in clinical practice. But, what does it mean to say that someone has a severe illness? Does severity refer to the number of symptoms a patient is experiencing? To the intensity of the symptoms? To symptom frequency or persistence? To the impact of symptoms on functioning or on quality of life? To the likelihood of the illness resulting in permanent disability or death? Putting aside the issue of how severity should be operationalized, another consideration is whether severity should be conceptualized similarly for all illnesses or be disorder specific. In this paper, we examine how severity is characterized in research and contemporary psychiatric diagnostic systems, with a special focus on depression and personality disorders. Our review shows that the DSM-5 has defined the severity of various disorders in different ways, and that researchers have adopted a myriad of ways of defining severity for both depression and personality disorders, although the severity of the former was predominantly defined according to scores on symptom rating scales, whereas the severity of the latter was often linked with impairments in functioning. Because the functional impact of symptom-defined disorders depends on factors extrinsic to those disorders, such as self-efficacy, resilience, coping ability, social support, cultural and social expectations, as well as the responsibilities related to one's primary role function and the availability of others to assume those responsibilities, we argue that the severity of such disorders should be defined independently from functional impairment.

Key words: Severity, psychiatric disorders, functional impairment, symptoms, depression, personality disorders, transdiagnostic models, HiTOP, DSM-5, ICD-10

(World Psychiatry 2018;17:258–275)

The determination of illness severity has important clinical implications. Depending on the disorder, severity affects decisions to seek treatment, the type and intensity of treatment, and whether to continue or stop treatment. Severity also impacts expectations in the fulfillment of role function and disability status. Measures of severity are used to evaluate outcome in treatment studies and may be used as meaningful endpoints in clinical practice.

But, what does it mean to say that someone has a severe illness? Of the various dictionary definitions of "severe", the one that is most relevant to the characterization of illness is "of great degree". This definition, however, does not convey what is meant when an illness is considered "severe". Does severity refer to the number of symptoms a patient is experiencing? To the intensity of the symptoms? To symptom frequency or persistence? To the impact of symptoms on functioning or quality of life? To the likelihood of the illness resulting in permanent disability or death?

Some of these questions about the meaning of severity can be further elaborated. For example, with regards to the prediction of mortality, does severity allude to imminent death, death in the near future, or death at any time in the future? Also, should the impact of intervention be considered? That is, is an illness severe only when death is likely if the illness is left untreated, or only if death is likely regardless of intervention?

Perhaps severity determinations should be independent of functional impact or prognosis and instead should be based on structural or morphological changes and damage to the diseased organ. To be sure, this is not relevant for many illnesses, but, when it can be measured, should this be the guiding principle for rating illness severity?

Putting aside the issue of how severity should be operationalized, another consideration is whether severity should be conceptualized similarly for all illnesses or be disorder specific. Should the severity of heart failure, rheumatoid arthritis, diabetes, an acute upper respiratory tract infection, and a headache be judged according to a common standard or metric, or should each disorder have its own respective guidelines for rating severity?

In this paper, we examine how severity is characterized in psychiatric research and contemporary psychiatric diagnostic systems. To illustrate some of the issues and controversies in determining the severity of psychiatric disorders, we focus on depression and personality disorders (PDs). The clinical significance of considering the severity of depression is reflected in official treatment guidelines wherein recommendations are based on illness severity[1,2]. The importance of considering the severity of PDs is reflected by the ICD-11 proposal to replace the specified criteria for different disorders by a single personality disorder category that is graded according to levels of severity[3,4].

Before discussing the issue of severity of psychiatric disorders, we present a brief overview of how severity has been conceptualized, assessed and measured for various physical illnesses, highlighting the variability of approaches.

SEVERITY OF PHYSICAL ILLNESSES

There is no consensus or uniform overriding principle in distinguishing between levels of severity of physical illnesses. In some cases, severity is defined by the degree of structural damage to the diseased organ. For example, the severity of rheumatoid arthritis has been defined according to radiographic evidence of joint damage[5]. The severity of diabetic retinopathy has been graded according to the degree of retinal damage assessed in a direct clinical eye exam[6]. In a related manner, physiological measures representing the impact of disease on the organ have been used to characterize the severity of some diseases. For example, left ventricular ejection fraction has been used as an index of the severity of cardiovascular disease[7-10]. Forced expiratory volume has been used as index of severity of cystic fibrosis[11]. Aminotransferase and bilirubin levels have been used to assess the severity of hepatitis[12].

Sometimes severity is defined by a disorder-specific clinical examination. For example, not only have radiographic assessments been used to evaluate the severity of rheumatoid arthritis, but severity has additionally been defined according to a count of the number of swollen and painful joints[13].

Illness severity has also been defined more broadly to encompass indices of the diseased organ as well as related and downstream effects. In a study of the prognostic implications of post-cardiac arrest illness severity, severity scores were based on cardiopulmonary dysfunction and neurologic status[14,15]. The severity of sickle cell disease has been based on the presence and frequency of complications such as renal failure, necrosis of hips and shoulders, and gallstones[16]. In studies of the severity of chronic obstructive pulmonary disease, the BODE index (B, body mass index; O, obstruction of airways as measured by forced expiratory volume in one second; D, dyspnea scale; E, exercise capacity as measured by a six-minute walk test) includes and goes beyond a direct, specific, assessment of pulmonary damage and has been found to be a better predictor of mortality, hospitalization, quality of life, and depression than forced expiratory volume alone[17]. The Unified Parkinson's Disease Rating Scale contains four subscales assessing mental state, activities of daily living, motor examination, and complications[18,19].

Moving further away from a direct or physiological assessment of the diseased organ, the New York Heart Association Functional Classification is a measure of cardiac disease severity based on limitations in physical activities and the presence of physical symptoms associated with varying degrees of activity[20].

In contrast to disorder-specific physical and physiological indicators of severity, there are composite measures of overall illness severity, such as the Acute Physiology and Chronic Health Evaluation (APACHE) scores and the Simplified Acute Physiology Score (SAPS), based on non-specific clinical and biological indicators of health status such as body temperature, age, history of organ failure, electrolytes, and hematocrit[21,22]. These illness severity measures have been used to predict mortality in heterogeneous and single disorder samples of acutely ill emergency department and hospitalized patients[23,24].

Finally, self-report questionnaires have been developed to assess the severity of some physical illnesses. The severity of benign prostatic hypertrophy as assessed by the American Urological Association Symptom Index is based on the frequency of symptoms[25]. The Tinnitus Severity Index is based on the frequency of functional impairment or psychological symptoms due to tinnitus[26]. The Bowel Symptom Severity Scale assesses the frequency, distress and disability of symptoms associated with irritable bowel syndrome[27]. The severity of headaches as measured by the Headache Impact Questionnaire is a composite measure of headache frequency, the average pain intensity of headaches, and the impairment resulting from headaches[28]. The Liverpool Seizure Severity Scale assesses perceptions of seizure control and severity of ictal and postictal symptoms[29].

Clark et al[30] summarized the approach taken to develop self-report measures of illness severity for six disease states studied in the Veterans Health Study. They defined illness severity in terms of patients' perceptions of the magnitude of symptoms or complications of the illness that are associated with reductions in health-related quality of life or health status. They distinguished disease severity from the impact of disease (e.g., impairment, life satisfaction, well-being), because the impact of disease is often mediated by personal characteristics (e.g., resiliency, self-efficacy) and social context.

SEVERITY OF PSYCHIATRIC DISORDERS AS DESCRIBED IN DSM-5

In contrast to some physical illnesses, there are no specific or non-specific biomarkers of psychiatric disorders that validly characterize the severity of the disorder. In the absence of such biological or structural indicators, researchers and clinicians are left to assess the epiphenomena of a psychiatric disorder to judge its severity.

Discussions of resource allocation in the public health sector often focus on patients with severe mental illness, though there is no consensus in how to define such an illness[31,32]. The DSM-5[33], like its immediate predecessors, defines severity for only some disorders. Table 1 lists the DSM-5 disorders with defined levels of severity.

The DSM-5 approach towards defining severity varies across disorders. The four severity levels of intellectual disability (mild, moderate, severe, profound) are the most elaborately defined, with three pages of descriptions of the adaptive functioning deficits characteristic of each level of severity. DSM-5 notes that severity was defined according to adaptive functioning level rather than IQ scores because the former is a better determinant of the level of supports that are needed. Similarly, the level of deficits and functional impairment defining the severity of autism spectrum disorders is linked to the supports required. The severity of learning disorders refers

Table 1 Characterization of disorder severity in DSM-5

DSM-5 disorder	Features used to define severity
Major depressive disorder	Number of symptoms, level of distress caused by intensity of symptoms, and impairment in social and occupational functioning
Mania, hypomania	Same as major depressive disorder
Alcohol use disorder	Number of criteria
Drug use disorder	Number of criteria
Bulimia nervosa	Frequency of compensatory behaviors per week
Anorexia nervosa	Body mass index
Binge eating disorder	Frequency of eating binges
Learning disorders	Severity of deficit in learning skills and likelihood of learning the skills with or without intervention
Attention-deficit/hyperactivity disorder	Number of symptoms, severity of individual symptoms, or level of impairment caused by the symptoms
Intellectual disability	Level of adaptive functioning
Autism spectrum disorder	Degree of impairment in functioning due to deficits in verbal and nonverbal communication, inflexibility of behavior, difficulty coping with change, or restricted/repetitive behaviors
Stereotypic movement disorder	The ease by which the symptoms can be suppressed and the need for intervention to prevent serious injury
Psychotic disorders	Quantitative assessment on 5-point scale of primary feature of the psychosis (delusions, hallucinations, disorganized speech, abnormal psychomotor behavior, and negative symptoms). Rating is based on symptom intensity or subjective distress due to symptom
Reactive attachment disorder	Only the severe type is defined. Severe is defined as all criteria met at a high level
Disinhibited social engagement disorder	Only the severe type is defined. Severe is defined as all criteria met at a high level
Somatic symptom disorder	Number of criteria and somatic complaints
Psychological factors affecting other medical conditions	Degree of impact on medical condition or medical risk
Hypersomnolence disorder	Number of days per week with difficulty maintaining daytime alertness
Narcolepsy	Frequency of cataplexy and responsiveness of cataplexy to medication, number of naps per day, degree of disturbance of nocturnal sleep
Obstructive sleep apnea/hypopnea	Apnea/hypopnea index score
Nightmare disorder	Frequency of nightmares per week
Sexual disorders	Degree of distress related to symptoms
Premature ejaculation	Time to ejaculation
Substance/medication-induced sexual dysfunction	Percentage of occasions of sexual activity that dysfunction occurs
Oppositional defiant disorder	Number of settings in which the symptoms occur
Conduct disorder	Number of conduct problems or the degree of harm caused to others
Neurocognitive disorders	Degree of difficulty with instrumental activities of daily living

to the difficulties in learning skills as well as the likelihood of learning those skills with or without intervention. For example, DSM-5 defines severe impairment of a learning disorder as "severe difficulties learning skills, affecting several academic domains, so that the individual is unlikely to learn those skills without ongoing intensive individualized and specialized teaching for most of the school years". For these disorders, then, the severity specifier is explicitly linked to suggested levels of intervention.

Depression and mania are classified as mild, moderate or severe according to the number of symptoms, the level of distress caused by the intensity of the symptoms, and the degree of impairment in social and occupational functioning. The severity of alcohol and drug use disorders is based on the number of criteria that are met (mild: 2 or 3 criteria; moderate: 4 or 5 criteria; severe: 6 or more criteria). The severity of attention-deficit/hyperactivity disorder is based on the number of symptoms, severity of individual symptoms, or level of impairment caused by the symptoms. The severity of bulimia nervosa is operationalized according to the number of inappropriate compensatory behaviors per week (mild: 1-3; moderate: 4-7; severe: 8-13; extreme: 14 or more), though the severity designation could be increased to reflect other symptoms or level of functional impairment. For anorexia nervosa, severity is defined according to body mass index, and for binge eating disorder it is defined by the number of binge eating episodes per

week, though, similar to bulimia nervosa, the severity designation can be increased to reflect other symptoms or degree of functional impairment. Severity of sexual disorders is based on the level of distress regarding the symptoms, except for premature ejaculation, for which severity is based on the time to ejaculation. The severity of cataplexy is based, in part, on lack of responsiveness to medication.

This brief overview illustrates the variability in the approaches taken in the DSM-5 towards defining degrees of severity, with some definitions emphasizing the number of criteria met, some others emphasizing the core feature of the disorder, some based on level of distress, and some focusing on response to intervention and prediction of course. In contrast to many physical illnesses, none of the definitions of severity refer to the likelihood of imminent or distal mortality, and most definitions do not refer to prognosis or future course. Rather, most definitions of severity in DSM-5 refer to the number of symptoms or criteria of the disorder, the frequency of symptoms, and the level of impairment or distress.

SEVERITY OF DEPRESSION

We focus on the severity of depression because it has received the most extensive research. While the research has not been entirely consistent, the severity of depression has been associated with health-related quality of life[34], functional impairment[35,36], suicidality[37-39], longitudinal course[40-43], and several biological variables[44-46]. Moreover, the severity of depression has been at the core of controversies regarding the efficacy of treatment and whether certain forms of treatment should be recommended as first line interventions. Almost all research on severity is based on scores on depression symptom scales, though most scales have been developed without consideration as to how to best conceptualize and assess the severity of depression.

Severity levels of depression in DSM-5 and ICD-10

Three elements are used to define the severity levels of depression in DSM-5: the number of symptoms, the level of distress caused by the intensity of the symptoms, and the degree of impairment in social and occupational functioning. The severity categorization applies to all depressive disorders, not just major depressive disorder (MDD). Mild depression is specified when "few, if any, symptoms in excess of those required to make the diagnosis are present, the intensity of the symptoms is distressing but manageable, and the symptoms result in minor impairment in social or occupational functioning". Severe depression is specified when "the number of symptoms is substantially in excess of that required to make the diagnosis, the intensity of the symptoms is seriously distressing and unmanageable, and the symptoms markedly interfere with social and occupational functioning". The DSM-5 does not explicitly define moderate depression other than to say that the number of symptoms, intensity of symptoms, and/or functional impairment are between mild and severe.

There are some problems with the DSM-5 specification of severity levels. The same definition of the severity specifier is used for MDD and persistent depressive disorder. This is a problem, because persistent depressive disorder requires fewer symptoms than does MDD to meet the DSM-5 diagnostic threshold. Thus, a patient with persistent depressive disorder who experiences the same number of symptoms as a patient with MDD, and with similar levels of functional impairment and distress, may be classified as more severe because the symptom count may be "substantially in excess" of the diagnostic threshold for persistent depressive disorder but not for MDD.

Another problem with the DSM-5 severity specifier is that the definition of functional impairment is limited to social or occupational functioning. This is inconsistent with the wording of the impairment criterion for the diagnosis of MDD and persistent depressive disorder, which refers to impairment in social, occupational, or *other important areas of functioning*. Thus, individuals who maintain social contacts, are not expected to be employed, but are unable to function as students or full-time parents, could be misclassified as less severe than they actually are.

While moderate severity is not specifically defined, the internal logic of the wording of the moderate severity description has a minor flaw. Mild depression requires low levels of symptoms, distress *and* functional impairment. Conversely, severe depression requires high levels of all three. Thus, moderate depression should be defined as lying between the mild and severe levels in symptoms, distress *or* functional impairment (not *and/or* as DSM-5 defines it).

Finally, two other variables often considered important in discussions about depression severity – suicidality and need for hospitalization – are not considered in DSM-5's definition of severity.

What evidence supports the validity of the DSM-5 approach towards defining severity in this manner? One study from a population-based registry of twins who experienced a major depressive episode in the year prior to the interview found that the three aspects of the severity specifier – number of symptoms, severity of symptoms, and degree of functional impairment – were significantly, albeit only modestly, correlated[47]. The authors concluded that the DSM severity construct was multifaceted and heterogeneous.

A study of psychiatric outpatients with a mood disorder[48], 84% of whom were in a major depressive episode, found that the number of DSM-IV symptoms of MDD was weakly correlated with clinicians' ratings on the Clinical Global Impression (CGI)[49] and the Global Assessment of Functioning (GAF)[50]. Moreover, the severity ratings of some individual symptoms of depression were as highly correlated with CGI and GAF scores as was the total number of depressive symptoms. A small study of psychiatric inpatients with MDD found that the number of MDD criteria was weakly correlated with the Glob-

al Assessment Scale[51]. Kessler et al[52] analyzed data from the National Comorbidity Study (NCS) and found that, compared to individuals who reported five or six MDD criteria during their worst episode of depression, individuals who reported seven to nine MDD criteria experienced more psychosocial impairment, more episodes of depression, and greater chronicity. Wakefield and Schmitz[53,54] examined the NCS database as well as another epidemiological survey and suggested that the number of depressive symptoms was less important than the type of depressive symptoms and other features of complicated depression in predicting future occurrence of a major depressive episode, seeking professional help for depression, a history of suicide attempt, and a history of psychiatric hospitalization. Thus, symptom count does not seem to be an adequate indicator of depression severity.

The ICD-10[55] designates three levels of severity – mild, moderate and severe – based on number of symptoms, severity of symptoms, functional impairment, level of distress and, indirectly, type of symptoms. In contrast to DSM-5, there is no symmetry in the descriptions of the three levels of severity. Mild depression refers to the presence of two or three symptoms that are distressing though the patient is likely to be able to continue with most activities. Moderate depression requires four or more symptoms with the patient having great difficulty to continue with ordinary activities. Severe depression requires "several symptoms that are marked and distressing, typically loss of self-esteem and ideas of worthlessness or guilt. Suicidal thoughts and acts are common and a number of 'somatic' symptoms are usually present".

As with the definition of the DSM-5 severity specifier, little research has been done on the ICD-10 severity specifier, perhaps because the reliability of making the severity distinctions is poor[56]. Poor reliability is not surprising, due to the impreciseness of the severity level definitions[57].

The severity definitions in the official diagnostic systems have not been used in treatment studies. Rather, in almost all those studies, severity is designated by a score on a symptom rating instrument – usually the Hamilton Depression Rating Scale (HAMD)[58] or the Montgomery-Åsberg Depression Rating Scale (MADRS)[59]. Thus, treatment studies generally do not consider other factors that have been used to characterize severity, such as level of functional impairment, degree of suicidality, or depressive subtype (i.e., presence of melancholic features or psychotic symptoms)[60,61].

Scales measuring the severity of depression

The severity of depression has been most frequently quantified on paper-and-pencil and clinician-administered rating scales. There is variability amongst the instruments in the time frame covered (the two most common time frames being the past one or two weeks), rating guidelines (most scales use Likert-type ratings based on symptom frequency, persistence or intensity), and item content.

Little research has examined which parameters provide the most valid indicator of depression severity. Is the severity of depression best conceptualized as the number of symptoms (i.e., present or absent), frequency of symptoms (e.g., every day vs. half the days vs. few days), persistence of symptoms (e.g., always present vs. often present vs. sometimes present), or intensity of symptoms (e.g., severe vs. moderate vs. mild)? Williams et al[62], in standardizing the scoring of the HAMD, created a grid scoring format to incorporate information regarding symptom frequency/persistence and intensity in the ratings. The only study to examine whether it is important to consider both intensity and frequency constructs found that symptom intensity was a better indicator of severity than symptom frequency[63]. In developing the Patient-Reported Outcomes Measurement Information System (PROMIS) depression scale, Pilkonis et al[64] reviewed studies comparing alternative response options and concluded that frequency scaling outperformed intensity ratings, though these were not studies of depression ratings. Thus, the most valid rating format of depression severity scales is unsettled, and has been little studied.

Should the content of a severity scale be based on the diagnostic criteria for the disorder, include other symptoms of depression that are not components of the diagnostic criteria (e.g., low motivation), or include symptoms that are frequent in depressed patients but are defining features of other disorders (e.g., anxiety, irritability)? And by what standard should one judge whether one approach or scale is a more valid indicator of severity? Statistical approaches such as item response theory have been used to construct scales[65,66]. While instruments derived from this approach may be psychometrically superior to measures based on the diagnostic criteria for MDD, such measures do not include symptoms that have long been considered to be core components of depression, such as appetite and sleep disturbances or suicidality. If a measure of severity is to be utilized for clinical purposes, and not just for administrative outcome measurement, it is important to include vegetative symptoms, as the presence of these symptoms affects medication selection[67], and to assess suicidality because of safety concerns.

While there are differences amongst the scales in how they were constructed, their intended purpose, item coverage, and rating guidelines, the one commonality is that the overall severity of depression is represented by the sum of the ratings of the individual items. For all but a few scales, all items on the scale are rated similarly and contribute equally to the total score. A notable exception is the HAMD[58], which includes some items rated 0 to 2, and some others rated 0 to 4. To be sure, measures differ in their emphasis on different content domains of depression[68]. Some measures have been criticized as being multidimensional, because a unidimensional construct of depression severity is better able to demonstrate treatment effects[69]. However, all scales, even multidimensional measures which yield subscale scores, as well as instruments that were initially intended to screen for depression ra-

ther than being used as indicators of severity, derive a total score that has been used to denote the severity of depression.

The *score summation approach* is based on some assumptions that have not been empirically supported. Adding up item scores to yield a total score as an indicator of overall depression severity assumes that all symptoms are equal indicators of the severity of depression. However, the different symptoms of depression are not similarly correlated with clinicians' global ratings of severity[48]. From the psychometric perspective, the rating options of individual items should convey valid information across the entire spectrum of severity[70]. Thus, severely depressed patients should more frequently receive the highest rating of a symptom than a low or zero rating, whereas mildly depressed patients should more frequently receive ratings indicating mild severity than the highest rating of a symptom. Santor and Coyne[70], using item response theory data analytic techniques, demonstrated that some of the items of the HAMD do not meet these assumptions.

In fact, scales based on item frequency ratings are unlikely to meet these assumptions and therefore may not be good measures of severity. For example, the items on the 9-item Patient Health Questionnaire (PHQ-9) are rated on a four-point scale of symptom frequency during the past two weeks: (0=not at all, 1=several days; 2=more than half the days; 3=nearly every day)[71]. Patients with MDD would be expected to score a 3 for most of the symptoms that are present, because the definition of MDD requires symptom presence for at least two weeks. Because of the ceiling effect, a patient with MDD seen in primary care who continues to work would score similarly to a depressed patient who is hospitalized because of difficulties with self-care. While there are several studies of the PHQ-9 using an item response theory approach, these have been of heterogeneous non-depressed psychiatric, medical or community samples[72-78]. We are unaware of any studies evaluating the performance of the PHQ-9 items in a sample of depressed patients presenting for treatment. We would predict that, in such a sample, some – perhaps many – items of the PHQ-9 would be highly skewed because of the aforementioned ceiling effect. No studies have examined the impact of different rating guidelines on the operating characteristics of items on a depression scale.

Implicit in the score summation approach is that low level ratings across many symptoms reflect equal severity to high ratings across a fewer number of symptoms. For example, someone who indicates that, in the past week, he/she has infrequently experienced low mood, insomnia, low self-esteem, guilt, reduced concentration, fatigue, psychomotor slowing, insomnia, reduced appetite, reduced concentration, impaired decision making, and reduced interest in usual activities would be considered at the same level of severity as someone who reports daily depressed mood, guilt, feelings of inferiority, and suicidal thoughts, but denies all somatic and vegetative symptoms of depression. Likewise, when item ratings are based on symptom intensity, a mild intensity rating of many symptoms is considered the same as a severe intensity rating of a more limited number of symptoms.

The score summation approach, in which all items are weighted equally, is not grounded in a specific overriding conceptualization of severity. If illness severity is conceptualized in terms of mortality risk, then one would expect a measure of depression severity to weight more heavily item ratings of suicidal thoughts, hopelessness and psychomotor agitation than ratings of impaired concentration and fatigue. On the other hand, if illness severity is conceptualized in terms of functional impairment, then one might expect items assessing impaired concentration and fatigue to be weighted more heavily than items assessing appetite reduction or guilt. To be sure, some measures assess functional impairment along with symptomatology[63,71,79-81]. No symptom-based measure, however, has been constructed by examining the association of individual items with indices of functional impairment and including on the scale only those items that are independently associated with impairment.

Few studies have examined the association between severity ratings of individual symptoms of depression and multiple external indicators of severity. Faravelli et al[48] found marked differences among symptoms in their association with CGI and GAF ratings. Moreover, the symptoms with the highest correlations with CGI ratings – such as depressed mood, psychic retardation, impaired concentration, and anhedonia – tended to have the highest correlations with GAF scores.

Most discussions of the problems with depression scales have focused on their limitations as outcome measures[69,82,83]. However, different aspects of outcome measurement may be of interest, and these differences might result in different approaches towards scale construction. Some measures of the severity of depression have been specifically designed to be sensitive to treatment effects[59,84]. Some measures are linked to the symptom criteria that are used to diagnose depression[71,79,85,86], whereas others assess a broad range of features that patients indicate are most important in measuring outcome[80] or assess a range of diagnostic and associated symptoms of depression[87]. Descriptions of scale construction typically focus on the content of the measure and rarely discuss the reason for choosing the rating format. For example, in developing the Multidimensional Depression Assessment Scale, Cheung and Power[68] reviewed the content of fifteen depression scales and how their scale would address a content gap. There was no discussion, however, of rating formats and why a symptom frequency format was chosen for their measure rather than a rating format assessing symptom intensity.

One of the commonly used clinician rated measures of severity, the MADRS, was designed to be particularly sensitive to change in treatment trials[59]. Items were selected if they were prevalent in the patients at the beginning of treatment (i.e., prevalence greater than 70%), showed the greatest change from baseline to week 4 of treatment, and change in scores from baseline to week 4 on the symptom showed the greatest correlation with change in total scores on the measure. While there is nothing inherently wrong with constructing a measure in this manner for this purpose, this should not be the basis for select-

ing items on a measure of depression severity, as the resulting scale can be biased towards the inclusion of items that are particularly sensitive to change for the medication(s) studied. The construction of the MADRS was based on response to mianserin, maprotiline, amitriptyline, and clomipramine – medications that are not commonly used today. Using the same approach to construct a measure today, when different medications are prescribed, might produce a scale that only partially overlaps with the items included on the MADRS. In the same vein, the HAMD, which was published more than 50 years ago, has been criticized for including items that are most responsive to the effects of sedating medications such as tricyclic antidepressants[88].

So, while there are many rating scales of depression, and several studies examining them, questions remain as to how to judge if one measure is a more valid indicator of depression severity than another measure. Should it be based on psychometric analyses indicating unidimensionality? Would a "better" measure of severity be more highly correlated with indices of impairment? Be more highly correlated with current suicidal ideation? Be more highly predictive of future suicidal behavior? Be more highly predictive of future mortality in general? Be more highly predictive of future course? Be better able to distinguish depressed patients who do and do not require hospitalization? Demonstrate a larger effect size in a treatment study? Have greater discriminative ability between depression and anxiety, and thus be a "purer" measure of depression?

A problem with depression scales: uncertain validity of cutoffs to define severity groupings

Putting aside the question of how to best conceptualize severity and construct a scale, a problem with the existing literature on depression severity is the inconsistency in the cutoff scores on symptom scales used to demarcate levels of severity, particularly severe depression. The use of various cutoff scores to define severity groups makes it difficult to compare the studies on the treatment implications of severity.

DeRubeis et al[89] conducted a mega-analysis of four studies comparing cognitive-behavioral therapy and medication, and defined severe depression as a cutoff of 20 or more on the 17-item HAMD. Likewise, the recent mega-analysis of placebo-controlled trials of fluoxetine and venlafaxine used a cutoff of 20 to define severe depression[90]. Both of these studies cited the landmark study by Elkin et al[91] to justify their definition of severe depression. However, Elkin et al did not cite empirical evidence for this cutoff and, in fact, did not refer to the patients scoring above 20 on the HAMD in absolute terms (i.e., having severe depression), but instead referred to these patients in relative terms (i.e., having more severe depression than the patients scoring 20 and below).

In Kirsch et al's[92] meta-analysis of the impact of severity on antidepressant-placebo differences, the authors noted that the mean baseline HAMD scores of the antidepressant efficacy trials were in the very severe range (i.e., > 23) based on the American Psychiatric Association (APA)'s Handbook of Psychiatric Measures[93] for all but two of the 35 studies included in the analysis. In a prior analysis of antidepressant efficacy studies in the Food and Drug Administration (FDA) data base, Khan et al[94] divided the studies into three groups based on pre-treatment HAMD scores (<24, 25-27, >28) without indicating the basis for using these cutoff scores to define the groups. Fournier et al[95] used the thresholds recommended in the APA's Handbook of Psychiatric Measures[93] to define grades of severity on the HAMD (mild to moderate: <18; severe: 19 to 22; very severe: >23). In contrast to these studies, and the APA guidelines, most pharmacotherapy studies have used a cutoff of 25 on the 17-item HAMD to define severe depression[96-101] and this cutoff has been recommended by several experts[102-104]. Thus, severe depression has not been consistently defined.

Fundamental to studies on the treatment implications of severity levels is the validity of the cutoffs on the HAMD to define the severity categories. In none of the discussion sections of the meta-analyses and pooled analyses of the reports on severity and treatment outcome were questions raised about the cutoffs used to define the grades of severity. The APA's Handbook of Psychiatric Rating Scales[93] cited only two small studies in support of the cutoff scores to identify severity subtypes, and neither study provided support for the APA guidelines. One was a study examining the validity of deriving a HAMD equivalent score on the Schedule for Affective Disorders and Schizophrenia[105]. This study did not attempt to determine the cutoff scores on the HAMD indicating grades of severity. The second study examined the association between HAMD scores and global ratings of severity in 59 depressed inpatients[106]. The authors did not derive (or recommend) cutoff scores corresponding to severity levels. Thus, it is unclear why a cutoff of 19 was recommended in the APA Handbook to identify severe depression. The UK National Institute for Health and Clinical Excellence (NICE) guidelines recommended a cutoff of 23 to identify severe depression on the HAMD, though no research was cited to support this recommendation[107].

Because of the limited amount of empirical research establishing cutoff scores for bands of severity on the HAMD, and the significance accorded to severity by treatment guidelines, our clinical research group also examined this issue in 627 psychiatric outpatients with MDD who were rated on the CGI[108]. The cutoff score on the HAMD that maximized the sum of sensitivity and specificity was 17 for the comparison of mild vs. moderate depression and 24 for the comparison of moderate vs. severe depression. Based on a review of the available evidence, as well as the recommendations that a cutoff of 7 be used to define remission, we recommended the following severity ranges for the 17-item HAMD: no depression (0-7); mild depression (8-16); moderate depression (17-23); and severe depression (>24).

Each of the above studies derived cutoff scores based on clinicians' global judgments of severity. A limitation of these studies is that it is not known on what basis the global judgments of severity were made. Were some symptoms of depres-

sion considered better indicators of severity than other symptoms? For example, are symptoms characteristic of melancholic or endogenous depression given greater weight in clinicians' CGI ratings? Are clinicians' global ratings disproportionately influenced by degree of suicidality? Do clinicians consider psychosocial impairment in making their CGI ratings? We are unaware of any studies that have attempted to derive severity ranges on the HAMD, or any other depression scale for that matter, based on degree of impairment or level of suicidality.

Another problem with depression symptom scales: different scales classify patients into different severity groups

In clinical practice, self-report questionnaires are preferable to clinician-rated scales because they take less time to administer. If self-report scales are to be used to classify patients into severity categories, and if treatment recommendations are to be based, in part, on severity classification, then it is important for different scales to classify individuals similarly. However, because the content of measures differ, it would not be surprising if there were significant differences between measures.

Cameron et al[109] compared the PHQ-9 and the Hospital Anxiety and Depression Scale (HADS) severity classifications in a sample of primary care patients referred by their general practitioners in the UK to a mental health worker[110]. No information was provided regarding the patients' psychiatric diagnoses. They found that the PHQ-9 overclassified severity compared to the HADS, with twice as many patients classified in the severe range. Other studies comparing the PHQ-9 and the HADS in medical patients found similar results[111,112]. However, these studies lack an external validator and it is therefore unclear if the PHQ-9 overclassifies, or the HADS underclassifies, severity. A second study by Cameron et al[107] included the second edition of the Beck Depression Inventory (BDI-II)[113] along with the PHQ-9 and HADS, and also assessed the patients with the HAMD. The participants were primary care patients who had been diagnosed by their general practitioner with depression. Both the PHQ-9 and BDI-II overclassified severity compared to the HAMD, whereas the HADS underclassified severity.

We are aware of only one study that compared self-report scales in a sample of psychiatric outpatients with MDD[114]. Our clinical research group compared severity classification on three measures that assess the DSM-IV/DSM-5 symptom criteria for MDD: the Clinically Useful Depression Outcome Scale (CUDOS)[79], the Quick Inventory of Depressive Symptomatology (QIDS)[85], and the PHQ-9[71]. The patients were also rated on the 17-item HAMD. In a study of depressed outpatients, we found that the correlations between the HAMD and all three self-report scale scores were nearly identical, and the average correlation among the three self-report scales was .73. However, the scales significantly differed in their distribution of patients into severity categories. Approximately one-third of the patients scored in the mild range on the HAMD and CUDOS, whereas approximately 10% of the patients were mildly depressed according to the PHQ-9 and QIDS. On the CUDOS and HAMD, moderate depression was the most frequent severity category, whereas on the PHQ-9 and QIDS the majority of the patients were classified as severe. The majority of the patients in the moderate range on the HAMD were in the severe range on the PHQ-9 and QIDS. Significantly fewer patients were classified as severely depressed on the CUDOS compared to the PHQ-9 and QIDS.

With the three self-report measures being highly correlated with each other, and equally correlated with the HAMD, what, then, might account for the marked differences between scales of similar content in the distribution of patients into severity groups?

The cutoffs on the three scales to define the severity groups were derived in different ways, and this was likely responsible for the differences between the scales in severity classification. For example, Kroenke et al[71] indicated that the cutoff scores on the PHQ-9 were chosen for the pragmatic reason of making them easier for clinicians to recall. They also noted that alternative cutoffs did not increase the association between increasing PHQ-9 severity and indices of construct validity. When selecting the cutoff scores to define the severity ranges on the PHQ-9, the developers of this questionnaire did not consider the potential impact of the broadness by which severity ranges were defined and how this might impact on treatment recommendations of official treatment guidelines.

Kroenke et al[71] indicated that, when severity groupings based on different cutoffs are equally associated with external variables, then the cutoffs can be chosen based on their ease of recall. We disagree with this reasoning. For all scales measuring the severity of depressive symptoms, the thresholds distinguishing patients with mild, moderate and severe depression do not represent well-demarcated lines separating the severity subtypes. As with other areas of psychopathology, the severity of depression better corresponds to a dimensional than a categorical model of classification[115]. Thus, alternative cutoffs to categorize severity groupings are likely to also be valid when the groupings are compared on an external variable such as psychosocial functioning. However, one should not be cavalier about the choice of cutoffs, because they impact on the relative broadness of each of the severity categories.

If clinicians are to follow official treatment guidelines' recommendations and base initial treatment selection on the severity of depression, then it is important to have a consistent method of determining depression severity. The marked disparity between standardized self-administered scales in the classification of depressed outpatients into severity groups indicates that there is a problem with the use of such instruments to classify depression severity. If official treatment guideline recommendations were followed, then use of measures such as the QIDS and PHQ-9, which broadly define the severe cat-

egory, would result in greater reliance on medication in preference to psychotherapy as the first line treatment option for MDD. Caution is thus warranted in the use of these scales to guide treatment selection until the thresholds to define severity ranges have been better established empirically.

The importance of severity of depression in treatment: official guideline recommendations

Notwithstanding the aforementioned problems with conceptualizing the severity of depression, and defining the cutoffs on scales for severity levels, depression severity is an important consideration in treatment decision-making. The severity of depression has influenced treatment recommendations in official guidelines. The third edition of the APA's guidelines for the treatment of MDD recommend both psychotherapy and pharmacotherapy as monotherapies for depression of mild and moderate severity, and pharmacotherapy (with or without psychotherapy) for severe depression[1]. The NICE updated guidelines for the treatment and management of depression discourage the use of antidepressant medication as the initial treatment option for mild depression, and recommend medication together with empirically supported psychotherapy for moderate and severe depression[2]. As reported by van der Lem et al[116], treatment guidelines in the Netherlands also recommend pharmacotherapy as the first treatment option for severely depressed patients, and either pharmacotherapy or psychotherapy for mildly and moderately depressed patients. While the recommendations in these guidelines are not entirely consistent, they are unanimous in recommending medication as the treatment of choice for severe depression.

The treatment significance of severity has been studied in several different ways. There are controlled studies, effectiveness studies, pooled analyses, and meta-analyses examining the impact of severity on particular treatments[117-122], comparing treatments across a range of severity[99,123-127], comparing medication and placebo across a range of severity[128,129], comparing psychotherapy and control groups across a range of severity[130,131], comparing treatments amongst severely depressed patients[96,101,102,132], and examining whether severity predicts short-term outcome[42,133-135], treatment resistance[136], longer-term outcome[40,137-139], and relapse[38].

Severity of depression and pharmacotherapy

In the past decade, questions have been raised whether selective serotonin reuptake inhibitors (SSRIs) and other new generation antidepressants are effective in non-severe depression. Khan et al[94] analyzed 45 clinical trials in the FDA database and found that in studies with a mean baseline 17-item HAMD score of 24 or less there was little evidence that antidepressant medication was superior to placebo, whereas in studies with a mean baseline HAMD score of 28 or greater there was clear evidence that medication was superior to placebo. Kirsch et al[92] similarly examined the FDA database, and they also examined the efficacy of antidepressants as a function of mean baseline HAMD score in the trial. Their results largely replicated the findings of Khan et al[94] that drug-placebo differences were largest in the studies with the highest baseline severity (i.e., HAMD >28). Kirsch et al[92] found that antidepressants were significantly more effective than placebo in the less severe cohorts, but they considered the difference in response to be modest and clinically insignificant.

In contrast to the analyses of the FDA database by Kirsch et al[92] and Khan et al[94], Fournier et al[95] pooled individual patient data from six published studies. Kirsch et al and Khan et al used aggregated mean scores for an entire study as the unit of analysis. That is, they compared studies with different mean severity scores at baseline. The problem with this approach is that a group of patients with a mean score in the severe range will also include some patients in the mild and moderate severity ranges. Likewise, a group of patients with a mean score in the mild or moderate severity range will include some patients scoring in the severe range. Pooling individual patient data avoids the problem of severity group misclassification at the individual patient level. Fournier et al[95] replicated the finding that drug-placebo differences were clinically significant only for severely depressed patients, and found only a small effect size for mildly and moderately depressed patients.

More recently, other pooled analyses of patient level data (rather than aggregated data from a trial) have been conducted. Using pharmaceutical company data bases, these analyses included all studies of a product, thereby avoiding the bias inherent in examining only published studies[140]. The results of three large, pooled analyses of published and unpublished studies, which included between 4,000 and 10,000 subjects each, indicated that antidepressants are effective across a range of severity[90,129,141]. These analyses, and the controversy that has been stirred regarding the efficacy of antidepressants, highlights the impact that considerations of severity might have on clinical practice.

Severity of depression and medication or psychotherapy as first line treatment

A second important severity related treatment question is whether the severity of depression should be used as the basis for recommending medication or psychotherapy as first line treatment. More specifically, the question is whether patients with severe depression should preferentially be treated with medication. A related question is whether psychotherapy is beneficial for severely depressed patients.

Symptom severity as a moderator of treatment response has been the subject of ongoing debate since the publication of the results from the US National Institute of Mental Health Treatment of Depression Collaborative Research Program (TDCRP),

suggesting that psychotherapy was not as effective as medication in the acute treatment of severe depression[91,142]. The first meta-analysis of studies directly comparing psychotherapy and pharmacological interventions included 30 published studies of more than 3,000 patients[143]. A meta-regression analysis examining whether effect sizes were associated with mean baseline scores on the HAMD or BDI found no evidence that baseline severity was associated with differential treatment outcome. A comparison of effect sizes in studies with baseline HAMD scores below 20 vs. 20 and above also found no differences.

A meta-analysis of 132 controlled psychotherapy studies of more than 10,000 patients found that greater mean baseline symptom severity did not predict poorer response[130]. More recently, Weitz et al[144] pooled individual patient data from 16 studies comparing antidepressants and cognitive behavior therapy. They defined the severe group according to the APA (HAMD ≥19) and NICE (HAMD >23) recommendations. Increased severity was associated with significantly lower remission rates (but not response rates) in both the medication and psychotherapy treatment conditions. Severity was not associated with differential treatment outcome, thus confirming the results of a prior pooled analysis based on a smaller number of studies[89]. In a follow-up study, the authors conducted a pooled analysis focused on the five studies that used placebo as the control condition[131]. The results were consistent with the larger pooled analysis: baseline symptom severity was not associated with change in symptom severity scores from baseline to endpoint between the cognitive behavior therapy and pill placebo groups.

The results of these more recent meta-analyses, based on severity classification according to symptom rating scales, are thus not consistent with official treatment guidelines which recommend medication as the first line treatment for severe depression.

SEVERITY OF PERSONALITY DISORDERS

Severity is clearly of import to PDs, though the current diagnostic systems do not include any formal severity ratings. PD patients identified as "severe" are more likely to exhibit high comorbidity with other psychiatric diagnoses, particularly mood, anxiety, substance use[145], and other PDs[146]. So-called "severe" cases are often in treatment for protracted periods of time[147-149], exhibit higher rates of hospitalization and suicide attempts[150], and self-injure with greater frequency[151]. They are likely to be incarcerated, unable to hold down a job, and have failed relationships[152]. It is generally agreed that they may present a public health burden, and therefore should be identified early and get treated often[3,4,153].

Nonetheless, the question remains: what is meant by "severe" PD? Severity has been assessed by counting the number of comorbid PD diagnoses overall, with higher comorbidity indicating higher severity[152,154-156]. However, this may better reflect the severity of overall personality pathology rather than the severity of a particular PD. More severe cases of personality pathology may further be identified by case complexity and specific comorbidity patterns. The main section of DSM-5 (i.e., Section II) identifies PDs as occurring in one of three clusters. Tyrer and Johnson[157] proposed that individuals with comorbid PDs from more than one cluster are more severe than those with comorbid PDs from the same cluster. The authors further identify antisocial PD as the most severe PD based on risk to others. Therefore, the most severe cases must be diagnosed with antisocial PD as well as PDs from other clusters. Using this model, severity of PD was associated with conduct disorder, criminal behavior, homelessness, institutionalization, unemployment, and delinquent behavior in childhood.

Severity of a specific PD may be measured by counting the number of criteria met. For example, cases of borderline PD for which nine criteria are endorsed would be viewed as more severe than patients endorsing only five criteria[147]. However, results from our clinical research group did not support this hypothesis, finding no differences in comorbidity or psychosocial functioning based on criteria count for patients diagnosed with borderline PD[158]. Alternatively, severity can be defined by the frequency of symptoms. For instance, patients with borderline PD who engage in self-injury multiple times daily would be more severe than those reporting only monthly self-injury[151].

Specific PDs have even been identified as more or less severe than others. Kernberg and Caligor[159] organized PDs into a hierarchy ranging from "more severe" (e.g., borderline PD) to less severe (e.g., obsessive compulsive PD, dependent PD). There has also been a strong push for conceptualizing PDs using constellations of pathology personality traits. From this perspective, a "severe" PD symptom or trait may be defined as one that is statistically extreme, or existing in only a very small proportion of the population[160].

Treatment research of "severe" personality disorders primarily emphasizes symptom characteristics (frequency, persistence, intensity) and functional impairment (social/occupational, or outcomes such as imprisonment)[161-163]. Maden and Tyrer[162] identify a category of "dangerous and severe" PD, which is characterized by having a high risk of causing unrecoverable harm to others. Confusingly, the first criterion for having a "dangerous and severe" PD is already being diagnosed with a "severe disorder of personality" which remains undefined itself. The authors do not clarify what severity means at the criterion level, although it appears this definition is legal in origin, and refers primarily to psychopathy and not to PDs as they are traditionally defined.

Severity of personality disorders and functioning

Although severity has been defined in various ways in the PD literature, a general consensus appears to have emerged

that PD severity is inherently linked with level of maladaptive functioning[164-169]. It is widely acknowledged that extreme trait or symptom variation is insufficient to diagnose PDs or to dictate diagnostic severity. Rather, the emphasis lies in having extreme personality traits in the presence of impairment associated with those traits. Unlike physical illnesses, or even depression, which are more focused on symptom presentation, personality diagnoses are intertwined with adaptive functioning. Like depression, PDs by definition must result in "distress or impairment" to be diagnosed[33]. In contrast to depression, however, the symptom criteria for diagnosing PDs include both affective/cognitive/emotional *and* functional components. For example, impoverished occupational and financial functioning is included in symptom criteria for antisocial PD, and failure to engage in social and leisure activities is part of the criteria for obsessive-compulsive PD.

The interrelationship between functional impairment and personality leads many to conclude that PD severity is a combination of extreme personality disturbance and maladaptive functioning associated with that disturbance[165,169]. In fact, functioning is so fundamental to determining PD presence and severity that some authors argue that assessing extreme traits/symptoms is unnecessary[170-173]. Thus, one need not demonstrate symptom severity if sufficient impairment is judged to be present. However, the dysfunction must be determined as due to the presence of the personality features, even if they are not extreme. For example, using the multiaxial DSM-IV, Livesley[174] proposed defining PD as present diagnostically on Axis I, and coding personality traits separately on Axis II. Widiger and Trull[169] proposed a similar model, only using the GAF score on Axis V as a stand in for severity.

Taken together, these models converge on defining severity as a generalized, adaptive failure of an intrapsychic system required to fulfill daily life tasks[166]. Although specific areas of impairment differ, there is convergence on impairment in three broad areas: identity formation, self-control (or direction), and interpersonal relationships[164]. However, some research indicates that pathological personality traits and functioning are so closely intertwined that they may not represent distinct domains[175].

Severity of personality disorders as described in DSM-5 and ICD-10

There is no clear mention of severity with respect to PDs in the main section II of DSM-5[33]. However, the overall description of PDs includes severity indicators common to other disorders. For example, PDs are specifically noted to be inflexible, maladaptive, pervasive, and associated with "clinically significant" functional impairment or subjective distress. Functional impairment is an indicator of severity in many physical and psychiatric disorders; pervasiveness is a severity indicator for depression; and subjective distress is identified as indicating a "severe case" for disorders of mood and sexual function. As it stands, there is no official method for indicating PD severity in DSM-5.

Section III (Emerging Measures and Models) of DSM-5 includes an alternative model for diagnosing PDs. Diagnosis is defined via a combination of severity levels of dysfunction and elevated personality traits, and severity is determined principally by dysfunction associated with elevated traits[33]. This model does not designate a measure for overall severity, but "moderate or greater impairment" is required for diagnosis. Impairment is operationalized as falling into one of five levels, with the extreme end indicative of severe personality pathology. The Level of Personality Functioning Scale (LPFS) is proposed to rate impairments in functioning, and therefore also PD severity. Ratings are made for self (identity and self-direction) and interpersonal (empathy and intimacy) functioning. Levels include: 0 (little or no impairment), 1 (some impairment), 2 (moderate impairment), 3 (severe impairment), 4 (extreme impairment). Individuals with extreme impairment are described as having an impoverished, unclear identity and self-direction with maladaptive self-concept, and completely lacking capacity to engage interpersonally.

Interestingly, DSM-5 Section III also includes discussion of an additional measure of personality traits, the Personality Inventory for DSM-5[176]. The items are clearly trait content related; however, the measure provides an overall summed score identified as measuring "overall personality dysfunction". The identification of extreme traits as indicative of dysfunction is curious, but not inconsistent with the significant overlap between functioning and PD traits/symptoms found elsewhere in the literature[175]. Nonetheless, this suggests that extreme traits are at least indicative of extreme dysfunction, which is the primary index of severity in this model.

Similar to the DSM-5, the ICD-10 does not make mention of severity in PD classifications. However, several papers have been published on changes proposed for ICD-11, which are substantial. Most notably, the primary classification of PDs will change to one based on severity of personality disturbance. Description of PD traits or features is optional but will not be required for diagnosis[3,4].

Consistent with the larger literature, the proposed changes to the ICD-11 conceptualize severity primarily as dysfunction, or the personality-related problems experienced by the individual. Again, five levels of severity are proposed, though they vary slightly from those in the DSM. Summed together, severity levels are dictated first by pervasiveness of the impairment (across situations or limited), and secondarily by the number of problematic personality traits (multiple or single). At the highest level of severity, risk to self or others is also assessed. Thus, the most severe cases are identified by functioning above all else. Symptoms/traits and risk of harm are secondary, but also considered. Unlike the DSM-5 alternative model proposal, dysfunction in self and identity is not included in severity ratings[3,4]. At the time of this writing, the ICD-11 has not yet been published, and therefore these definitions should be considered provisional. Nonetheless, the emphasis on func-

tioning via severity ratings has been criticized for insufficient research establishing its reliability and validity[177].

Measures of personality severity

As early as 1996, Tyrer and Johnson[157] developed a five-point scale assessing disorder severity similar to that in the ICD-11 proposal. Ratings were made based on information derived from a trait personality measure, the Personality Assessment Schedule (PAS)[153]. Thus, severity was weighted more towards extremity on traits than on functioning. The PAS has also been used to classify individuals into the four PD categories proposed by Tyrer and Johnson[157]: no PD, personality difficulty, simple PD, complex PD. PAS severity designations are primarily based on the frequency of DSM-IV and ICD-10 categories, and have been used in studies predicting treatment outcomes, albeit with mixed findings[178]. The General Assessment of Personality Disorder[179] has been used as an index of severity in multiple studies, and provides two main scales of severity – self-pathology and interpersonal problems – both of which reflect functional impairment as defined by the DSM-5[164,180,181]. Similarly, the Severity Indices of Personality Problems[173] defines severity as a combination of impoverished self and interpersonal functioning.

Relatively few measures of severity exist for individual PDs, and these largely focus on borderline PD. For example, the Borderline Personality Disorder Severity Index (BPDSI)[151,182] is a semi-structured clinical interview that operationalizes severity primarily by frequency of borderline PD symptom behaviors over the preceding three months. Frequency of symptoms is rated from 0 (never) to 10 (daily). Severity is averaged across these scores, yielding severity scores for individual borderline PD criteria as well as the diagnosis overall. Thus, the BPDSI largely measures severity as a function of symptom frequency, though many of the items also ask about behaviors that have implied functional consequences (e.g., going out instead of working).

Consistent with the severity of personality pathology often being linked with impairments in functioning, PD treatment outcome research has often focused on the degree to which various treatment approaches (e.g., dialectical behavioral therapy, mentalization-based treatments, transference-focused psychotherapy) improve day-to-day functioning and reduce specific, concrete maladaptive behaviors[147,183,184]. For instance, in the extensive borderline PD treatment literature, change in personality pathology is often assessed using measures such as the Zanarini Rating Scale for Borderline Personality Disorder[185] and the Barratt Impulsiveness Scale[186]. However, reduction in suicide attempts, self-harm behavior, and reliance on psychiatric emergency treatment services are often primary treatment outcome measures, as are improvements in maintaining meaningful relationships and improving workplace functioning[147,183,184,187,188].

Although the PD treatment literature has focused primarily on the treatment of borderline PD, other PDs also have received some attention, with functional impairment being identified as central to treatment outcomes. For instance, transference-focused psychotherapy has demonstrated some benefit for patients with comorbid narcissistic and borderline PD, and this treatment approach emphasizes interpersonal functioning in personal and workplace relationships when assessing outcome[189]. Treatment research on antisocial PD has focused on subsequent substance use and arrests[190]. Thus, across the treatment of various PDs, treatment outcome and a reduction in "severity" is understood not just as symptom reduction, but also reduction in specific deleterious behaviors (e.g., self-harm) and the promotion of interpersonal functioning and specific prosocial behaviors (e.g., maintaining employment).

TRANSDIAGNOSTIC MODELS AND SEVERITY: THE EMERGENCE OF PSYCHOPATHOLOGY SPECTRA

Many of the questions asked above as to how to compare the validity of depression scales in measuring severity also apply to determining if different diagnoses confer differential levels of severity. The likelihood of meeting criteria for different diagnoses confers standing on underlying genetic liabilities[191,192]. This is important to consider given that individuals who meet criteria for one diagnosis are very likely to meet criteria for multiple other diagnoses[193], such that various diagnoses may be thought to be manifestations of underlying spectra (e.g., antisocial PD, narcissistic PD and substance use all reflect an underlying externalizing spectrum).

Research examining the relations amongst various internalizing diagnoses characterized by subjective distress and fear suggests that it may be "easier" for individuals to meet criteria for diagnoses such as MDD than for more "severe" disorders such as generalized anxiety and panic disorders[194]. Put differently, meeting criteria for generalized anxiety or panic disorder reflect higher standing on the internalizing dimension than would simply meeting criteria for MDD. Interestingly, Krueger and Finger[194] also found that high standing on the internalizing dimension was linked robustly to lifetime number of inpatient hospitalizations and past month psychosocial functioning.

Other more recent research has also linked "severity" on the internalizing spectrum to key outcomes. For instance, Eaton et al[195] found that the likelihood of meeting criteria for various depressive disorders, anxiety disorders, and bipolar disorders can be represented by an underlying continuum. Individuals with high scores on this dimension, who would be characterized as having more "severe" levels of internalizing psychopathology, would thus be likely to meet criteria for many diagnoses and to report a broad range of symptoms (e.g., depressed mood, worry, concentration difficulties, irritability) characterizing the various DSM diagnoses defining this dimension.

Eaton et al[195] presented evidence indicating that scores on the internalizing spectrum predicted outcomes such as the future occurrence of internalizing symptoms (e.g., depressed mood, worry), suicide attempts, angina/chest pain, and ulcers. Moreover, standing on this underlying dimensionally-based internalizing spectrum predicted these outcomes much more strongly than did DSM-based conceptualizations of various internalizing disorders (e.g., MDD, generalized anxiety disorder), thereby providing evidence for the utility of this approach in capturing severity as it relates to important outcomes such as suicidality and physical health concerns[195].

In regard to other forms of psychopathology, Krueger et al[196] presented evidence indicating that symptoms and behaviors defining personality and substance use disorders can be captured by an underlying externalizing dimension. Other research also supports the presence of this underlying latent externalizing dimension, which explains why antisocial behaviors (e.g., various unlawful behaviors) and traits (e.g., impulsivity, callousness) and substance use issues are likely to co-occur[191,197]. Carragher et al[197] presented findings suggesting that meeting criteria for some disorders (e.g., cocaine dependence) confers higher standing and severity on this underlying externalizing dimension than other "less severe" disorders (e.g., nicotine and alcohol dependence). Similarly, overlap in disorders such as schizophrenia and schizotypal PD appears to be reflected by a thought disorder spectrum[191,198]. Standing on this spectrum has been linked to functional impairment and illness course[198].

Going forward, it will continue to be important for future research to further explicate how level of severity (i.e., how likely an individual is to meet criteria for different disorders and to meet criteria for "difficult" disorders such as cocaine dependence in the case of the externalizing spectrum) captured by broad internalizing, externalizing, and thought disorder dimensions predicts illness course and other key outcomes related to morbidity and mortality. These dimensions account for diagnostic comorbidity amongst various disorders and have been shown to predict various outcomes more strongly than diagnostic status on various DSM disorders, suggesting important merits to this approach[191,195]. In this regard, the Hierarchical Taxonomy of Psychopathology (HiTOP) has emerged as a dimensionally-based alternative to the DSM-5[191,199]. Thus, it will be important to determine the degree to which this framework adequately captures psychopathology "severity", however severity is defined, and is useful for researchers and practitioners.

CONCLUSIONS

The issue of severity has great clinical importance. Severity influences decisions about level of care and affects decisions to seek government assistance due to psychiatric disability. In outpatient settings, the importance of severity is reflected in the controversy about the efficacy of antidepressants across the spectrum of depression severity, and whether patients with severe depression should be preferentially treated with medication rather than psychotherapy.

We began this paper with a series of questions as to how the severity of psychopathology should be conceptualized. Some authors have suggested that the core indicator of the severity of mental illness is functional disability[200]. The DSM-5 has defined the severity of different disorders in different ways. Our review of the literature for depression and PDs demonstrated that researchers have adopted a myriad of ways of defining severity. The severity of depression has predominantly been defined according to scores on symptom rating scales. To be sure, there is some variability in how items are rated (i.e., symptom intensity vs. symptom frequency vs. symptom persistence), as well as some variability in the range of symptoms assessed by different measures of depression. Irrespective of the precise manner by which symptom severity is determined, most of the literature on the severity of depression is based on the parameters of symptoms. By contrast, the core of personality pathology is intertwined with its impact on functioning. Distinguishing extreme variants of personality traits from functioning has been challenging, therefore functional impairment has been fundamental to conceptualizing the severity of PDs.

Because the functional impact of symptom-defined disorders such as MDD depends on factors unrelated to the disorder such as self-efficacy, resilience, coping ability, social support, cultural and social expectations, as well as the responsibilities related to one's primary role function and the availability of others to assume those responsibilities, we would argue that the severity of such disorders should be defined independently from functional impairment. To those who would disagree, consider the following scenario: two individuals have an upper respiratory tract infection. They have the same elevation in body temperature, sneeze and cough with the same frequency, have the same level of mucus production and nasal discharge, and the same viral load. And the symptoms last for the same number of days. In short, they have the same intensity, frequency, and persistence of symptoms. Yet one person misses work for a week and the other does not miss work. Does the person who missed work have a more severe upper respiratory tract infection?

A distinction could be made between defining severity at the level of a disorder vs. overall global illness severity. As stated, at the level of disorder, severity should be determined by the factors that are intrinsic to the disorder. Thus, the severity of depression should be determined by the intensity, frequency, and/or persistence of the depressive symptoms. And the same is true for other disorders such as generalized anxiety disorder, post-traumatic stress disorder, mania/hypomania, and tic disorder. The severity of panic disorder should be based on the intensity and frequency of panic attacks. The severity of premature ejaculation should be based on time to ejaculation, the severity of hypoactive sexual desire based on the intensity (or lack thereof) of desire, the severity of binge eating disorder on the frequency and intensity of binges, etc..

The episodic nature of some psychiatric disorders and symptoms presents some measurement challenges. There may be day-to-day variability in symptom intensity as well as symptom persistence through the course of the day. Symptom frequency varies by disorder. Too little research has compared the validity of symptom intensity, frequency, and persistence assessments.

Severity, however, can be considered from another perspective: at the level of overall illness. A patient with depression, borderline PD, some anxiety disorders, substance use disorder and an eating disorder has a severe illness. It would likely be difficult to parse the levels of functional impairment to the separate disorders. The severity of the symptoms of depression may not be high, but the patient is nonetheless severely ill. How to take into account comorbidity when determining the severity of individual disorders is not clear. A global rating of overall illness severity was included in DSM-III through DSM-IV, but dropped from DSM-5. The global rating of illness severity can be considered to be akin to the composite measures of physical illness severity, described in the introduction, that have been used to predict mortality in emergency room and hospitalized patients. The problem with the GAF was that it was a single rating that required consideration of multiple constructs, including symptom frequency, type of symptom, level of impairment, suicidality, ability to care for oneself, and psychosis. Because of its complexity, there were problems with the reliability of its ratings[201]. Perhaps the dimensionally based measures of psychopathology articulated in HiTOP will yield clinically meaningful and useful approaches towards characterizing overall severity.

In the future, research on severity needs to be clear as to what correlates of a measure are expected. We noted above that too little research has compared the validity of symptom intensity, frequency, and persistence assessments. The question is how to evaluate validity. Should severity be a predictor of outcome? Should it help match patients to appropriate treatments or appropriate levels of care? Should it predict mortality? Should it reflect underlying pathophysiology? Should it confer genetic risk? Should it be used to guide the allocation of finite resources at either the insurance company or governmental funding agency level?

There are a wealth of papers in the psychiatric, medical and epidemiological literatures that refer to depression severity in the title and examine the correlates of a measure of depressive symptoms. But how to best measure severity has largely not been the subject of study. Numerous scales have been developed that purport to measure the severity of depression. When the authors of these scales discuss the reason behind developing their measure, the explanation usually focuses on item content and rarely discusses the reason for choosing a particular rating approach. Perhaps it does not make a meaningful difference how items are scaled. Perhaps the exact content of a scale does not make a meaningful difference either. Perhaps simplicity and clinical utility should trump any minor incremental validity that one measure shows over another.

However, some research suggests otherwise. The ability to detect differences between medication and placebo may be related to the content of the measure used[202]. Scales differ in severity classification[111,112,114], and treatment guidelines suggest that severity be used to select among treatment alternatives[1,2]. Thus, severity has real world implications in both the research and clinical communities. It is our hope that this paper stimulates more consideration and research into the issue of how to best conceptualize and measure the severity of psychiatric disorders.

REFERENCES

1. American Psychiatric Association. Practice guideline for the treatment of patients with major depressive disorder, 3rd ed. Washington: American Psychiatric Association, 2010.
2. National Collaborating Centre for Mental Health. Depression: the treatment and management of depression in adults. London: National Institute for Health and Clinical Excellence, 2009.
3. Tyrer P, Crawford M, Mulder R et al. Reclassifying personality disorders. Lancet 2011;377:1814-5.
4. Tyrer P, Crawford M, Mulder R et al. The rationale for the reclassification of personality disorder in the 11th revision of the International Classification of Diseases (ICD-11). Personal Ment Health 2011;5:246-59.
5. Minichiello E, Semerano L, Boissier MC. Time trends in the incidence, prevalence, and severity of rheumatoid arthritis: a systematic literature review. Joint Bone Spine 2016;83:625-30.
6. Hirai FE, Tielsch JM, Klein BE et al. Relationship between retinopathy severity, visual impairment and depression in persons with long-term type 1 diabetes. Ophthalmic Epidemiol 2012;19:196-203.
7. Greco A, Steca P, Pozzi R et al. Predicting depression from illness severity in cardiovascular disease patients: self-efficacy beliefs, illness perception, and perceived social support as mediators. Int J Behav Med 2014;21:221-9.
8. Steca P, Greco A, Monzani D et al. How does illness severity influence depression, health satisfaction and life satisfaction in patients with cardiovascular disease? The mediating role of illness perception and self-efficacy beliefs. Psychol Health 2013;28:765-83.
9. Pelletier R, Lavoie KL, Bacon SL et al. Depression and disease severity in patients with premature acute coronary syndrome. Am J Med 2014;127:87-93.
10. Carels RA. The association between disease severity, functional status, depression and daily quality of life in congestive heart failure patients. Qual Life Res 2004;13:63-72.
11. Snell C, Fernandes S, Bujoreanu IS et al. Depression, illness severity, and healthcare utilization in cystic fibrosis. Pediatr Pulmonol 2014;49:1177-81.
12. Deterding K, Gruner N, Buggisch P et al. Symptoms of anxiety and depression are frequent in patients with acute hepatitis C and are not associated with disease severity. Eur J Gastroenterol Hepatol 2016;28:187-92.
13. Euesden J, Matcham F, Hotopf M et al. The relationship between mental health, disease severity, and genetic risk for depression in early rheumatoid arthritis. Psychosom Med 2017;79:638-45.
14. Reynolds JC, Rittenberger JC, Toma C et al. Risk-adjusted outcome prediction with initial post-cardiac arrest illness severity: implications for cardiac arrest survivors being considered for early invasive strategy. Resuscitation 2014;85:1232-9.
15. Coppler PJ, Elmer J, Calderon L et al. Validation of the Pittsburgh Cardiac Arrest Category illness severity score. Resuscitation 2015;89:86-92.
16. Schaeffer JJ, Gil KM, Burchinal M et al. Depression, disease severity, and sickle cell disease. J Behav Med 1999;22:115-26.
17. Kim KU, Park HK, Jung HY et al. Association of depression with disease severity in patients with chronic obstructive pulmonary disease. Lung 2014;192:243-9.
18. van Dijk JP, Havlikova E, Rosenberger J et al. Influence of disease severity on fatigue in patients with Parkinson's disease is mainly mediated by symptoms of depression. Eur Neurol 2013;70:201-9.
19. Goetz CG, Tilley BC, Shaftman SR et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-

UPDRS): scale presentation and clinimetric testing results. Mov Disord 2008;23:2129-70.
20. Bennett JA, Riegel B, Bittner V et al. Validity and reliability of the NYHA classes for measuring research outcomes in patients with cardiac disease. Heart Lung 2002;31:262-70.
21. Knaus WA, Zimmerman JE, Wagner DP et al. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. Crit Care Med 1981;9:591-7.
22. Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. JAMA 1993;270:2957-63.
23. de Groot B, de Deckere ER, Flameling R et al. Performance of illness severity scores to guide disposition of emergency department patients with severe sepsis or septic shock. Eur J Emerg Med 2012;19:316-22.
24. Schneider AG, Lipcsey M, Bailey M et al. Simple translational equations to compare illness severity scores in intensive care trials. J Crit Care 2013;28: e881-8.
25. Barry MJ, Fowler FJ Jr, O'Leary MP et al. The American Urological Association symptom index for benign prostatic hyperplasia. J Urol 2017;197: S189-97.
26. Folmer RL, Shi YB. SSRI use by tinnitus patients: interactions between depression and tinnitus severity. Ear Nose Throat J 2004;83:107-8.
27. Gros DF, Antony MM, McCabe RE et al. Frequency and severity of the symptoms of irritable bowel syndrome across the anxiety disorders and depression. J Anxiety Disord 2009;23:290-6.
28. Stewart WF, Lipton RB, Simon D et al. Validity of an illness severity measure for headache in a population sample of migraine sufferers. Pain 1999; 79:291-301.
29. Baker GA, Smith DF, Jacoby A et al. Liverpool Seizure Severity Scale revisited. Seizure 1998;7:201-5.
30. Clark JA, Spiro A, Miller DR et al. Patient-based measures of illness severity in the Veterans Health Study. J Ambul Care Manage 2005;28:274-85.
31. Slade M, Powell R, Strathdee G. Current approaches to identifying the severely mentally ill. Soc Psychiatry Psychiatr Epidemiol 1997;32:177-84.
32. Ruggeri M, Leese M, Thornicroft G et al. Definition and prevalence of severe and persistent mental illness. Br J Psychiatry 2000;177:149-55.
33. American Psychiatric Association. Diagnostic and statistical manual of mental disorders, 5th ed. Arlington: American Psychiatric Association, 2013.
34. Fattori A, Neri L, Bellomo A et al. Depression severity and concentration difficulties are independently associated with HRQOL in patients with unipolar depressive disorders. Qual Life Res 2017;26:2459-69.
35. Goethe JW, Fischer EH, Wright JS. Severity as a key construct in depression. J Nerv Ment Dis 1993;181:718-24.
36. Luty SE, Joyce PR, Mulder RT et al. Social adjustment in depression: the impact of depression severity, personality, and clinic versus community sampling. J Affect Disord 2002;70:143-54.
37. Bradvik L, Mattisson C, Bogren M et al. Long-term suicide risk of depression in the Lundby cohort 1947-1997 – severity and gender. Acta Psychiatr Scand 2008;117:185-91.
38. Kessing LV. Severity of depressive episodes according to ICD-10: prediction of risk of relapse and suicide. Br J Psychiatry 2004;184:153-6.
39. Wang YY, Jiang NZ, Cheung EF et al. Role of depression severity and impulsivity in the relationship between hopelessness and suicidal ideation in patients with major depressive disorder. J Affect Disord 2015;183:83-9.
40. Katon W, Unutzer J, Russo J. Major depression: the importance of clinical characteristics and treatment response to prognosis. Depress Anxiety 2010; 27:19-26.
41. Keller MB, Lavori PW, Mueller TI et al. Time to recovery, chronicity, and levels of psychopathology in major depression. A 5-year prospective follow-up of 431 subjects. Arch Gen Psychiatry 1992;49:809-16.
42. Meyers BS, Sirey JA, Bruce M et al. Predictors of early recovery from major depression among persons admitted to community-based clinics: an observational study. Arch Gen Psychiatry 2002;59:729-35.
43. Melartin T, Rytsala H, Leskela U et al. Severity and comorbidity predict episode duration and recurrence of DSM-IV major depressive disorder. J Clin Psychiatry 2004;65:810-9.
44. Berent D, Zboralski K, Orzechowska A et al. Thyroid hormones association with depression severity and clinical outcome in patients with major depressive disorder. Mol Biol Rep 2014;41:2419-25.
45. de Diego-Adelino J, Pires P, Gomez-Anson B et al. Microstructural white-matter abnormalities associated with treatment resistance, severity and duration of illness in major depression. Psychol Med 2014;44:1171-82.
46. Zimmerman M, Coryell W, Pfohl B. The validity of the dexamethasone suppression test as a marker for endogenous depression. Arch Gen Psychiatry 1986;43:347-55.
47. Lux V, Aggen SH, Kendler KS. The DSM-IV definition of severity of major depression: inter-relationship and validity. Psychol Med 2010;40:1691-701.
48. Faravelli C, Servi P, Arends J et al. Number of symptoms, quantification, and qualification of depression. Compr Psychiatry 1996;37:307-15.
49. Guy W. ECDEU assessment manual for psychopharmacology. Rockville: National Institute of Mental Health, 1976.
50. American Psychiatric Association. Diagnostic and statistical manual of mental disorders, 4th ed. Washington: American Psychiatric Association, 1994.
51. Kitamura T, Nakagawa Y, Machizawa S. Grading depression severity by symptom scores: is it a valid method for subclassifying depressive disorders? Compr Psychiatry 1993;34:280-3.
52. Kessler RC, Zhao S, Blazer DG et al. Prevalence, correlates, and course of minor depression and major depression in the National Comorbidity Survey. J Affect Disord 1997;45:19-30.
53. Wakefield JC, Schmitz MF. Severity of complicated versus uncomplicated subthreshold depression: new evidence on the "monotonicity thesis" from the National Comorbidity Survey. J Affect Disord 2017;212:101-9.
54. Wakefield JC, Schmitz MF. Symptom quality versus quantity in judging prognosis: using NESARC predictive validators to locate uncomplicated major depression on the number-of-symptoms severity continuum. J Affect Disord 2017;208:325-9.
55. World Health Organization. International classification of diseases and related health problems, 10th revision. Geneva: World Health Organization, 2016.
56. Hiller W, Dichtl G, Hecht H et al. Evaluating the new ICD-10 categories of depressive episode and recurrent depressive disorder. J Affect Disord 1994; 31:49-60.
57. Montgomery S. Are the ICD-10 or DSM-5 diagnostic systems able to define those who will benefit from treatment for depression? CNS Spectr 2016;21:283-8.
58. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960;23:56-62.
59. Montgomery SA, Åsberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry 1979;134:382-9.
60. Nemeroff CB. The burden of severe depression: a review of diagnostic challenges and treatment alternatives. J Psychiatr Res 2007;41:189-206.
61. Sonawalla SB, Fava M. Severe depression: is there a best approach? CNS Drugs 2001;15:765-76.
62. Williams JB, Kobak KA, Bech P et al. The GRID-HAMD: standardization of the Hamilton Depression Rating Scale. Int Clin Psychopharmacol 2008; 23:120-9.
63. Parker G, Hadzi-Pavlovic D, Sengoz A et al. A brief self-report depression measure assessing mood state and social impairment. J Affect Disord 1994;30:133-42.
64. Pilkonis PA, Choi SW, Reise SP et al. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS): depression, anxiety, and anger. Assessment 2011;18: 263-83.
65. Balsamo M, Giampaglia G, Saggino A. Building a new Rasch-based self-report inventory of depression. Neuropsychiatr Dis Treat 2014;10:153-65.
66. Vaccarino AL, Evans KR, Kalali AH et al. The Depression Inventory Development Workgroup: a collaborative, empirically driven initiative to develop a new assessment tool for major depressive disorder. Innov Clin Neurosci 2016;13:20-31.
67. Zimmerman M, Posternak M, Friedman M et al. Which factors influence psychiatrists' selection of an antidepressant? Am J Psychiatry 2004;161:1285-9.
68. Cheung HN, Power MJ. The development of a new multidimensional depression assessment scale: preliminary results. Clin Psychol Psychother 2012;19:170-8.
69. Licht RW, Qvitzau S, Allerup P et al. Validation of the Bech-Rafaelsen Melancholia Scale and the Hamilton Depression Scale in patients with major depression; is the total score a valid measure of illness severity? Acta Psychiatr Scand 2005;111:144-9.
70. Santor D, Coyne J. Examining symptom expression as a function of symptom severity: item performance on the Hamilton Rating Scale for Depression. Psychol Assess 2001;13:127-39.
71. Kroenke K, Spitzer R, Williams J. The PHQ-9. Validity of a brief depression severity measure. J Gen Int Med 2001;16:606-13.

72. Downey L, Hayduk LA, Curtis JR et al. Measuring depression-severity in critically ill patients' families with the Patient Health Questionnaire (PHQ): tests for unidimensionality and longitudinal measurement invariance, with implications for CONSORT. J Pain Sympt Manage 2016;51:938-46.
73. Fischer HF, Tritt K, Klapp BF et al. How to compare scores from different depression scales: equating the Patient Health Questionnaire (PHQ) and the ICD-10-Symptom Rating (ISR) using item response theory. Int J Methods Psychiatr Res 2011;20:203-14.
74. Adler M, Hetta J, Isacsson G et al. An item response theory evaluation of three depression assessment instruments in a clinical sample. BMC Med Res Methodol 2012;12:84.
75. Barthel D, Barkmann C, Ehrhardt S et al. Screening for depression in pregnant women from Cote d'Ivoire and Ghana: psychometric properties of the Patient Health Questionnaire-9. J Affect Disord 2015;187:232-40.
76. Pedersen SS, Mathiasen K, Christensen KB et al. Psychometric analysis of the Patient Health Questionnaire in Danish patients with an implantable cardioverter defibrillator (The DEFIB-WOMEN study). J Psychosom Res 2016;90:105-12.
77. Umegaki Y, Todo N. Psychometric properties of the Japanese CES-D, SDS, and PHQ-9 depression scales in university students. Psychol Assess 2017;29:354-9.
78. Zhong Q, Gelaye B, Fann JR et al. Cross-cultural validity of the Spanish version of PHQ-9 among pregnant Peruvian women: a Rasch item response theory analysis. J Affect Disord 2014;158:148-53.
79. Zimmerman M, Chelminski I, McGlinchey JB et al. A clinically useful depression outcome scale. Compr Psychiatry 2008;49:131-40.
80. Zimmerman M, Galione J, Attiullah N et al. Depressed patients perspectives of two measures of outcome: the Quick Inventory of Depressive Symptomatology (QIDS) and the Remission from Depression Questionnaire (RDQ). Ann Clin Psychiatry 2011;23:208-12.
81. Bentley KH, Gallagher MW, Carl JR et al. Development and validation of the Overall Depression Severity and Impairment Scale. Psychol Assess 2014; 26:815-30.
82. Bagby RM, Ryder AG, Schuller DR et al. The Hamilton Depression Rating Scale: has the gold standard become a lead weight? Am J Psychiatry 2004; 161:2163-77.
83. Zimmerman M, Posternak M, Chelminski I. Is it time to replace the Hamilton Depression Rating Scale as the primary outcome measure in treatment studies of depression? J Clin Psychopharmacol 2005;25:105-10.
84. Tiplady B. A self-rating scale for depression designed to be sensitive to change. Neuropharmacology 1980;19:1211-2.
85. Rush A, Trivedi M, Ibrahim H et al. The 16-item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. Biol Psychiatry 2003;54:573-83.
86. Olsen LR, Jensen DV, Noerholm V et al. The internal and external validity of the Major Depression Inventory in measuring severity of depressive states. Psychol Med 2003;33:351-6.
87. Rush AJ, Gullion CM, Basco MR et al. The Inventory of Depressive Symptomatology (IDS). Psychol Med 1996;26:477-86.
88. Moller HJ. Methodological aspects in the assessment of severity of depression by the Hamilton Depression Scale. Eur Arch Psychiatry Clin Neurosci 2001;251(Suppl. 2):13-20.
89. DeRubeis RJ, Gelfand LA, Tang TZ et al. Medications versus cognitive behavior therapy for severely depressed outpatients: mega-analysis of four randomized comparisons. Am J Psychiatry 1999;156:1007-13.
90. Gibbons R, Hur K, Brown C et al. Benefits from antidepressants: synthesis of 6-week patient-level outcomes from double-blind placebo-controlled randomized trials of fluoxetine and venlafaxine. Arch Gen Psychiatry 2012; 69:572-9.
91. Elkin I, Gibbons R, Shea M et al. Initial severity and differential treatment outcome in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. J Consult Clin Psychol 1995;63:841-7.
92. 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;5:260-8.
93. Rush AJ, First MB, Blacker D. Handbook of psychiatric measures, 2nd ed. Washington: American Psychiatric Publishing, 2008.
94. Khan A, Leventhal RM, Khan SR et al. Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. J Clin Psychopharmacol 2002;22:40-5.
95. Fournier JC, DeRubeis RJ, Hollon SD et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. JAMA 2010;303:47-53.
96. Dunner D, Lipschitz A, Pitts C et al. Efficacy and tolerability of controlled-release paroxetine in the treatment of severe depression: post hoc analysis of pooled data from a subset of subjects in four double-blind clinical trials. Clin Ther 2005;27:1901-11.
97. Kasper S. Efficacy of antidepressants in the treatment of severe depression: the place of mirtazapine. J Clin Psychopharmacol 1997;17(Suppl. 1): 19S-28S.
98. Montgomery S, Ferguson J, Schwartz G. The antidepressant efficacy of reboxetine in patients with severe depression. J Clin Psychopharmacol 2003;23:45-50.
99. Schmitt AB, Bauer M, Volz HP et al. Differential effects of venlafaxine in the treatment of major depressive disorder according to baseline severity. Eur Arch Psychiatry Clin Neurosci 2009;259:329-39.
100. Shelton RC, Prakash A, Mallinckrodt CH et al. Patterns of depressive symptom response in duloxetine-treated outpatients with mild, moderate or more severe depression. Int J Clin Pract 2007;61:1337-48.
101. Versiani M, Moreno R, Ramakers-van Moorsel C et al. Comparison of the effects of mirtazapine and fluoxetine in severely depressed patients. CNS Drugs 2005;19:137-46.
102. Hirschfeld R. Efficacy of SSRIs and newer antidepressants in severe depression: comparison with TCAs. J Clin Psychiatry 1999;60:326-35.
103. Montgomery S, Lecrubier Y. Is severe depression a separate indication? Eur Neuropsychopharmacol 1999;9:259-64.
104. Schatzberg AF. Antidepressant effectiveness in severe depression and melancholia. J Clin Psychiatry 1999;60(Suppl. 4):14-21.
105. Endicott J, Cohen J, Nee J et al. Hamilton depression rating scale. Arch Gen Psychiatry 1981;38:98-103.
106. Kearns NP, Cruickshank CA, McGuigan KJ et al. A comparison of depression rating scales. Br J Psychiatry 1982;141:45-9.
107. Cameron IM, Cardy A, Crawford JR et al. Measuring depression severity in general practice: discriminatory performance of the PHQ-9, HADS-D, and BDI-II. Br J Gen Pract 2011;61:e419-26.
108. Zimmerman M, Martinez JH, Young D et al. Severity classification on the Hamilton Depression Rating Scale. J Affect Disord 2013;150:384-8.
109. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. Acta Psychiatr Scand 1983;67:361-70.
110. Cameron IM, Crawford JR, Lawton K et al. Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. Br J Gen Pract 2008;58:32-6.
111. Hansson M, Chotai J, Nordstom A et al. Comparison of two self-rating scales to detect depression: HADS and PHQ-9. Br J Gen Pract 2009;59: e283-8.
112. Reddy P, Philpot B, Ford D et al. Identification of depression in diabetes: the efficacy of PHQ-9 and HADS-D. Br J Gen Pract 2010;60:e239-45.
113. Beck A, Steer R, Brown G. The Beck Depression Inventory, 2nd ed. San Antonio: The Psychological Corporation, 1996.
114. Zimmerman M, Martinez J, Friedman M et al. How can we use depression severity to guide treatment selection when measures of depression categorize patients differently? J Clin Psychiatry 2012;73:1287-91.
115. Ruscio J, Zimmerman M, McGlinchey JB et al. Diagnosing major depressive disorder XI: a taxometric investigation of the structure underlying DSM-IV symptoms. J Nerv Ment Dis 2007;195:10-9.
116. van der Lem R, van der Wee NJ, van Veen T et al. The generalizability of antidepressant efficacy trials to routine psychiatric out-patient practice. Psychol Med 2011;41:1353-63.
117. Bielski RJ, Friedel RO. Prediction of tricyclic antidepressant response: a critical review. Arch Gen Psychiatry 1976;33:1479-89.
118. Grammer GG, Kuhle AR, Clark CC et al. Severity of depression predicts remission rates using transcranial magnetic stimulation. Front Psychiatry 2015;6:114.
119. Jones NP, Siegle GJ, Thase ME. Effects of rumination and initial severity on remission to cognitive therapy for depression. Cogn Ther Res 2008;32: 591-604.
120. Lisanby SH, Husain MM, Rosenquist PB et al. Daily left prefrontal repetitive transcranial magnetic stimulation in the acute treatment of major depression: clinical predictors of outcome in a multisite, randomized controlled clinical trial. Neuropsychopharmacology 2009;34:522-34.
121. Bower P, Kontopantelis E, Sutton A et al. Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. BMJ 2013;346:f540.
122. Sugawara Y, Higuchi H, Yoshida K et al. Response rate obtained using milnacipran depending on the severity of depression in the treatment of major depressive patients. Clin Neuropharmacol 2006;29:6-9.

123. Friedman ES, Davis LL, Zisook S et al. Baseline depression severity as a predictor of single and combination antidepressant treatment outcome: results from the CO-MED trial. Eur Neuropsychopharmacol 2012;22:183-99.
124. Kennedy S, Andersen H, Lam R. Efficacy of escitalopram in the treatment of major depressive disorder compared with conventional selective serotonin reuptake inhibitors and venlafaxine SR: a meta-analysis. J Psychiatry Neurosci 2006;31:122-31.
125. Angst J, Amrein R, Stabl M. Moclobemide and tricyclic antidepressants in severe depression: meta-analysis and prospective studies. J Clin Psychopharmacol 1995;15:16S-23S.
126. Wiles NJ, Mulligan J, Peters TJ et al. Severity of depression and response to antidepressants: GENPOD randomised controlled trial. Br J Psychiatry 2011;200:130-6.
127. Kilts CD, Wade AG, Andersen HF et al. Baseline severity of depression predicts antidepressant drug response relative to escitalopram. Exp Opin Pharmacother 2009;10:927-36.
128. Khan A, Sambunaris A, Edwards J et al. Vilazodone in the treatment of major depressive disorder: efficacy across symptoms and severity of depression. Int Clin Psychopharmacol 2014;29:86-92.
129. Mosca D, Zhang M, Prieto R et al. Efficacy of desvenlafaxine compared with placebo in major depressive disorder patients by age group and severity of depression at baseline. J Clin Psychopharmacol 2017;37:182-92.
130. Driessen E, Cuijpers P, Hollon SD et al. Does pretreatment severity moderate the efficacy of psychological treatment of adult outpatient depression? A meta-analysis. J Consult Clin Psychol 2010;78:668-80.
131. Furukawa TA, Weitz ES, Tanaka S et al. Initial severity of depression and efficacy of cognitive-behavioural therapy: individual-participant data meta-analysis of pill-placebo-controlled trials. Br J Psychiatry 2017;210:190-6.
132. Guelfi JD, Ansseau M, Timmerman L et al. Mirtazapine versus venlafaxine in hospitalized severely depressed patients with melancholic features. J Clin Psychopharmacol 2001;21:425-31.
133. Henkel V, Seemuller F, Obermeier M et al. Relationship between baseline severity of depression and antidepressant treatment outcome. Pharmacopsychiatry 2011;44:27-32.
134. Hirschfeld RM, Russell JM, Delgado PL et al. Predictors of response to acute treatment of chronic and double depression with sertraline or imipramine. J Clin Psychiatry 1998;59:669-75.
135. Madhoo M, Levine SZ. Initial severity effects on residual symptoms in response and remission: a STAR*D study during and after failed citalopram treatment. J Clin Psychopharmacol 2015;35:450-3.
136. Souery D, Oswald P, Massat I et al. Clinical factors associated with treatment resistance in major depressive disorder: results from a European multicenter study. J Clin Psychiatry 2007;68:1062-70.
137. Brown C, Schulberg HC, Prigerson HG. Factors associated with symptomatic improvement and recovery from major depression in primary care patients. Gen Hosp Psychiatry 2000;22:242-50.
138. Enns MW, Cox BJ. Psychosocial and clinical predictors of symptom persistence vs. remission in major depressive disorder. Can J Psychiatry 2005;50:769-77.
139. Sargeant JK, Bruce ML, Florio LP et al. Factors associated with 1-year outcome of major depression in the community. Arch Gen Psychiatry 1990; 47:519-26.
140. Turner EH, Matthews AM, Linardatos E et al. Selective publication of antidepressant trials and its influence on apparent efficacy. N Engl J Med 2008;358:252-60.
141. Rabinowitz J, Werbeloff N, Mandel FS et al. Initial depression severity and response to antidepressants vs. placebo: patient-level data analysis from 34 randomised controlled trials. Br J Psychiatry 2016;209:427-8.
142. Elkin I, Shea M, Watkins J et al. NIMH treatment of depression collaborative research program: general effectiveness of treatments. Arch Gen Psychiatry 1989;46:971-82.
143. Cuijpers P, van Straten A, van Oppen P et al. Are psychological and pharmacologic interventions equally effective in the treatment of adult depressive disorders? A meta-analysis of comparative studies. J Clin Psychiatry 2008;69:1675-85.
144. Weitz ES, Hollon SD, Twisk J et al. Baseline depression severity as moderator of depression outcomes between cognitive behavioral therapy vs. pharmacotherapy: an individual patient data meta-analysis. JAMA Psychiatry 2015;72:1102-9.
145. Links PS, Eynan R. The relationship between personality disorders and Axis I psychopathology: deconstructing comorbidity. Annu Rev Clin Psychol 2013;9:529-54.
146. Lenzenweger MF, Lane MC, Loranger AW et al. DSM-IV personality disorders in the National Comorbidity Survey Replication. Biol Psychiatry 2007;62:553-64.
147. Bateman AW, Fonagy P. The effectiveness of partial hospitalization in the treatment of borderline personality disorder – a randomised controlled trial. Am J Psychiatry 1999;156:1563-9.
148. National Collaborating Centre for Mental Health. Borderline personality disorder: recognition and management. London: Department of Health, 2009.
149. Blum N, St John D, Pfohl B et al. Systems Training for Emotional Predictability and Problem Solving (STEPPS) for outpatients with borderline personality disorder: a randomized controlled trial and 1-year follow-up. Am J Psychiatry 2008;165:468-78.
150. Zanarini MC, Yong L, Frankenburg FR et al. Severity of childhood sexual abuse and its relationship to severity of borderline psychopathology and psychosocial impairment among borderline inpatients. J Nerv Ment Dis 2002;190:381-7.
151. Giesen-Bloo JH, Wachters LM, Schouten E et al. The borderline personality disorder severity index-IV: psychometric evaluation and dimensional structure. Pers Ind Diff 2010;49:136-41.
152. Yang M, Coid J, Tyrer P. Personality pathology recorded by severity: national survey. Br J Psychiatry 2010;197:193-9.
153. Tyrer P, Alexander MS, Cicchetti D et al. Reliability of a schedule for rating personality disorders. Br J Psychiatry 1979;135:168-74.
154. Dolan B, Evans C, Norton K. Multiple axis-II diagnoses of personality disorder. Br J Psychiatry 1995;166:107-12.
155. Oldham JM, Skodol AE, Kellman HD et al. Diagnosis of DSM-III-R personality disorders by two structured interviews: patterns of comorbidity. Am J Psychiatry 1992;149:213-20.
156. Zimmerman M, Galione JN, Chelminski I et al. Does the diagnosis of multiple Axis II disorders have clinical significance? Ann Clin Psychiatry 2012; 24:195-201.
157. Tyrer P, Johnson T. Establishing the severity of personality disorder. Am J Psychiatry 1996;153:1593-7.
158. Asnaani A, Chelminski I, Young D et al. Heterogeneity of borderline personality disorder: do the number of criteria met make a difference? J Pers Disord 2007;21:615-25.
159. Kernberg OF, Caligor E. A psychoanalytic theory of personality disorders. In: Clarkin JF, Lenzenweger MF (eds). Major theories of personality disorder. New York: Guilford, 2005:114-56.
160. Paris J. Dimensional diagnosis and the DSM-5. J Clin Psychiatry 2005;72: 1340.
161. Moran P. Dangerous severe personality disorder – bad tidings from the UK. Int J Soc Psychiatry 2001;48:6-10.
162. Maden A, Tyrer P. Dangerous and severe personality disorders: a new personality concept from the United Kingdom. J Pers Disord 2003;17: 489-96.
163. Tyrer P, Cooper S, Rutter D et al. The assessment of dangerous and severe personality disorder: lessons from a randomised controlled trial linked to qualitative analysis. Forensic Psychol Psychiatry 2009;20:132-46.
164. Berghuis H, Kamphuis JH, Verheul R. Specific personality traits and general personality dysfunction as predictors of the presence and severity of personality disorders in a clinical sample. J Pers Assess 2014;96: 410-6.
165. Crawford MJ, Koldobsky N, Mulder R et al. Classifying personality disorder according to severity. J Pers Disord 2011;25:321-30.
166. Livesley WJ. Practical management of personality disorders. New York: Guilford, 2003.
167. Widiger TA, Costa PT, McCrae RR. A proposal for Axis II: diagnosing personality disorders using the five factor model. In: Costa PT, Widiger TA (eds). Personality disorders and the five factor model of personality. Washington: American Psychological Association, 2002:431-52.
168. Widiger TA, Mullins-Sweatt SN. Five-factor model of personality disorder: a proposal for DSM-V. Annu Rev Clin Psychol 2009;5:197-220.
169. Widiger TA, Trull TJ. Plate tectonics in the classification of personality disorder: shifting to a dimensional model. Am Psychol 2007;62:71-83.
170. Livesley W, Schroeder M, Jackson D et al. Categorical distinctions in the study of personality disorder: implications for classification. J Abnorm Psychol 1994;103:6-17.
171. Parker G, Hadzi-Pavlovic D, Both L et al. Measuring disordered personality functioning: to love and to work reprised. Acta Psychiatr Scand 2004; 110:230-9.

172. Trull TJ. Dimensional models of personality disorder: coverage and cutoffs. J Pers Disord 2005;19:262-82.
173. Verheul R, Andrea H, Berghout CC et al. Severity Indices of Personality Problems (SIPP-118): development, factor structure, reliability, and validity. Psychol Assess 2008;20:23-34.
174. Livesley WJ. Suggestions for a framework for an empirically based classification of personality disorder. Can J Psychiatry 1998;43:137-47.
175. Ro E, Clark LA. Interrelations between psychosocial functioning and adaptive- and maladaptive-range personality traits. J Abnorm Psychol 2013;122:822-35.
176. Krueger RF, Derringer J, Markon KE et al. Initial construction of a maladaptive personality trait model and inventory for DSM-5. Psychol Med 2012;42:1879-90.
177. Gunderson J, Zanarini MC. Commentary: deceptively simple – or radical shift? Pers Ment Health 2011;5:260-2.
178. Kelly BD, Nur UA, Tyrer P et al. Impact of severity of personality disorder on the outcome of depression. Eur Psychiatry 2009;24:322-6.
179. Livesley WJ. General Assessment of Personality Disorder (GAPD). Vancouver: University of British Columbia, 2006.
180. Berghuis H, Kamphuis JH, Verheul R. Core features of personality disorder: differentiating general personality dysfunctioning from personality traits. J Pers Disord 2012;26:704-16.
181. Berghuis H, Kamphuis JH, Verheul R et al. The General Assessment of Personality Disorder (GAPD) as an instrument for assessing the core features of personality disorders. Clin Psychol Psychother 2013;20:544-57.
182. Arntz A, van den Hoorn M, Cornelis J et al. Reliability and validity of the borderline personality disorder severity index. J Pers Disord 2003;17:45-59.
183. Clarkin JF, Levy KN, Lenzenweger MF et al. Evaluating three treatments for borderline personality disorder: a multiwave study. Am J Psychiatry 2007;164:922-8.
184. Linehan MM, Comtois KA, Murray AM et al. Two-year randomized controlled trial and follow-up of dialectical behavior therapy vs. therapy by experts for suicidal behaviors and borderline personality disorder. Arch Gen Psychiatry 2006;63:757-66.
185. Zanarini MC, Vujanovic AA, Parachini EA et al. Zanarini Rating Scale for Borderline Personality Disorder (ZAN-BPD): a continuous measure of DSM-IV borderline psychopathology. J Pers Disord 2003;17:233-42.
186. Patton JH, Stanford MS, Barratt ES. Factor structure of the Barratt Impulsiveness Scale. J Clin Psychol 1995;51:768-74.
187. Doering S, Horz S, Rentrop M et al. Transference-focused psychotherapy v. treatment by community psychotherapists for borderline personality disorder: randomised controlled trial. Br J Psychiatry 2010;196:389-95.
188. McMain SF, Links PS, Gnam WH et al. A randomized trial of dialectical behavior therapy versus general psychiatric management for borderline personality disorder. Am J Psychiatry 2009;166:1365-74.
189. Diamond D, Yeomans FE, Stern B et al. Transference focused psychotherapy for patients with comorbid narcissistic and borderline personality disorder. Psychoanal Inq 2013;33:527-51.
190. Messina NP, Wish ED, Hoffman JA et al. Antisocial personality disorder and treatment outcomes. Am J Drug Alcohol Abuse 2002;28:197-212.
191. Kotov R, Krueger RF, Watson D et al. The Hierarchical Taxonomy of Psychopathology (HiTOP): a dimensional alternative to traditional nosologies. J Abnorm Psychol 2017;126:454-77.
192. Stanton K, Rozek DC, Stasik-O'Brien SM et al. A transdiagnostic approach to examining the incremental predictive power of emotion regulation and basic personality dimensions. J Abnorm Psychol 2016;125:960-75.
193. Zimmerman M. A review of 20 years of research on overdiagnosis and underdiagnosis in the Rhode Island Methods to Improve Diagnostic Assessment and Services (MIDAS) project. Can J Psychiatry 2016;61:71-9.
194. Krueger RF, Finger MS. Using item response theory to understand comorbidity among anxiety and unipolar mood disorders. Psychol Assess 2001; 13:140-51.
195. Eaton NR, Krueger RF, Markon KE et al. The structure and predictive validity of the internalizing disorders. J Abnorm Psychol 2013;122:86-92.
196. Krueger RF, Markon KE, Patrick CJ et al. Externalizing psychopathology in adulthood: a dimensional-spectrum conceptualization and its implications for DSM-V. J Abnorm Psychol 2005;114:537-50.
197. Carragher N, Krueger RF, Eaton NR et al. ADHD and the externalizing spectrum: direct comparison of categorical, continuous, and hybrid models of liability in a nationally representative sample. Soc Psychiatry Psychiatr Epidemiol 2014;49:1307-17.
198. Kotov R, Chang SW, Fochtmann LJ et al. Schizophrenia in the internalizing-externalizing framework: a third dimension? Schizophr Bull 2011; 37:1168-78.
199. Kotov R, Krueger RF, Watson D. A paradigm shift in psychiatric classification: the Hierarchical Taxonomy Of Psychopathology (HiTOP). World Psychiatry 2018;17:24-5.
200. Gaebel W, Zaske H, Baumann AE. The relationship between mental illness severity and stigma. Acta Psychiat Scand 2006;113(Suppl. 429): 41-5.
201. Grootenboer EM, Giltay EJ, van der Lem R et al. Reliability and validity of the Global Assessment of Functioning Scale in clinical outpatients with depressive disorders. J Eval Clin Pract 2012;18:502-7.
202. Bech P, Boyer P, Germain JM et al. HAM-D17 and HAM-D6 sensitivity to change in relation to desvenlafaxine dose and baseline depression severity in major depressive disorder. Pharmacopsychiatry 2010;43:271-6.

DOI:10.1002/wps.20569

Digital phenotyping: a global tool for psychiatry

In 2050, when psychiatrists look back at the first two decades of the 21st century, what will they recognize as having the greatest impact? No doubt the revolution in genomics, which has given us new insights into the risk architecture of mental illness, and the revolution in neuroscience, which has given us a new view of mental illnesses as circuit disorders, will be considered important. But perhaps the revolution in technology and information science will prove more consequential for global mental health.

If this sounds like hyperbole, consider two supportive data points. First, in the past decade smartphones have become nearly ubiquitous. There are over three billion smartphone Internet subscriptions, each device with the information processing capacity of the supercomputers of the 1990s[1]. In many parts of the world that lack credit cards, phones have become the primary way to conduct commerce. Second, broadband access to social media and search platforms is becoming global. In 2016, 3.3 billion people had Internet access, one third of whom were in India and China[2]. Even in areas without easy access to clean water, ownership of a smartphone and rapid access to information have become the symbols of modernity.

The smartphone and the Internet can solve specific problems that we face in psychiatry, but their clinical use also raises new ethical challenges.

What specific problems can be addressed by the smartphone? Our lack of objective measurement has handicapped both diagnosis and treatment in psychiatry. As just one example, our assessment of depression depends largely on self-reports of sleep, appetite and emotional state, although we recognize that people with depression are biased in their assessments. The smartphone offers us an objective and ecological source of measurement. This approach, now called digital phenotyping, is based on sensors (activity and location), voice and speech (sentiment and prosody), and, perhaps most important, human-computer interaction[3].

Human-computer interaction measures not what you type but how you type. Subtle aspects of typing and scrolling, such as the latency between space and character or the interval between scroll and click, are surprisingly good surrogates for cognitive traits and affective states[4]. If this seems improbable, remember that many of our neuropsychological tests, such as the Trails A and B tests or the Digit Symbol Substitution, are not substantially different from the psychomotor requirements of operating a smartphone. In a sense, those gold standard tests of cognitive control and information processing are attempting to assess how we function. In a world where we spend so much of our lives on our smartphones, could it become possible to assess how we function directly and continuously rather than using laboratory measures at a single point in time?

The promise of digital phenotyping is that this objective measure happens in the context of a patient's lived experience, reflecting how he/she functions in his/her world, not in our clinic. Signals from a new mother struggling with depression may look quite different during a 3 am feeding compared to what she reports to her clinician the next day. This kind of ecological and continuous measurement addresses some of the central issues that challenge our field. We know that most people with a mental illness do not seek help, and those who do seek help usually arrive after considerable delay. For populations at risk, such as post-partum women or victims of trauma, could digital phenotyping signal the transition from risk to the need for care? For people in care, too often we fail to preempt relapse. For patients in treatment, could digital phenotyping serve as a "smoke alarm" providing early signals of relapse or recovery?

Digital phenotyping is still being developed as a clinical tool. It seems clear from the early results that, although activity and geolocation data are non-specific and noisy, for some people changes in activity can be an early sign of mania or depression[5]. Speech and voice may also yield clinically relevant signals. We have known for decades that when people are depressed their pronouns shift to first person singular[6]. But again, the sensitivity and specificity of these findings still need to be defined. Putting sensor data, speech and voice data, and human-computer interaction together might provide a digital phenotype that could do for psychiatry what HgbA1c or serum cholesterol has done for other areas of medicine, giving precision to diagnosis and accuracy to outcomes.

The opportunity of this new approach to measurement is matched by an ethical challenge. When does measurement become surveillance? Is tracking geolocation or collecting speech too intrusive? How can patients trust that digital phenotyping data will be protected? Even if patients consent to have their smartphone monitored, is there full transparency and a deep understanding of what data will be collected and how these data will be used? Who owns the data? For psychiatry, one of the most informative phone signals might reside in the "digital exhaust", such as search history or social media posts. Those signals might confess suicidal intent or early signs of psychosis. Does the value of this information outweigh the intrusion of privacy required to obtain it?

All of these issues are part of an active debate, as merits any new promising technology. To be clear, digital phenotyping is still a research project conducted on small numbers of consented volunteers. While researchers hope this approach will solve global mental health problems, the scientific and ethical issues need to be resolved before digital phenotyping becomes a tool for population health.

Some of the most vexing issues may have technical solutions. For instance, human-computer interaction is "content-free". This approach collects how you type, not what you type and, therefore, might be less intrusive than monitoring geolocation or search history. Tools that can analyze smartphone

signals on the phone rather than sending data to the cloud have the advantage of keeping raw data local and private. And other approaches, such as Google takeout, that empower users to monitor their own data can avoid the sense of surveillance.

Some have claimed that the smartphone is more the source than the solution for mental disorders[7]. As phones kidnap our attention and remove us from real world interaction, this worry seems increasingly urgent, especially in young people who are the most intensive smartphone users. On the other hand, the smartphone may be an unprecedented opportunity to measure real-world functioning and potentially to offer just-in-time interventions.

All new technologies face this dual-use dilemma between risk and benefit. For digital phenotyping, this is the time for patients, families, providers and researchers to define together the balance between clinical value and public trust.

Thomas R. Insel
Mindstrong Health, Palo Alto, CA, USA

1. Ericsson. https://www.ericsson.com/en/news/2018/2/5.2-billion-mobile-broadband-subscriptions.
2. International Telecommunication Union. https://www.itu.int/en/ITU-D/Statistics/Pages/stat/treemap.aspx.
3. Insel TR. JAMA 2017;318:1215-6.
4. Dagum P. npj Digital Medicine 2018;1:10.
5. Reinertsen E, Clifford GD. Physiol Meas 2018;39:05TR01.
6. Pennybaker JW, Mehl MR, Niederhoffer KG. Annu Rev Psychol 2003;54:547-77.
7. Twenge JM, Martin GN, Campbell WK. Emotion (in press).

DOI:10.1002/wps.20550

Telemental health: why the revolution has not arrived

Mental illness is often underdiagnosed and undertreated. Several obstacles help explain this public health problem, including provider shortage, difficulty accessing care, cost, stigma, and a variety of diagnosis-specific issues. By promising to broaden access, increase efficiency, decrease costs, and remove stigma, telemental health has been touted as a solution[1].

However, despite three decades of often encouraging investigations across several technology platforms (computerized therapy, Internet-delivered video- or chat-based treatment, mobile therapy, "serious games", and virtual reality therapy), significant challenges continue to limit the wide adoption of telemental health interventions. They include: the present state of research; the rise of "coaching"; attrition rates; security concerns; legal confusion; insufficient guidance from professional organizations; comparisons with gaming; and the still relevant obstacles of infrastructure cost and technical know-how.

Most telepsychiatry studies are too small and unrepresentative, and lack the control of in-person treatment. Consequently, they limit broad recommendations in favor of adoption. The discrepancy between the slow pace of research (the process of funding procurement, protocol design, institutional review board approval, recruitment, testing, data analysis, peer review, and publication) and the breakneck pace of technology also limits the value of existing studies. By the time a well-designed trial generates data, the platform may be outdated or less appealing given more sophisticated alternatives now available. This can mean that research-based recommendations often lag available offerings. It can also mean that marketing by well-funded health technology companies can be divorced from the evidence base, with serious regulatory consequences[2].

Paradoxically, the rise of "coaching" may have also limited telemental health adoption. Many studies have set out to prove that adequate psychotherapy can be implemented with little or no support from professionally trained providers[3]. This mirrors the move in broader psychotherapy from the interpretive therapist to one following a standardized cognitive-behavioral therapy (CBT) model. Less reliance on therapists would be laudable if it democratized care. However, one consequence may have been to depend on "coaches" who need no particular training or licensing, and who provide support while eschewing direct "treatment". As a result, medical professionals can now be entirely bypassed: many patients already self-diagnose via "Dr. Google" and, now, they can self-treat using telepsychiatry tools, with or without the help of a "coach". This can lead mental health providers to view telepsychiatry as a potential competitor that aims to supplant them with lesser-trained individuals (or standalone platforms). Consequently, they may hesitate to recommend telemental health services.

Treatment adherence represents another challenge, and studies have suggested higher attrition rates compared to conventional treatment[1]. While analyzing the patient-therapist relationship is no longer a cornerstone of treatment, having no relationship (e.g., standalone computerized CBT) or a very limited one (e.g., online CBT modules with minimal therapist contact) may preclude a "therapeutic alliance", thereby perhaps decreasing motivation to engage in treatment. Ingrained online habits, where relationship "termination" is as easy as the click of a button (e.g., "unfriending" or "blocking"), may also contribute to poor adherence to a telepsychiatry provider and telepsychiatry interventions.

With frequent news of hacks into supposedly secure networks, questions arise about the possibility of safeguarding digital platforms, presenting another challenge to the practice of telepsychiatry. Telemental health research has not prioritized testing limits, expectations and views around security. Yet, this is a crucial determinant of adoption for both patients and providers. Simply stating a platform is encrypted is insuffi-

cient, and making platform security a design and research priority may help reassure reluctant users.

Another challenge is the confusing legal landscape within which telepsychiatry practice occurs. Depending on the country, this may involve adhering to a complex web of federal and regional legislation. In the US, for example, treatment must adhere to federal laws that predate current telemental health tools (e.g., the Health Insurance Portability and Accountability Act of 1996). The result may be that crucial questions in telepsychiatry practice remain unanswered, such as whether ubiquitous tools like FaceTime and Skype meet the requirements of health care technology legislation. Also, in the US, where licensing laws are regional and deem care to occur in the state where the patient resides, cross-state treatment is severely limited, a reality that neutralizes a key telemedicine value proposition – correcting shortages in access to care.

The dearth of guidance from leading professional organizations has also limited telemental health adoption. The first major telemental health initiatives by the American Psychiatric Association and the American Psychological Association, for example, date back only to 2015 and 2011, respectively. This has contributed to confusion among practitioners regarding treatment "best practices", remote management of emergencies, reimbursement, insurance coverage, malpractice protection, documentation, product vetting, and security. More guidance is required if providers are to embrace promising novel treatments that may come with heightened risks.

Further, certain telemental health tools have not escaped automatic comparisons with video games or other online or technology-enabled entertainment. This is particularly true within the field of "serious games", defined as video games with educational or therapeutic goals[4], and virtual reality therapy. Especially when infrastructure investment can be significant, interventions that are perceived as entertaining but not necessarily therapeutic will struggle to gain footing.

Indeed, infrastructure, while significantly less expensive now, as evidenced by the decrease in the price of virtual reality equipment[5], is still not universally affordable. This represents an ongoing challenge to wider adoption; one that mirrors technical know-how, which – while no longer the obstacle it was, due to increased technology literacy and ever more "plug and play" models – still represents a challenge in certain populations.

The unmet needs in mental heath care are too large to be addressed without leveraging technological innovations. Mental health care is particularly well suited to benefit from telemedicine advances, but several obstacles have made it so that the telemental health revolution, with its promised solutions, has not yet arrived. Concerted efforts by funding agencies, researchers, engineers, public health authorities, professional organizations, and legislative bodies are needed if the hope is to translate into real-life improvement.

Elias Aboujaoude
Stanford University School of Medicine, Stanford, CA, USA

1. Aboujaoude E, Salame W, Naim L. World Psychiatry 2015;14:223-30.
2. Federal Trade Commission Bureau of Consumer Protection. Lumosity to pay $2 million to settle FTC deceptive advertising charges for its "brain training" program. https://www.ftc.gov/news-events/press-releases.
3. Aboujaoude E. J Obsessive Compuls Relat Disord 2017;14:65-70.
4. Djaouti D, Alvarez J, Jessel JP. In: Felicia P (ed). Handbook of research on improving learning and motivation through educational games: multidisciplinary approaches. Hershey: IGI Global, 2011:118-36.
5. Fink C. Behind those high end VR price cuts. www.forbes.com.

DOI:10.1002/wps.20551

The brain's center of gravity: how the default mode network helps us to understand the self

The self is an elusive concept. We have an intuitive sense as to what it refers to, but it defies simple definition. There is some consensus that the self can be broadly separated into what W. James referred to as the "I" and the "me" – the self that experiences, and the self that extends outwards in space and in time, allowing it to be perceived as an object[1]. This includes the self as physical object (the body), and as an abstract object with beliefs and attitudes. Divisions of the self similar to James's have been suggested by Damasio (the core and the autobiographical self)[2] and Gallagher (the minimal and the narrative self)[3].

The philosopher D. Dennett has defined the self as "the center of narrative gravity"[4]. This definition encapsulates the idea of the self as both the center of experience, and one that is situated in a broader and ongoing narrative. In using the center of gravity as a metaphor for the self, Dennett wanted to highlight that it – like the self – is an abstraction, having no physical properties. The center of gravity exists only as a concept, but one that is useful for predicting an object's characteristics (at what point will it tip over?). So it is that the self can be viewed: as a useful abstraction that we can all agree exists in a broad sense, but which cannot be precisely defined in physical terms.

Dennett argued that "it is a category mistake to start looking around for the self in the brain"; and that he couldn't imagine us ever saying: "that cell there, right in the middle of the hippocampus (or wherever) – that's the self!"[4]. He is right in the sense he discusses: we cannot locate the self in a particular

region of the brain. But modern neuroimaging techniques have been able to reveal that aspects of the self are associated with the dynamic coordinated activity of a large-scale brain network. This network is referred to as the default mode network (DMN).

The DMN is composed primarily of medial prefrontal cortex (MPFC) and posterior cingulate cortex (PCC), both situated along the brain's midline, together with inferior parietal and medial temporal regions. The network was first observed in nuclear imaging studies, where it was noted that the regions consistently showed *reduced* levels of activity when participants performed various goal-directed tasks[5]. The regions were described as comprising a "default mode" because it was thought that the pattern of activity was what the brain defaulted to in the absence of particular task demands[6]. This hypothesis has since been confirmed by other observations, including studies that have examined resting-state functional activity of the DMN.

The idea that DMN function underlies self-related processes has been demonstrated by experimental tasks, as well as by studies of participants who show reduced self-awareness (for example, as they enter sleep or anesthetic states). Overlapping regions of the DMN are generally activated by tasks that encourage self-reflection, with evidence of differential patterns of activation to task components.

The anterior DMN – and especially dorsal MPFC – is more broadly activated by self-directed thoughts: for example, by the effortful appraisal of one's attributes, or thinking about the self in past and future contexts. The posterior DMN, on the other hand, is more broadly active during passive resting-state conditions. It integrates spatial and interoceptive representations of the body, along with low-level surveillance of one's surroundings.

We have recently examined how MPFC and PCC act in concert during self-referential processing, showing that PCC appears to coordinate the generation of relevant self-representations, while MPFC acts to select and gate the representations into conscious awareness[7].

Imaging "connectomic" approaches, which explore how regions of the brain interact with one another from a dynamic whole-brain perspective, have shown that the MPFC and PCC have among the highest degrees of global connectivity, serving as hubs in the brain's overall network organization[8]. The regions act at the intersection of large-scale networks, where they integrate information from diverse sources – including from self-relevant sources such as autobiographical memory and interoceptive processes. Evidence from connectomic studies suggests that the DMN is unique in its capacity to integrate information processing across the brain, allowing it to support the generation of higher-order, self-related mental activity.

Brain networks must affect motor output to influence behavior. The MPFC has rich connections with the hypothalamus and midbrain autonomic control centers, thereby influencing affective, visceral and behavioral responses to events[9]. The hypothalamus drives tendencies to fight, flee, feed and fornicate (the famous "4 Fs"), as well as influencing sleep, energy levels, and other neuroendocrine processes. By means of these systems, the DMN influences the state of the body, and the way it is represented by internal processes, which we hypothesize become dynamically re-integrated with higher-level DMN self-representations. The DMN therefore coordinates a sense of self that spans cognitive abstractions about the self with a more grounded awareness of the state of the body in the here and now.

The center of gravity was introduced by Dennett as a metaphor for how we might understand the self; as a useful abstraction that we cannot define in terms related to its physical properties. Here, we propose extending that metaphor to illustrate the role of the DMN.

The center of gravity is a dynamic property of complex moving objects, such as the human body. It is created from the sum of variables related to the mass, shape, acceleration and rotation of the object's interacting parts, and shifts with movement. In the act of bipedal walking, for example, the center of gravity is propelled forward with the generation of movement, and must be constantly adjusted so that our bodies remain upright over uneven terrain.

It is in this light that we can recognize the role of the default mode network: as a dynamic entity that sums the activity of, and interaction between, other large-scale systems across the brain. The DMN acts to coordinate network integration to influence the body's response to events, thereby supporting flexible, adaptive behavior in complex environments. It is from this activity – which creates "a center of narrative gravity" – that our sense of ourselves emerges.

Christopher G. Davey[1,2], Ben J. Harrison[3]
[1]Orygen, The National Centre of Excellence in Youth Mental Health, Parkville, VIC, Australia; [2]Centre for Youth Mental Health, University of Melbourne, Parkville, VIC, Australia; [3]Melbourne Neuropsychiatry Centre, Department of Psychiatry, University of Melbourne, Parkville, VIC, Australia

1. James W. The principles of psychology. New York: Holt & Co., 1890.
2. Damasio A. Self comes to mind: constructing the conscious brain. New York: Pantheon, 2010.
3. Gallagher S. Trends Cogn Sci 2000;4:14-21.
4. Dennett D. In: Kessel F, Cole P, Johnson D (eds). Self and consciousness: multiple perspectives. Hillsdale: Erlbaum, 1992.
5. Shulman GL, Corbetta M, Buckner RL et al. J Cogn Neurosci 1997;9:648-63.
6. Raichle ME, MacLeod AM, Snyder AZ et al. Proc Natl Acad Sci USA 2001; 98:676-82.
7. Davey CG, Pujol J, Harrison BJ. Neuroimage 2016;132:390-7.
8. van den Heuvel MP, Sporns O. J Neurosci 2011;31:15775-86.
9. Price JL, Drevets WC. Trends Cogn Sci 2012;16:61-71.

DOI:10.1002/wps.20553

Why are savant skills and special talents associated with autism?

Portrayals of autism spectrum disorder in film, television and literature often show special or "savant" skills: a young child who can crack advanced codes, an adult with astonishing memory, or a musician who can play any tune by ear after a single hearing. Are such portrayals realistic or helpful?

Special abilities *are* more common in autism than in other groups, with one study[1] finding that a third of autistic adults showed superior skills in one or more areas by parental report and on psychometric tests. Some well-documented skills are as astounding as any in fiction, such as the renowned artist S. Wiltshire's ability to draw in beautiful detail the cityscape of Tokyo from memory after a single 20 min helicopter ride over the city.

Special skills typically fall into a narrow range of areas. A recent study[2] suggested that more than 70% of autistic children and adults had a special isolated skill in memory (52% of the sample), visuo-spatial abilities (32%), calculation, drawing or music (about 17% for each area).

Parents are often understandably annoyed by the emphasis on savant skills in the media: "When neighbours ask me what my autistic son's special talent is, I tell them it's having a meltdown in the store because the fluorescent lights flicker!", one mother told me. On the other hand, identifying and fostering special interests and abilities can increase self-esteem, opportunities for interaction and appreciation, and employment options for those on the autism spectrum.

What can the study of special skills tell us about autism, and what can their raised incidence in this group teach us about the nature of extreme talents more generally? My first autism research experience was assessing memory skills in mnemonists for N. O'Connor and B. Hermelin, the pioneers of cognitive research in autism, who first demonstrated that no child was ineducable or untestable. Their student and colleague U. Frith continued this tradition, showing that more could be learnt about autism through exploration of assets than of deficits. Her "weak central coherence" theory was built on demonstrations of superior jigsaw-puzzle-like disembedding skills in autism. This "eye for detail" – the tendency to process local rather than global information – may be an important starting engine for talent.

In a study[3] of more than 6,000 8-year-old twins, parent-reported talents in music, maths, art or memory were positively associated with parent-reported autistic-like traits, and specifically with rigid and repetitive interest and activities. Children reported to have special talents were said to show more autistic traits, and in particular to notice and remember details that others miss. Twin modelling (comparing monozygotic and dizygotic cross-twin cross-trait correlations) suggested largely overlapping genetic effects on these two variables: much of the genetic influence on talents was also affecting individual differences in autistic-like rigid/repetitive and eye for detail traits.

That – crudely put – genes "for" talent overlap with genes "for" autism fits with recent evidence suggesting that common alleles associated with autism have been positively selected for during human evolution, and correlate with childhood intelligence and educational attainment in the general population[4].

Detail-focused processing may lie at the root of autistic musical and artistic talent. All musical savants tested to date have shown absolute pitch, but autistic children even without musical training or proficiency are much better able than neurotypical children to hold exact pitch information in mind over days and weeks. While in typical development children move from focus on exact notes to tunes (which can be recognized across different transcriptions), in autism the ability to maintain exact representations appears to be preserved. One multiply talented autistic man could not only name musical tones but also identify the pitch of spoken words or environmental sounds[5].

Many autistic artists have a beautifully detailed style, and some draw from part to neighbouring part rather than sketching the outline first. Eye for detail, and a relative preference for local versus global processing, is characteristic of autism, but can of course be found in neurotypical people. The different cognitive facets of autism appear to "fractionate", with different underlying genetic and neural underpinnings. This means that highly talented people may share a cognitive style with autism, but may not share the socio-communicative difficulties.

Eye for detail is part of the cognitive phenotype of autism, but does not explain everything. Autistic social and communication difficulties appear to result from impaired "theory of mind": while neurotypical children spontaneously and unconsciously read others' behaviour as evidence of invisible mental states ("she's looking in the drawer because she *wants* something she *thinks* is in there"), autistic children and adults do not automatically do so. For them, "reading minds" is hard, conscious calculation – as exhausting as doing long division might be for neurotypicals. The absence of intuitive and even obligatory "mentalizing" in autism may also contribute to talent. People with autism may be less subject to herd thinking, and more able to take original perspectives. Certainly, new evidence suggesting increased perceptual capacity (both visual and auditory) in autism provides a new avenue to understand and explore talent, and possible links to sensory hypersensitivity, distraction and aversion[6].

Along with eye for detail and reduced mentalizing, autism is also characterized by executive dysfunction, most evident when having to deal with change and novelty. The autistic "insistence on sameness" may reflect difficulty in frontal lobe-based executive skills such as planning ahead, monitoring, flexibly shifting set and inhibiting habitual responses. Could executive dysfunction also contribute to talent in autism? Repetition is certainly not the enemy of creativity, as evident in

Monet's haystacks or waterlillies. Beyond this, studies of fronto-temporal dementia have – controversially – suggested a release of talent with the diminishing of frontal functions[7]. This idea has been tested most extremely by A. Snyder, who has claimed to release savant skills (improved drawing, calculation) in healthy adults by inducing a temporary lesion to the left anterior temporal lobe using transcranial magnetic stimulation[8]. While such bold attempts have not been replicated to date, there is considerable interest in the positive aspects of reduced cognitive control for creativity[9].

Increasing genetic evidence supports a dimensional view of autism, with similar genetic influences on diagnosed autism and on autistic traits in the general population[10]. Further research on special skills and talents in autism, therefore, has the promise of finding routes to increase talent beyond those with autism, as well as providing new insights to help recognize, respect and release the "beautiful otherness of the autistic mind".

Francesca Happé
Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK

1. Howlin P, Goode S, Hutton J et al. Philos Trans R Soc Lond B Biol Sci 2009; 364:1359-67.
2. Meilleur AAS, Jelenic P, Mottron L. J Autism Dev Disord 2015;45:1354-67.
3. Happé F, Vital P. Philos Trans R Soc Lond B Biol Sci 2009;364:1369-75.
4. Polimanti R, Gelernter J. PLoS Genet 2017;13:e1006618.
5. Heaton P, Davis RE, Happé F. Neuropsychologia 2008;46:2095-8.
6. Remington A, Fairnie J. Cognition 2017;166:459-65.
7. Miller BL, Hou CE. Arch Neurol 2004;61:842-4.
8. Snyder A. Philos Trans R Soc Lond B Biol Sci 2009;364:1399-405.
9. Amer T, Campbell KL, Hasher L. Trends Cogn Sci 2016;20:905-15.
10. Colvert E, Tick B, McEwen F et al. JAMA Psychiatry 2015;72:415-23.

DOI:10.1002/wps.20552

Progress in achieving quantitative classification of psychopathology

Robert F. Krueger[1], Roman Kotov[2], David Watson[3], Miriam K. Forbes[4], Nicholas R. Eaton[5], Camilo J. Ruggero[6], Leonard J. Simms[7], Thomas A. Widiger[8], Thomas M. Achenbach[9], Bo Bach[10], R. Michael Bagby[11], Marina A. Bornovalova[12], William T. Carpenter[13], Michael Chmielewski[14], David C. Cicero[15], Lee Anna Clark[3], Christopher Conway[16], Barbara DeClercq[17], Colin G. DeYoung[1], Anna R. Docherty[18], Laura E. Drislane[19], Michael B. First[20], Kelsie T. Forbush[21], Michael Hallquist[22], John D. Haltigan[11], Christopher J. Hopwood[23], Masha Y. Ivanova[9], Katherine G. Jonas[2], Robert D. Latzman[24], Kristian E. Markon[25], Joshua D. Miller[26], Leslie C. Morey[27], Stephanie N. Mullins-Sweatt[28], Johan Ormel[29], Praveetha Patalay[30], Christopher J. Patrick[31], Aaron L. Pincus[22], Darrel A. Regier[32], Ulrich Reininghaus[33], Leslie A. Rescorla[34], Douglas B. Samuel[35], Martin Sellbom[36], Alexander J. Shackman[37], Andrew Skodol[38], Tim Slade[39], Susan C. South[35], Matthew Sunderland[39], Jennifer L. Tackett[40], Noah C. Venables[1], Irwin D. Waldman[41], Monika A. Waszczuk[2], Mark H. Waugh[42], Aidan G.C. Wright[43], David H. Zald[44], Johannes Zimmermann[45]

[1]Department of Psychology, University of Minnesota, Minneapolis, MN, USA; [2]Department of Psychiatry, Stony Brook University, Stony Brook, NY, USA; [3]Department of Psychology, University of Notre Dame, Notre Dame, IN, USA; [4]Department of Psychology, Macquarie University, Sydney, NSW, Australia; [5]Department of Psychology, Stony Brook University, Stony Brook, NY, USA; [6]Department of Psychology, University of North Texas, Denton, TX, USA; [7]Department of Psychology, University at Buffalo, State University of New York, New York, NY, USA; [8]Department of Psychology, University of Kentucky, Lexington, KY, USA; [9]Department of Psychiatry, University of Vermont, Burlington, VT, USA; [10]Psychiatric Research Unit, Slagelse Psychiatric Hospital, Slagelse, Denmark; [11]Department of Psychiatry, University of Toronto, Toronto, ON, Canada; [12]Department of Psychology, University of South Florida, Tampa, FL, USA; [13]Department of Psychiatry, University of Maryland, Baltimore, MD, USA; [14]Department of Psychology, Southern Methodist University, Dallas, TX, USA; [15]Department of Psychology, University of Hawaii, Honolulu, HI, USA; [16]Department of Psychology, College of William and Mary, Williamsburg, VA, USA; [17]Department of Developmental, Personality, and Social Psychology, Ghent University, Ghent, Belgium; [18]Department of Psychiatry, University of Utah, Salt Lake City, UT, USA; [19]Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA; [20]Department of Psychiatry, Columbia University, New York, NY, USA; [21]Department of Psychology, University of Kansas, Lawrence, KS, USA; [22]Department of Psychology, Pennsylvania State University, State College, PA, USA; [23]Department of Psychology, University of California at Davis, Davis, CA, USA; [24]Department of Psychology, Georgia State University, Atlanta, GA, USA; [25]Department of Psychology, University of Iowa, Iowa City, IA, USA; [26]Department of Psychology, University of Georgia, Athens, GA, USA; [27]Department of Psychology, Texas A&M University, College Station, TX, USA; [28]Department of Psychology, Oklahoma State University, Stillwater, OK, USA; [29]Department of Psychiatry, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; [30]Institute of Psychology, Health and Society, University of Liverpool, Liverpool, UK; [31]Department of Psychology, Florida State University, Tallahassee, FL, USA; [32]Department of Psychiatry, Uniformed Services University, Bethesda, MD, USA; [33]School for Mental Health and Neuroscience, Maastricht University, Maastricht, The Netherlands; [34]Department of Psychology, Bryn Mawr College, Bryn Mawr, PA, USA; [35]Department of Psychology, Purdue University, West Lafayette, IN, USA; [36]Department of Psychology, University of Otago, Dunedin, New Zealand; [37]Department of Psychology, University of Maryland, College Park, MD, USA; [38]Department of Psychiatry, University of Arizona, Tucson, AZ, USA; [39]National Drug and Alcohol Research Centre, University of New South Wales, Randwick, NSW, Australia; [40]Department of Psychology, Northwestern University, Evanston, IL, USA; [41]Department of Psychology, Emory University, Atlanta, GA, USA; [42]Oak Ridge National Laboratory, University of Tennessee, Oak Ridge, TN, USA; [43]Department of Psychology, University of Pittsburgh, Pittsburgh, PA, USA; [44]Department of Psychology, Vanderbilt University, Nashville, TN, USA; [45]Psychologische Hochschule Berlin, Berlin, Germany

Shortcomings of approaches to classifying psychopathology based on expert consensus have given rise to contemporary efforts to classify psychopathology quantitatively. In this paper, we review progress in achieving a quantitative and empirical classification of psychopathology. A substantial empirical literature indicates that psychopathology is generally more dimensional than categorical. When the discreteness versus continuity of psychopathology is treated as a research question, as opposed to being decided as a matter of tradition, the evidence clearly supports the hypothesis of continuity. In addition, a related body of literature shows how psychopathology dimensions can be arranged in a hierarchy, ranging from very broad "spectrum level" dimensions, to specific and narrow clusters of symptoms. In this way, a quantitative approach solves the "problem of comorbidity" by explicitly modeling patterns of co-occurrence among signs and symptoms within a detailed and variegated hierarchy of dimensional concepts with direct clinical utility. Indeed, extensive evidence pertaining to the dimensional and hierarchical structure of psychopathology has led to the formation of the Hierarchical Taxonomy of Psychopathology (HiTOP) Consortium. This is a group of 70 investigators working together to study empirical classification of psychopathology. In this paper, we describe the aims and current foci of the HiTOP Consortium. These aims pertain to continued research on the empirical organization of psychopathology; the connection between personality and psychopathology; the utility of empirically based psychopathology constructs in both research and the clinic; and the development of novel and comprehensive models and corresponding assessment instruments for psychopathology constructs derived from an empirical approach.

Key words: Psychopathology, mental disorder, personality, nosology, classification, dimensions, clinical utility, Hierarchical Taxonomy of Psychopathology, ICD, DSM, RDoC

(World Psychiatry 2018;17:282–293)

Throughout the history of psychiatric classification, two approaches have been taken to delineating the nature of specific psychopathologies[1]. A first one might be termed authoritative: experts gather under the auspices of official bodies, and delineate classificatory rubrics through group discussions and associated political processes. This approach characterizes official nosologies, such as the DSM and the ICD. It also often characterizes official efforts to influence the constructs and conceptualizations that frame the perspectives of funding bodies. For example, the US National Institute of Mental Health's Research Domain Criteria (RDoC) effort involved the delineation of constructs that were shaped and organized by panels of experts[2].

A second approach might be termed empirical. In this approach, data are gathered on psychopathological building blocks. These data are then analyzed to address specific research questions. For example, does a specific list of symptoms delineate a single psychopathological entity or, by contrast, do those symptoms delineate multiple entities? This approach is sometimes characterized as more "bottom up", compared

with the more "top down" approach of official nosologies. This is because the approach generally starts with basic observations and works to assemble them into classificatory rubrics, rather than working from a set of assumed rubrics to fill in the detailed features of those rubrics.

Obviously, these approaches, although distinguishable, are not entirely separable. Authoritative classification approaches have relied on specific types of empiricism as part of their construction process, and an empirical approach begins with the expertise needed to assemble and assess specific psychopathological building blocks (e.g., signs and symptoms). Nevertheless, it is clear that authoritative approaches tend to weigh putative expertise, disciplinary background, and tradition heavily.

To pick a specific example, the construction of DSM-5 was primarily a psychiatric endeavor, by virtue of the disciplinary background of most participants and by the nature of the body that served to generate and publish the manual (i.e., the American Psychiatric Association). As part of the DSM-5 construction process, field trials were undertaken to evaluate the reliability of specific mental disorder diagnoses. Interestingly, these trials produced a wide range of reliability estimates, encompassing evidence of weak reliability for many common diagnostic entities, such as major depressive disorder and generalized anxiety disorder[3]. In spite of questionable reliability, these constructs remain enshrined in DSM-5 and constitute the official "diagnostic criteria and codes" in Section II of the manual.

Because of these types of sociopolitical dynamics (e.g., asserting the existence of specific psychopathological categories *ex cathedra* despite questionable evidence), authoritative approaches have come under increased scrutiny. Many types and sources of scrutiny coalesce around the scientific disappointments that have accompanied research on diagnostic categories. Simply put, the categories of official nosologies have not provided compelling guidance in the search for etiology and pathophysiology. As a result, the empirical approach to classification is now attracting great interest as a potential alternative to diagnosis by presumed authority and fiat.

In the present paper, we summarize some key types of evidence that have emerged from the burgeoning literature on empirical approaches to psychiatric classification. We focus in particular on: a) evidence pertaining to the continuous versus discrete nature of psychopathological constructs; b) evidence for the hierarchical organizational structure of psychopathological constructs; and c) evidence for specific empirically-based organizational rubrics.

In our discussion of specific empirically-based organizational rubrics, we focus on a consortium that has recently formed to organize and catalyze empirical research on psychopathology, the Hierarchical Taxonomy of Psychopathology (HiTOP) Consortium. As we discuss the work of this consortium, we consider major issues that confront an empirical approach to classification, as it continues to evolve. These issues correspond to existing workgroups in the consortium, and hence, we use the foci of those workgroups to organize our discussion.

Specifically, those workgroups and our discussion are organized around: a) continued research on the organization of broad spectra of psychopathology; b) the connection between personality and psychopathology; c) the utility of constructs derived from an empirical approach (e.g., the ability of these constructs to organize research on pathophysiology); d) translation of empirical research into clinical practice; e) the development of novel and comprehensive models and corresponding assessment instruments for constructs derived from an empirical approach.

THE CONTINUOUS VS. DISCRETE NATURE OF PSYCHOPATHOLOGICAL PHENOTYPES

Perhaps the most fundamental difference between current authoritative psychiatric nosologies and empirical research on psychopathology classification pertains to the continuous vs. discrete nature of constructs. Through tradition and putative authority, authoritative nosologies claim that psychopathologies are organized into discrete diagnostic entities. By contrast, an empirical approach to classification treats the discrete vs. continuous nature of psychopathology as a research question[4]. When treated as a research question, evidence points toward the generally continuous nature of psychopathological variation.

Taxometric evidence

Taxometric methods originated in the writings of P. Meehl, and evaluate the possibility that a set of symptoms (or other indicators of psychopathology) delineate a discrete group. These methods have been used extensively, such that there is now a considerable literature on their application. This literature was summarized quantitatively by Haslam et al[5]. Based on findings from 177 articles, encompassing data from over half a million research participants, psychopathological variation was found to be continuous as opposed to discrete, i.e., there was little consistent evidence for taxa.

Subsequent taxometric reports in diverse areas also tend to reveal greater evidence for continuity as opposed to discreteness. For example, recent taxometric investigations have provided evidence for the continuity of subclinical paranoia and paranoid delusions[6], adolescent substance use[7], and depression in youth[8]. Occasional evidence for potential discreteness is also reported[9,10], emphasizing the importance of ongoing quantitative summaries of this literature.

Psychometric studies of putative taxa are important to establish their validity, such as evaluating stability over time. That is, longitudinal stability of putative taxon membership is also a key means of evaluating a taxonic conjecture, inasmuch as psychopathology taxon membership is conceptualized as a stable property over modest time intervals (e.g., weeks or months). For example, Waller and Ross[11] reported evidence that patho-

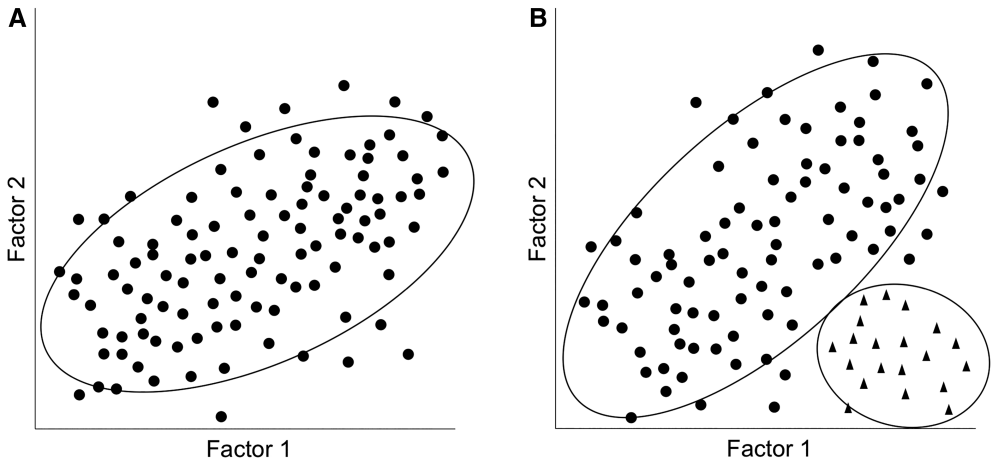

Figure 1 Illustration of hypothetical data compatible with fully continuous and partially discrete models of psychopathological variation. In Panel A, the data points are generally well captured by positing a single group, in which Factor 1 and Factor 2 are positively correlated. In Panel B, the data are better captured by positing two groups, one in which Factor 1 and Factor 2 are positively correlated (the circles), and a second smaller group in which Factor 1 and Factor 2 are weakly negatively correlated (the triangles).

logical dissociation might be taxonic. Watson[12] investigated this putative taxon and found that taxon membership was not stable across a two-month interval, whereas continuous indicators of dissociation were strongly stable.

In sum, extensive evidence suggests that the likelihood of identifying discrete psychopathology groups empirically via taxometrics is not high. By contrast, the taxometrics literature generally points to the continuity of psychopathological variation, emphasizing the greater relative utility and empirical accuracy of continuous as opposed to discrete conceptualizations of psychopathology.

Model-based evidence

Taxometric procedures originally evolved to some extent outside of the mainstream statistical literature. Within the more mainstream literature, approaches have emerged that rely on the ability to fit models to raw data on symptom patterns, and to use all of the extensive information in those data to adjudicate between continuous, discrete and hybrid accounts of psychopathology constructs. These approaches are often termed *model-based*, because they rely on formal statistical models that describe the distributional form of the constructs that underlie symptoms.

Generally, direct comparison of continuous and discrete models via these approaches have indicated that psychopathological constructs tend to be more continuous than discrete[13-19]. Nevertheless, there are also occasional suggestions of potentially meaningful discontinuities, particularly as conceptualized in models that have both continuous and discrete features[20-22].

For example, Figure 1 depicts a bivariate distribution similar to the results found in Forbes et al[20]. Panel A shows a sample where the two continuous factors are moderately correlated for all participants (i.e., all participants are drawn from a single underlying population, akin to the results Forbes et al found for the relationships among depression, anxiety and sexual dysfunctions for women). In contrast, Panel B shows a discontinuity in the data where two groups emerge: the majority of the sample has a strong positive correlation between the factors, but a subgroup of the sample has a weak negative correlation (i.e., participants are drawn from two distinct underlying populations, akin to the results Forbes et al found for men). Generally speaking, the development and comparison of models of latent structure remains a profitable and active area of inquiry, because this approach provides an empirical means of directly comparing and potentially integrating categorical and continuous conceptions of psychopathology[23,24].

However, similar to the situation with potential taxa, the discontinuities need to map truly discrete features of psychopathology (i.e., be reliable and replicable) to be meaningful. Consider, for example, how these requirements played out in a project reported by Eaton et al[25]. In this project, *model based clustering* was used to discern potential discrete personality disorder groups. This approach works well in a variety of scientific areas, when there are actual discontinuities to be detected (e.g., character recognition, tissue segmentation; see http://www.stat.washington.edu/mclust/). Eaton et al therefore applied this approach to a large data set (N=8,690) containing samples from four distinguishable populations (clinical, college, community and military participants). Potential discontinuities observed in each sample were not replicated across samples. By contrast, a dimensional model of the data was readily replicated across the samples. The authors interpreted these findings as suggesting that personality dis-

order features did not delineate replicable discontinuities, but instead, represented replicable continuities.

In sum, efforts to identify potential discontinuities on the basis of data are important endeavors, because they continue to expose dimensional conjectures to risky and direct tests. Nevertheless, similar to what has been learned from decades of taxometric research, the bulk of the existing model-based evidence points to the dimensional nature of psychopathology.

Implications of dimensionality

Evidence to date, stemming from multiple empirical approaches, generally points to the continuity of psychopathological phenotypes. As a result, contemporary empirical approaches often conceptualize psychopathological constructs as dimensional, which has a number of implications. For example, it highlights the extent to which the categories of official nosologies are out of sync with data on the dimensional nature of psychopathology. This disparity is well recognized, and also, very challenging to navigate in a sociopolitical sense, because so many professional endeavors are firmly intertwined with the category labels enshrined in official nosologies[26]. In this paper, we do not detail specific events that have recently played out surrounding this challenge (e.g., pertaining to DSM-5 and ICD-11), but we do note that the challenge needs to be faced head-on if official nosologies aim to be founded on solid empirical footing[27].

We also note here another key implication of the dimensional nature of psychopathology, pertaining to relations between manifest psychopathology and its correlates. Specifically, the continuous nature of psychopathological variation provides a framework for understanding the form and nature of relations between cumulative risk factors, manifest psychopathology, and important outcomes[28]. Consider distal and putatively etiologic correlates, such as specific genetic and environmental risk factors. Continuous phenotypic variation suggests (but does not prove) that the relevant etiologic elements are likely multiple and numerous. Multiple relatively independent causes give rise to continuous phenotypic variation, as is observed with many human phenotypes, e.g. height[29,30]. Similar to physical phenotypes, psychopathological phenotypes are likely the result of specific mixtures of numerous etiologic influences, with both proportions of influence and the resulting phenotypes varying continuously across persons[31].

In sum, the concept of continuous variation among persons in etiologic mixture dovetails well with the observation of continuous phenotypic variation, and provides generative strategies for etiologic research. For example, persons with similar phenotypic values may have arrived at those values in distinct ways. Hence, profitable research strategies might focus less on "cases" and "controls", and more on developing multivariate models of the joint distribution of etiologic (e.g., genomic polymorphisms) and continuous phenotypic observations in larger samples[32].

Turning from causes to consequences, thinking about continuous variation and the public health consequences of psychopathology may also provide novel insights. Although psychopathology appears to be a continuous predictor, the nature of its relationship with public health consequences could take numerous forms, at least in theory. Thinking about this situation may provide insights that go well beyond an artificial "cases vs. controls" research strategy. For example, continuous psychopathology may very well show a monotonically-increasing and generally linear relationship with impairment[33,34]. Or, the relationship could have non-linear features, e.g., accelerating in a certain region of continuous psychopathological variation[22,35].

Again, the key point here is that these possibilities are empirically tractable when psychopathology is modeled dimensionally, yet obscured through the artificial dichotomization that characterizes traditional psychiatric nosologies. Somewhat ironically, continuous measurement of psychopathology is essential to evaluating the possibility that there are meaningful thresholds, beyond which social and occupational dysfunction becomes increasingly more likely.

HIERARCHICAL ORGANIZATIONAL STRUCTURE OF PSYCHOPATHOLOGICAL DIMENSIONS

One perennial issue in developing an empirically-derived and dimensional approach to psychopathology pertains to general organizing principles. In traditional authoritative and categorical approaches to classification, this issue is tacitly addressed by the organizational structure of the classificatory effort. For example, the specific workgroup structure of the DSM-5 construction effort implies an organization of psychopathology into rubrics that reflect the workgroup names, and that structure trickles down into the chapter structure of the printed classification.

Might organizational issues also be addressed empirically? Evidence described in the foregoing section stems from asking if a specific set of signs and symptoms delineates a specific dimension as opposed to a specific category. This evidence suggests that psychopathology is generally dimensional in nature, but how many dimensions are there, and how are these dimensions organized?

Work in this area has generally progressed from asking "what is the correct number of dimensions" to realizing that this question is somewhat specious, because individual difference dimensions (e.g., individual differences in the propensity to experience specific psychopathological signs and symptoms) are organized *hierarchically*. This understanding has been important in resolving a variety of classificatory conundrums, typically focused in areas where two or more psychopathological constructs contain variation that is both shared and unique.

Perhaps the most classic example pertains to anxiety and depression[36]. The tendency to experience pathological anxiety is clearly correlated with the tendency to experience pathological depression, yet these tendencies are also distinguishable. Categorical nosologies

have difficulty managing these situations, because they tend to lead to proposals of "mixed categories" (e.g., a category of mixed anxiety and depression that is putatively distinguishable from a category of anxiety only and a category of depression only). If anxiety and depression are more dimensional than categorical, as well as correlated but not perfectly correlated, then most patients will not fit neatly into any of these three categories. This tends to lead to difficulties making categorical diagnostic determinations in practice. For example, a mixed anxiety-depression category was proposed for DSM-5, but did not emerge from the field trials as a reliable diagnosis[37].

The key to resolving these sorts of dilemmas is to realize that the evidence is most readily compatible with conceptualizing anxiety and depressive phenomena (as well as other dimensional phenomena) as encompassed by *hierarchically organized dimensions*. To illustrate this point concretely, consider a model developed by Waszczuk et al[38], portrayed in Figure 2. This model, which is based on extensive data, shows how specific anxiety and depressive phenomena are associated with continuous degrees of similarity and distinctiveness, across four hierarchically arranged levels of generality vs. specificity. These hierarchical levels reflect the overall degree of empirical co-occurrence vs. distinctiveness of the phenomena encompassed by the model. Concepts higher in the figure are more general and broad, whereas concepts lower in the figure are more specific and narrow.

At the most general level, diverse anxious and depressive phenomena are understood to be aspects of a general domain of internalizing psychopathology. However, as is apparent in both data and clinical work in this area, although anxious and depressive phenomena are indeed correlated, they are not perfectly correlated and, therefore, are distinguishable from one another. Hence, one level down, distinctions emerge among distress, fear, and obsessive-compulsive (OCD)/manic phenomena. Note that this is a more refined and empirically based understanding when compared with DSM chapter headings, because, rather than being delineated by individual committees, this model uses data to encompass the breadth of phenomena that fall into the internalizing domain.

Accordingly, at a third level of specificity, key distinctions emerge among aspects of the three distress, fear and OCD/mania domains. OCD and mania are distinguishable at this level, as are specific aspects of these broader domains, such as the cognitive and vegetative aspects of depression. Indeed, considered across levels, these patterns have fundamental conceptual and clinical implications. For example, these patterns highlight the connection between OCD and manic phenomena, as well as their distinctiveness from distress and fear. This may be traceable to the connection that OCD and manic phenomena share with the broad spectrum of psychosis, and how this psychotic aspect both drives OCD and mania together, and separates them from other parts of the internalizing spectrum[39]. Finally, at the lowest level of the hierarchy lie specific symptom clusters, such as checking, lassitude, and so on.

In sum, the Figure 2 model solves the problem of "comorbidity between anxiety and depression" by using data to model the empirical organization of emotional disorder phenomena. Rather than forcing these phenomena into committee-derived categories, they are modeled as they are in nature. As a result, "complex presentations" (e.g., persons who present with a mix of emotional disorder symptoms) are handled because these presentations can be readily represented by a specific profile of problems. This understanding then drives case conceptualization in the clinic[40], and strategies for identifying key correlates (e.g., neural response) in the laboratory[41].

Evidence for dimensional hierarchies can be found throughout psychopathology, and is not limited to anxiety and mood phenomena. Indeed, this evidence is sufficiently comprehensive that it has formed the basis for a consortium of researchers interested in empirical approaches to psychopathology, the HiTOP Consortium[42]. We turn now to describe the main features of the model that frames HiTOP, as well as the issues and topics that are currently being pursued within HiTOP.

EVIDENCE FOR SPECIFIC EMPIRICALLY-BASED ORGANIZATIONAL RUBRICS

Given evidence that psychopathological phenotypes are dimensional in nature, and that these dimensions are organized hierarchically, what types of classificatory rubrics emerge in an empirical hierarchy of psychopathological dimensions? The HiTOP Consortium focuses on these and related issues.

The consortium currently consists of 70 investigators with backgrounds in diverse disciplines (e.g., psychology, psychiatry and philosophy), and this group has proposed a working dimensional and hierarchical model, derived from the literature on empirical psychopathology classification. This model is portrayed in Figure 3.

The model is not intended to be the final word on empirical psychopathology classification. Indeed, the purpose of articulating this model was to provide a first draft that might frame continued inquiry, and thereby move discourse away from tendentious debates about various reified classification schemes. Nevertheless, the model does summarize a substantial literature, reviewed by Kotov et al[43] as background for the hierarchical structure portrayed in Figure 3. Here, we will briefly outline the main features of the model, and then turn to discuss various workgroups within the consortium, which formed to address major issues in the field of empirical psychopathology classification.

As portrayed in Figure 3, the working HiTOP model is hierarchical in nature. Constructs higher in the figure summarize the tendencies for constructs lower in the figure to co-occur in specific patterns. For example, consistent with Figure 2, the broad internalizing spectrum in Figure 3 encompasses more specific "sub-spectra" such as the fear, distress and mania spectra. However, the model in Figure 3 was intended to synthesize the entire available literature on empiri-

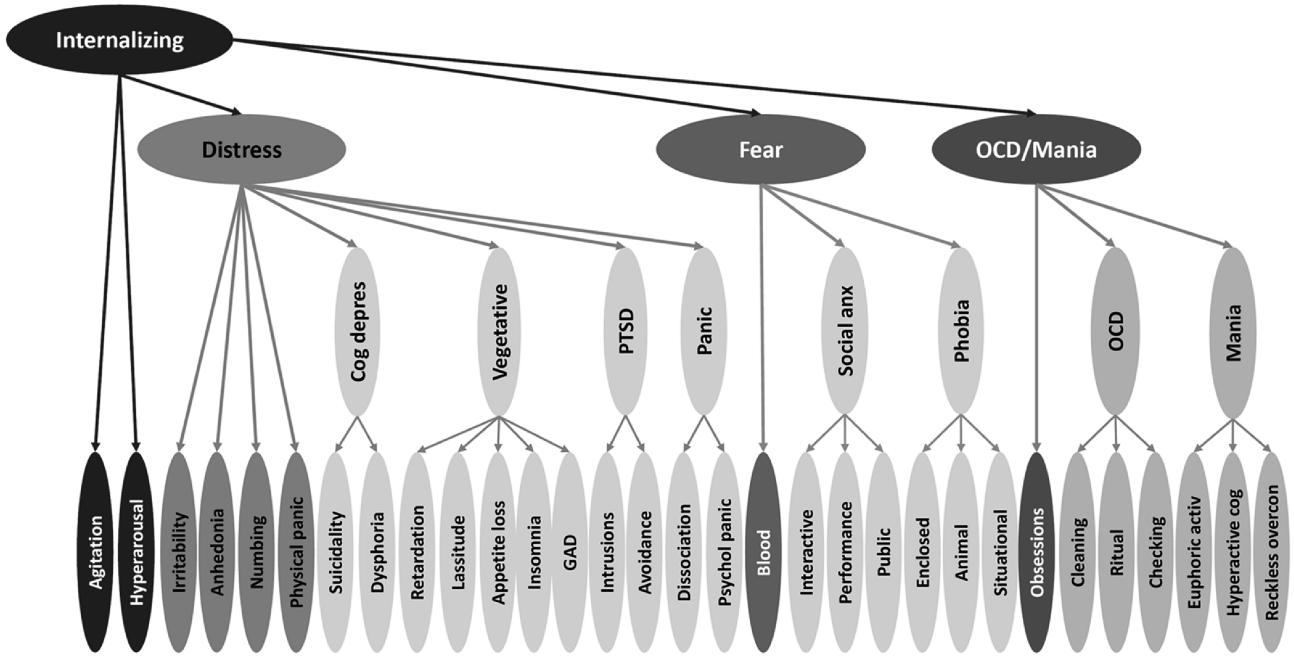

Figure 2 Illustration of an empirically based model of the internalizing spectrum. Constructs higher in the figure are broader and more general, whereas constructs lower in the figure are narrower and more specific (adapted from Waszczuk et al[38]). PTSD – post-traumatic stress disorder, Social anx – social anxiety, OCD – obsessive-compulsive disorder, GAD – generalized anxiety disorder, Cog depress – cognitive depression, Psychol panic – psychological panic, Euphoric activ – euphoric activation, Hyperactive cog – hyperactive cognition, Reckless overcon – reckless overconfidence.

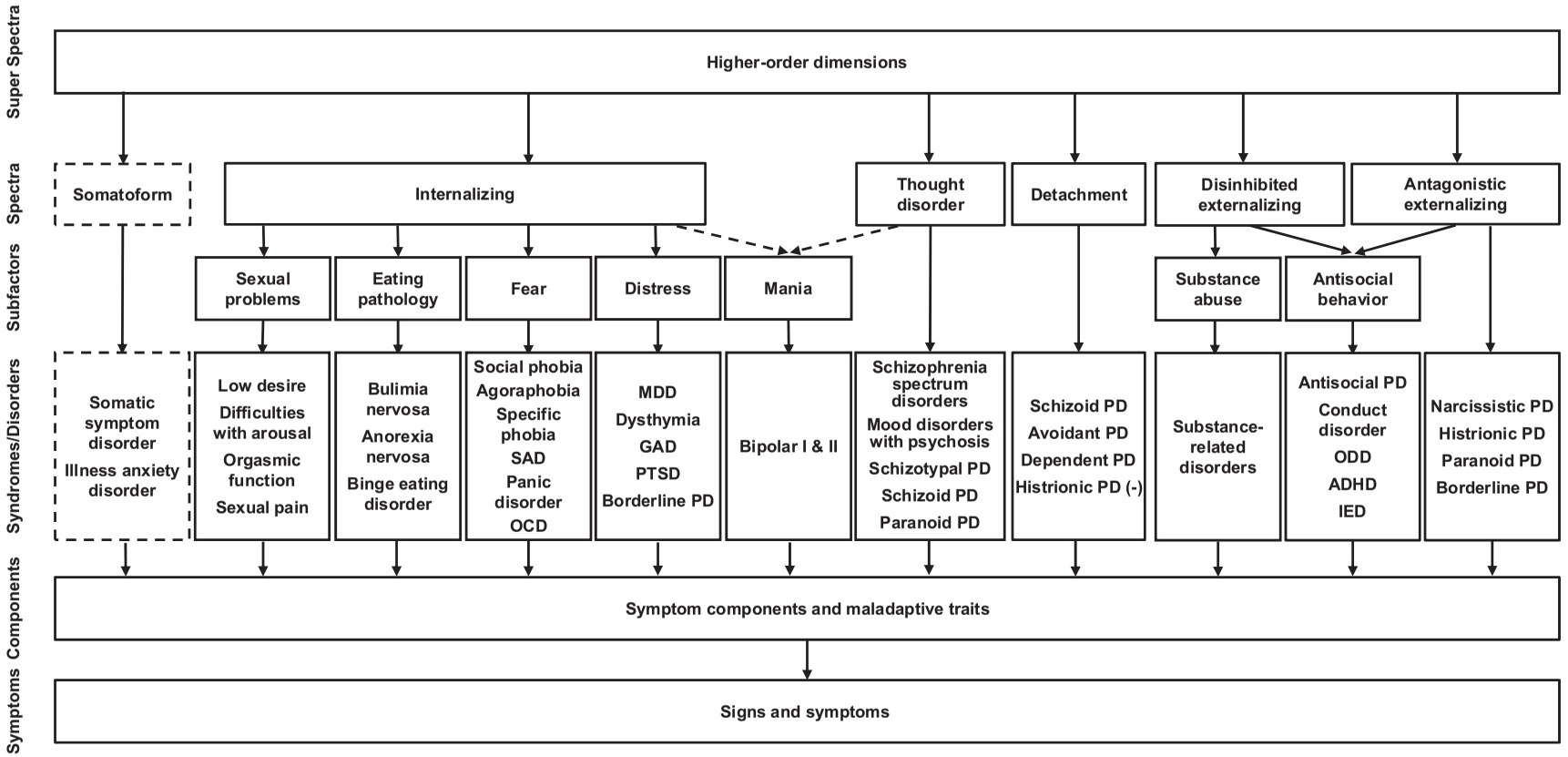

Figure 3 Working Hierarchical Taxonomy of Psychopathology (HiTOP) consortium model. Constructs higher in the figure are broader and more general, whereas constructs lower in the figure are narrower and more specific (adapted from Kotov et al[43]). SAD – separation anxiety disorder, OCD – obsessive-compulsive disorder, MDD – major depressive disorder, GAD – generalized anxiety disorder, PTSD – post-traumatic stress disorder, PD – personality disorder, ODD – oppositional defiant disorder, ADHD – attention-deficit/hyperactivity disorder, IED – intermittent explosive disorder.

cal classification and, as a result, its scope and breath is considerably larger than the Figure 2 model, which was designed specifically to delineate the internalizing spectrum.

Consider spectra adjacent to internalizing in the Figure 3 model. In addition to the internalizing spectrum, five other major empirical divisions of psychopathology are portrayed on the same level. Currently, the model posits major spectra labeled *somatoform, thought disorder, detachment, disinhibited externalizing,* and *antagonistic externalizing.* These concepts are reminiscent of, but not necessarily coterminous with, similar constructs in existing authoritative nosologies such as the DSM and ICD. For example, the current HiTOP model posits the existence of a somatoform spectrum that is separable from other major psychopathology spectra, and roughly similar in content to somatoform diagnoses in DSM-5.

While the evidence for the somatoform spectrum is limited (as indicated by the dashed lines in Figure 3), this spectrum illustrates a general principle of empirical classification research. Phenomena that are not explicitly considered within a specific scope can be considered by expanding that scope accordingly. For example, somatoform constructs are not as heavily researched as other phenomena on the level of major spectra (e.g., internalizing and externalizing), and this provides an important opportunity for targeted and focused research[44]. Specifically, how closely do somatoform concepts align with other spectrum concepts, and what are the shared and distinguishing features of these concepts?

Rather than being handled in relatively insular literatures aligned with traditional classificatory rubrics, the HiTOP framework provides novel opportunities for more targeted and synthetic research on key empirical questions in classification. For example, how do somatoform phenomena covary with other phenomena in the HiTOP model? Are they better understood as an aspect of the broader internalizing spectrum, or are they sufficiently distinguished to form their own separate spectrum? If they have both shared and distinctive features, are intervention efforts more effective if focused on the shared features, or on the distinctive features? Such questions are posed and framed by thinking about somatoform phenomena in the context of psychopathology broadly, in ways that go well beyond a more piecemeal approach to parsing and conceptualizing psychopathology.

Similar to the situation with the somatoform spectrum, other constructs on the spectra level have varying volumes of associated literature, as well as being associated with specific arrangements portrayed in Figure 3. Recognizing these hypothesized arrangements provides generative avenues for novel research. Consider examples pertinent to each of the spectra in Figure 3. The thought disorder spectrum reflects the close empirical connections among psychotic phenomena that have historically been divided between more dispositional vs. more acute manifestations[45,46]. This empirical distinction thereby becomes a topic for continuing empirical inquiry, and not an issue presumably settled by the unfortunate tradition of studying personality and clinical disorders in separate literatures[47].

For example, the ICD-11 proposal for personality disorders does not encompass a psychoticism domain, not because psychotic phenomena are outside of a comprehensive multivariate model of maladaptive personality, but rather because tradition places them in a different chapter within the ICD (and in contrast with the DSM, which assigns schizotypal disorder primarily to the personality disorders chapter, with a secondary assignment as part of the schizophrenia spectrum in the schizophrenia and other psychotic disorders chapter[48]). Likewise, antisocial personality disorder is assigned both to the personality disorder and the disruptive, impulse control and conduct disorders chapter. In the HiTOP approach, these sorts of fundamental issues become topics for empirical inquiry.

Similar issues are addressed by the two externalizing spectra portrayed in Figure 3. The current HiTOP model reflects the distinction between the two major aspects of externalization: antagonism (hurting others intentionally) and disinhibition (acting on impulse or in response to a current stimulus, with little consideration of consequences[49]). As such, it also reflects the ways in which these separable aspects are both present in traditional DSM diagnostic criteria sets. For example, DSM-IV defined antisocial personality disorder, and similar DSM diagnostic concepts, represent a mix of antagonistic and disinhibited features[50]. The HiTOP model posits that separating these empirically-based features may result in greater clarity regarding the classification of specific phenomena. For example, the model posits a closer connection between substance related disorders and disinhibition than between substance related disorders and antagonism. In addition, the model ties together closely aligned externalizing phenomena that are spread throughout DSM chapters and various literatures (e.g., child and adult manifestations of basic antagonistic tendencies, as well as phenomena such as intermittent explosive disorder).

Finally, consider the detachment (avoidance of socioemotional engagement) spectrum portrayed in Figure 3. Similar to somatoform phenomena, detachment phenomena have not been as heavily studied as other major spectra. In addition, similar to externalizing phenomena, detachment has been somewhat diffused throughout traditional nosologies, being captured within the features of a number of traditional personality disorders. The HiTOP model recognizes the evidence that detachment appears to be a major spectrum of adult psychopathology. As such, the model underlines the importance of understanding the public health significance of pathological socioemotional avoidance, as opposed to spreading this feature across constructs that have attracted relatively less clinical and research attention, compared with more florid manifestations of psychopathology.

Below the level of spectra in Figure 3 are levels encompassing subfactors and disorders. These concepts reflect a mix of more traditional and more empirically based rubrics. The presence of traditional diagnostic labels on Figure 3 is not to reify these concepts (many of which are highly heterogeneous, and therefore in need of

empirical refinement), but rather, to provide a cross walk to traditional and familiar DSM-style labels. As the model implies, the heterogeneity of these phenomena provides important opportunities for clarifying investigations.

Consider, for example, borderline personality disorder (BPD), which is listed below both the distress and antagonistic externalizing rubrics in the working HiTOP model. BPD encompasses a number of distinguishable elements and, as a result, tends to be associated with diverse psychopathology spectra[51,52]. Indeed, the majority of the variance in BPD is shared with other forms of psychopathology (rather than being unique to it), emphasizing the importance of reducing BPD and similar constructs to their constituent elements, and working to reconstitute those elements in an empirical manner.

This type of refinement endeavor has been clarifying in specific literatures where it has been undertaken. For example, empirical efforts underlie large segments of the DSM-5 alternative personality disorder model, and frame the essential structure of the ICD-11 personality disorder approach, in ways that go fundamentally beyond traditional personality disorder rubrics. Thinking broadly, the HITOP model underlines the general utility of this type of empirical refinement endeavor, pursued with regard to psychopathology writ large.

THE HIERARCHICAL TAXONOMY OF PSYCHOPATHOLOGY CONSORTIUM (HiTOP) AS A FRAMEWORK FOR CONTINUED PROGRESS

HiTOP is intended to serve as a consortium to organize and stimulate progress on an empirical approach to classifying psychopathology. To facilitate this progress, the consortium is organized into a series of workgroups. The workgroup rubrics do not exhaust all the important issues that might be addressed in empirical psychopathology classification. Nevertheless, they do reflect themes that have emerged to organize current HiTOP efforts. Importantly, membership in HiTOP is not closed, and there are many opportunities to get involved in various aspects of the endeavor[42].

Higher-order dimensions workgroup

A significant challenge posed by the model in Figure 3 is its breadth. As implied by the distinction between Figure 2 and Figure 3 (i.e., the distinction between detail and breadth), many empirical classification efforts have been understandably focused on specific spectra of psychopathology. Above the level of internalizing in Figure 3 is the "super spectra" level, which is currently open, largely because relations among various psychopathology spectra remains an active area of empirical inquiry. For example, there has been recent interest in a general psychopathology dimension, akin to the general dimension found in the cognitive abilities literature[53,54].

Although there is little doubt that variation in psychopathology spectra is generally correlated (i.e., multi-morbidity is encountered frequently), important issues remain to be addressed in contemplating the organizational structure of psychopathology above the spectrum level. For example, for a hierarchical construct to be "truly general", its influence on constructs below it in a hierarchy should be relatively uniform. Contrary to this conceptualization, the magnitude of influence of the general psychopathology factor on specific constructs below it has not been necessarily uniform. For example, Caspi et al[53] modeled a general factor of psychopathology and found it to be associated primarily with psychotic phenomena. Lahey et al[54] also modeled a general factor of psychopathology, but found it to be associated primarily with phenomena that fall generally into the distress subdomain of internalizing (albeit they did not specifically study psychotic phenomena).

These distinctions between various representations of the general factor of psychopathology may relate to important technical issues surrounding the meaning and interpretation of a general factor. For example, technical issues have arisen in the literature on individual differences in cognitive test performance. In that literature, it is now understood that ways of modeling general factors (e.g., using a bifactor versus a hierarchical structural model), and ways of comparing models (e.g., based on fit indices), differ in subtle but important ways from many traditional approaches to structural modeling[55-57]. These issues have yet to be addressed thoroughly in the psychopathology literature, and are therefore a focus of current activity in the higher order workgroup.

Furthermore, we note that the breadth of psychopathology in various studies of potential general factors is less than the breadth of psychopathology encompassed in Figure 3. How to efficiently assess (and thereby have the opportunity to model) the entire breadth of psychopathology covered by Figure 3 presents an important – and daunting – challenge. In addition, the current model does not encompass the neurodevelopmental spectrum (e.g., intellectual disability, autism spectrum disorders, learning disorders), the neurocognitive disorders, and the paraphilic disorders.

Measures development workgroup

Many existing measures assess different aspects of the HiTOP scheme (see https://psychology.unt.edu/hitop). Nevertheless, as of this writing, a comprehensive measure designed to assess the entire breadth of psychopathology covered in Figure 3 does not exist. The measures development workgroup in HiTOP was created to address this issue directly. The related but distinct goals of the measurement workgroup are to: a) simultaneously develop measures for all proposed symptom dimensions and personality traits encompassed by HiTOP in the service of empirically refining the model through psychometrically rigorous structural work, and b) based on this work, developing clinical useful tools designed to permit researchers and mental health practitioners to reliably, validly and efficiently assess all components of the HiTOP model.

In the service of building clinically useful tools, which is an important translational goal of HiTOP more generally, a number of fundamental measurement issues arise. We list just a few here to give a feel for some of the challenges ahead. For example, if the conceptualization of psychopathology is dimensional, should skip-outs (or other adaptive techniques) be employed to enhance the efficiency of assessment (akin to skip-outs designed on a rational basis to enhance the efficiency of traditional category assessment via structured interview)? Traditionally, dimensional approaches to psychopathology have been more closely associated with questionnaire as opposed to interview assessment strategies (because of the close intellectual and historical connections between psychometrics and questionnaire development). How can interview approaches – often favored in clinical research contexts – be developed that reflect more dimensional conceptualizations (e.g., the Structured Interview for the Five Factor Model[58] and the Interview for Mood and Anxiety Symptoms[38])? In addition, assessment of traditional categories via interview is typically modularized; only specific modules are used in many assessments, consistent with the constructs targeted. Can or should dimensional assessment be similarly modularized? Is this even possible or desirable, given the evidence portrayed in Figure 3, that all varieties of psychopathology are positively correlated? Finally, how can transient symptom manifestations and chronic maladaptive trait characteristics be seamlessly integrated within a single instrument?

Normal personality workgroup

The resemblance between the model portrayed in Figure 3 and well-established models of human personality variation, particularly the prominent Five Factor Model[59], is clear. This resemblance is not accidental, but rather reflects the ways in which personality forms the empirical psychological infrastructure for the development of specific varieties of psychopathological symptoms[59]. Nevertheless, a number of interesting and important issues arise in recognizing the intertwined nature of variation in personality and psychopathology.

For example, as noted earlier, the model in Figure 3 reflects empirical connections based on extant literature that was framed by constructs that vary in their associated presumed periodicity. By tradition, DSM frames some disorders as more episodic (e.g., mood disorders), and other disorders are more dispositional (e.g., personality disorders). Stepping back from this act of historical fiat, what in actuality are the distinctions between more dispositional personality constructs, and more acute symptom constructs? Both seem important in comprehensive case conceptualization but, practically and empirically, what strategies might help to parse similarities and differences, yet also unify them in a more comprehensive model? These are the sorts of issues that fall into the bailiwick of the HiTOP normal personality workgroup.

Utility workgroup

Implicit in articulating the type of model portrayed in Figure 3 is the idea that this model has utility, i.e., that it can do some useful work in the world that will help to propel research and clinical practice. The role of the utility workgroup is to realize this potential explicitly. A number of examples might be mentioned, but those that seem particularly salient involve connections of empirical psychopathological phenotypes with neural mechanisms and genomic variants, given contemporary funding priorities. The biomedical research enterprise (e.g., the basic paradigm framing funding bodies such as the US National Institutes of Health) prioritizes the role of fundamental biological processes in addressing issues in public health. This prioritization reflects the success of this paradigm in addressing many health problems during the 20th century. Accordingly, there is substantial interest and financial investment in understanding the neural bases of manifest psychopathology.

HiTOP constructs have a key role to play in furthering this endeavor. For example, the RDoC initiative has sometimes been criticized for providing limited guidance in conceptualizing clinical psychopathology *per se*. This may in some ways reflect a disjunction between what RDoC has aimed to achieve, and what investigators are seeking. To our reading, RDoC aimed to focus attention and effort on more fundamental neurobiological constructs as promising topics for research. The intent was not necessarily to re-conceptualize phenotypic psychopathology[60]. In this way, HiTOP represents a necessary and desirable counterpart to RDoC. The interface between the neurobiological constructs of RDoC and the more phenotypic constructs of HiTOP represents a key means of connecting structure and process in understanding psychopathology.

Clinical translation workgroup

Although traditional nosologies are framed by their category labels, dimensional approaches to psychopathology are also clearly part and parcel of clinical practice. Psychosocial and pharmacological intervention strategies often are effective because they track clinically salient clusters of symptom dimensions[61]. Indeed, dimensional conceptualization and corresponding intervention strategies are arguably (if not always explicitly) the essence of clinical practice[62]. Triage is often a matter of matching the intensity of the presentation with the intensity of intervention. In routine clinical practice, the key decision is not typically "to treat or not to treat". Rather, the key decision is "what level of intervention best suits this level of need?".

To pick a specific example, persons presenting with substance use problems are not clinically homogenous in their level of problems and corresponding need for a specific treatment approach (indeed, the DSM-5's more dimensional conceptualization of substance use disorder reflects this reality). Instead, milder presentations can often be treated effectively through outpatient detoxification (assuming medi-

cal stabilization); more severe presentations often benefit from more structured approaches (e.g., partial hospitalization); and very severe presentations often require at least an initial inpatient stay (e.g., for purposes of medical stabilization). As this example makes clear, conceptualizing substance use presentations as "present vs. absent" would be fundamentally at odds with routine and responsible clinical practice[63]. The clinical translation workgroup serves to make these sorts of dimensional considerations more explicit, and to help disseminate specific dimensionally-oriented approaches to front-line clinicians.

SUMMARY AND CONCLUSIONS

There has been considerable recent interest in empirical approaches to psychopathology classification. This interest has arisen for various reasons, but arguably, the overarching consideration and motive is to place classification on an empirical playing field, as opposed to relying more on the political considerations that influence traditional nosological endeavors, such as the DSM revision process.

This empirical classification movement is well intended, but numerous challenges remain. For example, will progress result more from a distributed approach, or from a more centrally organized approach? In many sciences, a distributed approach facilitates progress. Laboratories compete for resources, and seek to replicate other laboratories' work. Classification of psychopathology, however, presents different kinds of scientific and practical challenges. For example, there is a need for coherence in conceptualizing the entire breadth of the subject matter. This need is arguably more acute than in many more focal scientific endeavors. That is, a piecemeal classification would have limited utility in portraying the entire picture, and portraying the entire picture is a key goal in addressing the limitations of extant schemes (e.g., the generally piecemeal nature of category-driven research efforts).

The HiTOP Consortium formed as a way of addressing this need for breadth and coherence, closely tethered to data. However, HiTOP, like endeavors before it, is a consortium of human clinicians, scientists and scholars, each with their own unique perspectives, in addition to their shared goals. Although focused squarely on the role of data in adjudicating nosological controversies via its principles[42], how will HiTOP navigate new evidence, which, after all, is not self-interpreting? We are optimistic that these challenges can (and indeed must) be surmounted, because moving toward a more empirical approach is critical to the ultimate intellectual health and credibility of the field.

The next phase in the development of HiTOP and the broader field of empirical psychopathology classification may prove to be a watershed in arriving at a data-based approach to age old questions in classification, and therefore, a system that bridges and unifies both research and clinical practice in mental health.

ACKNOWLEDGEMENTS

R.F. Krueger is supported by the US National Institutes of Health, NIH (R01AG053217, U19AG051426) and the Templeton Foundation; A.J. Shackman by the US NIH (DA040717 and MH107444) and the University of Maryland, College Park; A. Wright by the US National Institute of Mental Health, NIMH (L30MH101760); N.C. Venables by the US National Institute of Drug Abuse (T320A037183); U. Reininghaus by the Netherlands Organization for Scientific Research (451-13-022). The opinions expressed in this paper are those of the authors and do not necessarily reflect those of the funding sources.

REFERENCES

1. Blashfield R. The classification of psychopathology: neo-kraepelinian and quantitative approaches. Berlin: Springer, 1984.
2. Cuthbert BN. The RDoC framework: facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. World Psychiatry 2014;13:28-35.
3. Regier DA, Narrow WE, Clarke DE et al. DSM-5 field trials in the United States and Canada, Part II: test-retest reliability of selected categorical diagnoses. Am J Psychiatry 2013;170:59-70.
4. Lubke GH, Miller PJ. Does nature have joints worth carving? A discussion of taxometrics, model-based clustering and latent variable mixture modeling. Psychol Med 2015;45:705-15.
5. Haslam N, Holland E, Kuppens P. Categories versus dimensions in personality and psychopathology: a quantitative review of taxometric research. Psychol Med 2012;42:903-20.
6. Elahi A, Perez Algorta G, Varese F et al. Do paranoid delusions exist on a continuum with subclinical paranoia? A multi-method taxometric study. Schizophr Res 2017;190:77-81.
7. Liu RT. Substance use disorders in adolescence exist along continua: taxometric evidence in an epidemiological sample. J Abnorm Child Psychol 2017;45:1577-86.
8. Liu RT. Taxometric evidence of a dimensional latent structure for depression in an epidemiological sample of children and adolescents. Psychol Med 2016;46:1265-75.
9. Morton SE, O'Hare KJM, Maha JLK et al. Testing the validity of taxonic schizotypy using genetic and environmental risk variables. Schizophr Bull 2017;43:633-43.
10. Witte T, Holm-Denoma J, Zuromski K et al. Individuals at high risk for suicide are categorically distinct from those at low risk. Psychol Assess 2017;29:382-93.
11. Waller NG, Ross CA. The prevalence and biometric structure of pathological dissociation in the general population: taxometric and behavior genetic findings. J Abnorm Psychol 1997;106:499-510.
12. Watson D. Investigating the construct validity of the dissociative taxon: stability analyses of normal and pathological dissociation. J Abnorm Psychol 2003;112:298-305.
13. Carragher N, Krueger RF, Eaton NR et al. ADHD and the externalizing spectrum: direct comparison of categorical, continuous, and hybrid models of liability in a nationally representative sample. Soc Psychiatry Psychiatr Epidemiol 2014;49:1307-17.
14. Conway C, Hammen C, Brennan PA. A comparison of latent class, latent trait, and factor mixture models of DSM-IV borderline personality disorder criteria in a community setting: implications for DSM-5. J Pers Disord 2012;26:793-803.
15. Krueger RF, Markon KE, Patrick CJ et al. Externalizing psychopathology in adulthood: a dimensional-spectrum conceptualization and its implications for DSM-V. J Abnorm Psychol 2005;114:537-50.
16. Vrieze SI, Perlman G, Krueger RF et al. Is the continuity of externalizing psychopathology the same in adolescents and middle-aged adults? A test of the externalizing spectrum's developmental coherence. J Abnorm Child Psychol 2012;40:459-70.
17. Walton KE, Ormel J, Krueger RF. The dimensional nature of externalizing behaviors in adolescence: evidence from a direct comparison of categorical, dimensional, and hybrid models. J Abnorm Child Psychol 2011;39:553-61.
18. Wright AGC, Krueger RF, Hobbs MJ et al. The structure of psychopathology: toward an expanded quantitative empirical model. J Abnorm Psychol 2013;122:281-94.
19. Eaton NR, Krueger RF, Markon KE et al. The structure and predictive validity of the internalizing disorders. J Abnorm Psychol 2013;122:86-92.
20. Forbes MK, Baillie AJ, Schniering CA. Where do sexual dysfunctions fit into the meta-structure of psychopathology? A factor mixture analysis. Arch Sex Behav 2016;45:1883-96.
21. Kramer MD, Arbisi PA, Thuras PD et al. The class-dimensional structure of PTSD before and after deployment to Iraq: evidence from direct comparison of dimensional, categorical, and hybrid models. J Anxiety Disord 2016;39:1-9.

22. Klein DN, Kotov R. Course of depression in a 10-year prospective study: evidence for qualitatively distinct subgroups. J Abnorm Psychol 2016;125:337-48.
23. Lubke GH, Luningham J. Fitting latent variable mixture models. Behav Res Ther 2017;98: 91-102.
24. Masyn KE, Henderson CE, Greenbaum PE. Exploring the latent structures of psychological constructs in social development using the dimensional-categorical spectrum. Soc Dev 2018;12:82-6.
25. Eaton NR, Krueger RF, South SC et al. Contrasting prototypes and dimensions in the classification of personality pathology: evidence that dimensions, but not prototypes, are robust. Psychol Med 2011;41:1151-63.
26. Zachar P, Krueger R, Kendler K. Personality disorder in DSM-5: an oral history. Psychol Med 2015;46:1-10.
27. Hopwood CJ, Kotov R, Krueger RF et al. The time has come for dimensional personality disorder diagnosis. Personal Ment Health 2018;12:82-6.
28. Forbes M, Tackett J, Markon K et al. Beyond comorbidity: toward a dimensional and hierarchical approach to understanding psychopathology across the life span. Dev Psychopathol 2016;28:971-86.
29. Lello L, Avery SG, Tellier L et al. Accurate genomic prediction of human height. bioRxiv (in press).
30. von Eye A, DeShon RP. Directional dependence in developmental research. Int J Behav Dev 2012;36:303-12.
31. Kendler KS. A joint history of the nature of genetic variation and the nature of schizophrenia. Mol Psychiatry 2015;20:77-83.
32. Okbay A, Baselmans B, De Neve J et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. Nat Genet 2016;48:624-33.
33. Markon KE. How things fall apart: understanding the nature of internalizing through its relationship with impairment. J Abnorm Psychol 2010;119:447-58.
34. Jonas K, Markon K. A model of psychosis and its relationship with impairment. Soc Psychiatry Psychiatr Epidemiol 2013;48:1367-75.
35. Wakschlag LS, Estabrook R, Petitclerc A et al. Clinical implications of a dimensional approach: the normal: abnormal spectrum of early irritability. J Am Acad Child Adolesc Psychiatry 2015;54: 626-34.
36. Clark LA, Watson D. Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. J Abnorm Psychol 1991; 100:316-36.
37. Möller HJ, Bandelow B, Volz HP et al. The relevance of 'mixed anxiety and depression' as a diagnostic category in clinical practice. Eur Arch Psychiatry Clin Neurosci 2016;266:725-36.
38. Waszczuk M, Kotov R, Ruggero C et al. Hierarchical structure of emotional disorders: from individual symptoms to the spectrum. J Abnorm Psychol 2017;126:613-34.
39. Chmielewski M, Watson D. The heterogeneous structure of schizotypal personality disorder: item-level factors of the schizotypal personality questionnaire and their associations with obsessive-compulsive disorder symptoms, dissociative tendencies, and normal personality. J Abnorm Psychol 2008;117:364-76.
40. Barlow DH, Farchione TJ, Bullis JR et al. The unified protocol for transdiagnostic treatment of emotional disorders compared with diagnosis-specific protocols for anxiety disorders: a randomized clinical trial. JAMA Psychiatry 2017; 74:875-84.
41. Venables NC, Yancey JR, Kramer MD et al. Psychoneurometric assessment of dispositional liabilities for suicidal behavior: phenotypic and etiological associations. Psychol Med 2018;48: 463-72.
42. Stony Brook School of Medicine. The Hierarchical Taxonomy of Psychopathology (HiTOP). https://medicine.stonybrookmedicine.edu/HITOP.
43. Kotov R, Krueger RF, Watson D et al. The Hierarchical Taxonomy of Psychopathology (HiTOP): a dimensional alternative to traditional nosologies. J Abnorm Psychol 2017;126:454-77.
44. Forbes M, Kotov R, Ruggero C et al. Delineating the joint hierarchical structure of clinical and personality disorders in an outpatient psychiatric sample. Compr Psychiatry 2017;79:19-30.
45. Carpenter WT, Bustillo JR, Thaker GK et al. The psychoses: cluster 3 of the proposed metastructure for DSM-V and ICD-11. Psychol Med 2009;39:2025-42.
46. Guloksuz S, van Os J. The slow death of the concept of schizophrenia and the painful birth of the psychosis spectrum. Psychol Med 2018; 48:229-44.
47. Kotov R, Watson D, Krueger RF et al. Thought disorder spectrum of the Hierarchical Taxonomy of Psychopathology (HiTOP): bridging psychosis and personality pathology. Submitted for publication.
48. Tyrer P, Crawford M, Mulder R. Reclassifying personality disorders. Lancet 2011;377:1814-5.
49. Krueger RF, Markon KE, Patrick CJ et al. Linking antisocial behavior, substance use, and personality: an integrative quantitative model of the adult externalizing spectrum. J Abnorm Psychol 2007;116:645-66.
50. Krueger R, Tackett J. The externalizing spectrum of personality and psychopathology: an empirical and quantitative alternative to discrete disorder approaches. In: Beauchaine T, Hinshaw S (eds). Oxford handbook of externalizing disorders. Oxford: Oxford University Press, 2015:79-89.
51. Eaton NR, Krueger RF, Keyes KM et al. Borderline personality disorder co-morbidity: relationship to the internalizing-externalizing structure of common mental disorders. Psychol Med 2011; 41:1041-50.
52. Sharp C, Wright AG, Fowler JC et al. The structure of personality pathology: both general ('g') and specific ('s') factors? J Abnorm Psychol 2015; 124:387-98.
53. 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.
54. Lahey BB, Krueger RF, Rathouz PJ et al. Validity and utility of the general factor of psychopathology. World Psychiatry 2017;16:142-4.
55. Morgan GB, Hodge KJ, Wells KE et al. Are fit indices biased in favor of bi-factor models in cognitive ability research?: a comparison of fit in correlated factors, higher-order, and bi-factor models via Monte Carlo simulations. Intelligence 2015;3:2-20.
56. Eid M, Geiser C, Koch T et al. Anomalous results in G-factor models: explanations and alternatives. Psychol Methods 2017;22:541-62.
57. Murray AL, Johnson W. The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. Intelligence 2013;41:407-22.
58. Trull TJ, Widiger TA, Burr R. A structured interview for the assessment of the five-factor model of personality: facet-level relations to the axis II personality disorders. J Pers 2001;69:175-98.
59. Widiger TA, Presnall JR. Clinical application of the five-factor model. J Pers 2013;81:515-27.
60. Clark LA, Cuthbert B, Lewis-Fernández 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.
61. Waszczuk MA, Zimmerman M, Ruggero C et al. What do clinicians treat: diagnoses or symptoms? The incremental validity of a symptom-based, dimensional characterization of emotional disorders in predicting medication prescription patterns. Compr Psychiatry 2017;79:80-8.
62. Helzer J, Kraemer H, Krueger R. The feasibility and need for dimensional psychiatric diagnoses. Psychol Med 2006;36:1671-80.
63. Gastfriend DR, Mee-Lee D. The ASAM patient placement criteria: context, concepts and continuing development. J Addict Dis 2003;1:1-8.

DOI:10.1002/wps.20566

Quantitative classification as (re-)descriptive psychopathology

Consider these contrasts identified in Krueger et al's paper[1]: authoritative vs. empirical, *ex cathedra* (dogmatic) vs. evidence-based, and tradition vs. empiricism. It is powerful verbiage, suggesting that the members of the Hierarchical Taxonomy of Psychopathology (HiTOP) consortium are arguing for their generation's Copernican turn in distinction to the system that preceded it.

Freud famously described his achievement as an intellectual revolution[2], the neo-kraepelinians used revolutionary idioms against the psychoanalysts that preceded them[3], and now they are being used against the neo-kraepelinians.

Doubtlessly, readers will have a range of reactions to these contrasts. If the reaction "this is a coup by clinical researchers to replace the DSM and ICD with the factor analytic, dimensional models used in psychological testing" formed one end of a bipolar continuum, the other end would be "this is a heroic scientific revolution". I doubt many readers would assent to either pole wholeheartedly, but they may lean more toward one side or the other. I will argue that "coup" is too antagonistic an attribution and the proposed transition would be more appealing to psychiatrists if diplomatic alternatives to the "revolution" metaphor were used.

With respect to the coup, importing research traditions from scientific psychology into psychiatry not only has historical precedent; it has been historically important. To illustrate, consider E. Kraepelin, a groundbreaking architect of psychiatric classification, and R. Spitzer, who was the driving force behind the DSM-III and DSM-III-R.

Kraepelin's career plan was inspired by his contact with the founder of scientific psychology, W. Wundt. From his earliest days in the field, Kraepelin wanted to orient psychiatry away from speculative anatomical hypotheses and reductionism, and replace them with the experimental methods and concepts used in scientific psychology[4]. His descriptive psychopathology owed much to Wundt's strategy of decomposing complex psychological states into components that are more measurable.

Spitzer majored in psychology at Cornell University. In her biography of him, H. Decker[5] reports that Spitzer was trained as a psychoanalyst, but his interests lay in developing structured interviews and rating scales. He began his academic career at the New York State Department of Mental Hygiene in the Biometrics Research Unit, under the psychologist J. Zubin. The unit's purpose was to advance the quantitative study of psychopathology[6]. The psychological nature of Spitzer's early work is further documented by his collaborations with J. Endicott – a psychologist who had training in psychometrics.

In current terms, Kraepelin and Spitzer each had an interdisciplinary focus. With respect to classification, it has not been such a bad thing for psychiatry to occasionally take note of what the scientific psychologists are doing and rethink current practices – and it does not require a coup.

Turning to the revolution, many psychiatrists, including Spitzer[7], would assert that they are aware that psychiatric distress occurs with degrees of severity and that the distinction between normal and abnormal can be fuzzy. Indeed, one could argue that a manifest dimensionality is fundamental to descriptive psychopathology. Understanding it is a prerequisite for the competent use of a categorical classification system. If so, rather than a revolution, the HiTOP model is better seen as an attempt to translate common background knowledge of psychopathology into something more precise and substantive. One disadvantage of revolutionary talk is that it emphasizes the discontinuity between past and present, often drawing attention away from the many continuities[8].

Illuminated by the light of dimensionality, our understanding of psychopathology can be expanded in useful and interesting ways. Krueger et al's paper emphasizes an expansion in the scope of research questions asked. Here I would like to discuss another area of expansion. In doing so I will explain what is meant by my title "Quantitative classification as (re-)descriptive psychopathology".

I begin by giving an example of descriptive psychopathology: the depiction of panic disorder. After imipramine was introduced in the late 1950s, working at Hillside Hospital on Long Island, D. Klein and M. Fink began prescribing the drug to patients to learn about its mode of action[9]. In a historical retrospective based on interviews with Klein, F. Callard[10] recounts Klein and Fink's treatment of the man who would become the ur-patient for panic disorder.

The referring therapist believed that this patient had schizophrenia, but Klein disagreed, describing him as anxious, dependent and demanding. After four weeks of treatment with imipramine, neither the patient, his resident therapist, nor the supervising psychiatrist believed that the medication had made any difference. The ward staff did not concur, but they were not sure why. Eventually one nurse noticed that the patient no longer ran to the nurse's station several times a day asking for help because he feared he was dying.

For much of the 20th century, the symptoms of panic were a commonly manifested feature in the population of psychiatric phenotypes, but they were seen as parts of a coherent anxiety neurosis. Klein and Fink re-described these symptoms by putting a boundary around them, thus separating what they called episodic anxiety from anticipatory anxiety. With this re-description, even though panic had long been a background feature of the psychiatric landscape, it came into the foreground.

Descriptive psychopathology has been derided as a shallow emphasis on surface features. A successful re-description, however, is also a conceptual achievement of a synthetic nature in Kant's sense – it guides the way to the acquisition of information that is not contained in the description itself. For example, once Klein

and Fink saw panic as distinct from worry and avoidance, they learned that the primary problem in agoraphobia is not fear of open spaces, but fear of having another panic attack. Their discovery that the same patients also avoided crowded theaters would have been a puzzling feature of agoraphobia, but not of panic disorder.

The Research Domain Criteria (RDoC) initiative, with its focus on causality, might represent the abandonment of descriptive psychopathology, but it is equally consistent with RDoC's anti-reductionist aspiration that mechanisms will be maps for locating new descriptions in the psychiatric landscape. The same is true for HiTOP. Proposing a meta-structure for how things fit together affords some options for recognizing new patterns. HiTOP has an immediate advantage over RDoC because it does not have to translate biological findings into psychological descriptions; it is already psychological.

Using a taxonomy, however, is only a part of understanding psychopathology, including descriptive psychopathology. It is unrealistic, therefore, to expect that statistical correlations can do all the descriptive work. With respect to panic disorder, Klein claimed that the ward nurse who reported that the ur-patient no longer ran to the nurse's station was a good observer. This was their first clue to describing what they called a psychiatric reaction pattern. It was followed by prolonged observations of what the patient did and said, how he reacted to others, and how others reacted to him.

Hopefully, good observers will notice some of the clues that a comprehensive dimensional hierarchy presents, recognize patterns, and subject them to validation studies. Concepts like borderline and narcissistic personality disorder are so entrenched that they assert themselves when certain features are present. HiTOP offers a way to take a second look. Ideally, clinicians and scientists could learn to see anew something that has been there before them all along – and let it guide them to other things that they did not recognize before.

Peter Zachar
Auburn University Montgomery, Montgomery, AL, USA

The author would like to thank S. Lilienfeld for thoughtful comments on an earlier draft of this commentary.

1. Krueger RF, Kotov R, Watson D et al. World Psychiatry 2018;17:282-93.
2. Freud S. A general introduction to psychoanalysis. New York: Liveright, 1920.
3. Wender PH, Klein DF. Mind, mood, and medicine. New York: Farrar Straus Giroux, 1981.
4. Engstrom EJ, Kendler KS. Am J Psychiatry 2015; 172:1190-6.
5. Decker HS. The making of DSM-III: a diagnostic manual's conquest of American psychiatry. New York: Oxford University Press, 2013.
6. Zubin J, Salzinger K, Fleiss JL et al. Annu Rev Psychol 1975;26:621-71.
7. Spitzer RL. Schizophr Bull 1982;8:592.
8. Zachar P. In: Kendler KS, Parnas J (eds). Philosophical issues in psychiatry II: Nosology – definition of illness, history, validity, and prospects. New York: Oxford University Press, 2012:21-34.
9. Klein DF, Fink M. Am J Psychiatry 1962;119: 432-8.
10. Callard F. Osiris 2016;31:203-26.

DOI:10.1002/wps.20558

Dimensions fit the data, but can clinicians fit the dimensions?

Krueger et al's paper[1] is impressive and erudite. One might say it is too erudite, because the average clinician will find it difficult to anchor his or her clinical practice to the attachments offered. But the arguments put forward are scientifically incontrovertible; the data for almost, if not all, psychiatric disorders indicate that their dimensional description is nearer to truth than a categorical one.

The key section in this paper to most readers in practice is "clinical translation", and here the work group is going to have to work extra hard. To what extent can the dimensional system be adapted, transformed, or forced, depending on your starting point, into clinical decision-making?

There is an interesting historical parallel here. In the UK, in the late 1950s and early 1960s, there was what is commonly called the Platt-Pickering debate, played out in the columns of *The Lancet*. This pitted the cerebral (dimensional) champion, G. Pickering, in one corner, against the clinical (categorical) pugilist, R. Platt, in the other. Although there were no apparent knockout blows, the debate was a riveting spectacle, illustrated by rapier-like thrusts and counter-punches by two austere but slightly irritable protagonists, always polite but each showing incredulity at the apparent stupidity of the other.

Their debate was over the classification of high blood pressure. Was it best regarded as a continuous variable[2] or better described as two categorical populations, a larger one with normal blood pressure, and a smaller one with hypertension[3]? Pickering made the case that blood pressure is a continuously distributed characteristic with no clean separation between abnormal and normal. Platt insisted that those with very high blood pressure were a discrete group who represented the disease, hypertension, and that this fact could potentially be explained by genetic characteristics; he proposed a Mendelian dominant gene. This genetic theory was not supported and the Pickering power-house swept away the old arguments: "The new view, for which we and our colleagues are largely responsible, is that essential hypertension represents a quantitative and not a qualitative deviation from the norm"[4].

This resonates strongly with the current debate about dimensions in psychiatry. In the Platt-Pickering debate, the clinicians – and, dare one say, *The Lancet* itself[5] – were on Platt's side. After all, if he was right, it would make their job so much easier. Clean categorical diagno-

sis is always better than a dimensional fudge. What the Hierarchical Taxonomy of Psychopathology (HiTOP) investigators need to do is to show the clinician that there is genuine clinical value in the dimensional approach; that it is not a fudge. We have some clues. Thus, in the case of personality disorder, shortly to be a dimensional diagnosis in ICD-11 and regarded as a genuine paradigm shift[6], it is important to know that the more severe the disorder the greater its persistence and its impact on long-term social functioning[7].

But this only describes prognosis. Can a dimensional diagnosis help treatment? Again we have some encouraging findings. Sub-clinical depression is not a formal diagnosis but it causes a lot of suffering. It is easily accommodated on a dimensional continuum and could be a suitable condition for treatment, and in a recent meta-analysis there is some evidence, not yet strong, that psychological treatments are effective here[8]. Would this apply to drug treatment too? Probably not, and, for this to be appropriate, a higher point on the dimension would probably have to be chosen[9].

Clinicians are naturally conservative when it comes to diagnosis and classification, and change is always resisted at first. But if it can be shown that there is definite advantage in a dimensional approach, that it can lead to better and more fine-tuning of management, then it may win approval. It would probably be necessary to have parallel systems at first to allow comparisons to be made between categorical and dimensional approaches.

But there will be continuing concerns in clinical practice if there are not clear indications for decision-making offered by the diagnostic system. Krueger and his colleagues rightly note that the recent elegant Research Domain Criteria proposals, whatever their value in identifying neurobiological constructs, do not help such decision-making. Although the HiTOP team may go further and succeed in their aim of "connecting structure and process" in explaining psychopathology, the clinician at the coal-face can only look on with bemusement at any system which, however well grounded in empirical science, still does not provide answers to key questions. When is apparent pathology within the range of normality? At what stage in a dimensional system of a major diagnosis is coercion justified in management? When is it right to regard co-occurrence of disorders as comorbidity or instead as part of the same spectrum (e.g., anxiety-depression)?

These are not academic talking points. Therapeutic advancement often happens by serendipity, but we also need to have a classification system that helps empirical science to focus on specific aspects of efficacy. So, instead of psychiatry using the current pot-pourri of general interventions into heterogeneous populations giving equivocal results, we could look forward to "focused diagnosis-specific gain". The possible value of quinine in malaria was discovered by chance but, because malaria was a clearly identified disease, it was possible, even in the mid-1860s, to show that all the cinchona alkaloids – quinine, quinidine, cinchonine and cinchonidine – were equally effective in treating the disease. Remember, at this time in history, malaria was identified by the same procedures that we use in psychiatry today.

The HiTOP investigators may feel it is far outside their remit to enter the therapeutic and other intervention arenas, but they need to be aware of their importance. The oldest and most successful classification in psychiatry has been the dimensional one of intellectual disability based on IQ. Although this has been rightly modified in several ways to take account of adaptive functioning, for more than a century this classification has allowed appropriate placement, support and management to take place for people in each of the dimensional groups.

What about the long-term outcome of the Platt-Pickering debate, which Pickering was generally assumed to have won? Currently the most common diagnosis in cardiology in the ICD-10 classification is essential hypertension, so the Platt supporters may now claim some sort of victory. So, in 60 years hence, will it be seen that dimensions have triumphed or will psychiatric classification be essentially the same as now? If Krueger and his colleagues can come forward with more clinical meat to add to their helping of science, things will certainly change.

Peter Tyrer
Centre for Psychiatry, Imperial College, London, UK

1. Krueger R, Kotov R, Watson D et al. World Psychiatry 2018;17:282-93.
2. Pickering G. Lancet 1960;275:170.
3. Platt R. Lancet 1959;274:1092.
4. Oldham PD, Pickering G, Fraser Roberts JA et al. Lancet 1960;275:1085-93.
5. Editorial. Lancet 1959;274:895-6.
6. Widiger TA, Oltmanns JR. Personal Ment Health 2016;10:120-2.
7. Tyrer P, Tyrer H, Yang M et al. Personal Ment Health 2016;10:76-83.
8. Cuijpers P, Koole SL, van Dijke A et al. Br J Psychiatry 2014;205:268-74.
9. Maj M. Br J Psychiatry 2011;199:85-6.

DOI:10.1002/wps.20559

HiTOP must meet the use requirements of the ICD before it can aspire to replace it

As described by Krueger et al[1], the approach being taken by the Hierarchical Taxonomy of Psychopathology (HiTOP) consortium in attempting to elucidate the underlying dimensions of psychopathology is an important one. I agree particularly about the immediate importance of identifying connections between overt expressions of psychopathology

and neural mechanisms and genomic variance, and believe that HiTOP has an important contribution to make in this regard.

At the same time, I do not believe that HiTOP can be successful as a sole approach. As with the Research Domain Criteria (RDoC) project promoted by the US National Institute of Mental Health (NIMH), it seems important not to oversell HiTOP or to pretend that it describes a classification system *per se* that will be capable of replacing the ICD or the DSM at any point in the immediate future. Although the NIMH has walked back its initial rhetoric[2] to clarify that RDoC is actually a framework for research[3], Krueger et al's paper makes the same mistake with HiTOP.

The paper is also marred by tendentious repetition of the claim that the ICD and the DSM are "consensus-based", "authoritative", "political" classifications, in contrast to HiTOP, which is "empirical" and "scientific". Such characterizations, although perhaps rhetorically useful in promoting a new approach, are actually inaccurate, as with the widely repeated and false characterization of DSM-I and DSM-II as psychoanalytic[4], or the initial messaging about RDoC that characterized the DSM explicitly and the ICD by implication as responsible for the lack of dramatic breakthroughs in understanding the etiology of mental disorders and providing curative treatments[2]. This paper's similar denigration of "authoritative" as opposed to "empirical" classification systems appears to be based, thinly, on the facts that: a) the ICD-11 and DSM-5 (and RDoC) are institutionally sponsored; b) expert working groups developed the initial proposals for changes to the previous versions of the classifications; and c) there was an institutional demand for some degree of continuity across versions.

With regard to the first point, the development and maintenance of international classifications for health and the standardization of diagnostic procedures are core constitutional functions assigned to the World Health Organization (WHO) through international treaty by 194 member states. It is unclear why being a "consortium" without a clear formal authority structure or a responsible institution would make HiTOP inherently superior in relation to these tasks. With regard to the second point, an explicit charge of working groups for both the ICD-11 and the DSM-5 was to perform a rather rigorous analysis of the state of the current evidence. Krueger et al are correct, though, that the range of possibilities for transforming the classifications was to some extent limited by the adoption of *a priori* elements of the existing structure, such as the existence of separate groupings of mood disorders and anxiety disorders.

Most of the results presented in the paper in support of HiTOP's hierarchical dimensional models are based on a set of inter-related techniques including taxometric analysis, latent class analysis, cluster analysis, and factor analysis. While these can be powerful and sophisticated statistical tools, they do not serve up the truth like Venus on a clamshell. They still require interpretation by human experts. The fact that HiTOP's authority structure and the specific criteria for evaluation are not transparent or explicit (at least based on this paper) does not mean that the evidence is not being synthesized and interpreted based on expert judgments.

For the WHO, a demand for explicit continuity between the ICD-10 and the ICD-11, at a minimum in the form of clear cross-walking, is based on one of the ICD's main purposes – to provide a framework for the collection and reporting of health statistics – as well as on the need for longitudinal global, national and local health information. The governments of WHO member states have increasingly integrated the ICD into clinical processes and policies related to health care coverage and reimbursement, social services, and disability benefits[5], and are also concerned about the continuity of health data and the continuous application of laws and policies. However, the paper suffers from a lack of familiarity with the functioning of the WHO and the purposes of the ICD-11. Even though Krueger et al include the ICD-11 in the sweep of their characterizations, all of the specific information in the paper about "traditional", "authoritative" classifications is taken from the DSM-5. This perhaps reflects the fact that only ten of the paper's 45 authors are from outside the US and none is from a developing country.

The WHO does not, in fact, "claim, through tradition and putative authority, that psychopathologies are organized into discrete diagnostic entities". We have recently written explicitly and at great length about the better correspondence of dimensional approaches to the observed data[3]. The categorical nature of the ICD system is necessary for its application in global health statistics and in many instances for its use in clinical settings (e.g., eligibility, treatment selection). In most countries, provision of medical care other than routine examinations and preventive services is contingent on a qualifying diagnosis. Other relevant decisions are typically categorical (yes/no); even if the information that underlies them is dimensional, a threshold must be imposed. Inclusion of mental disorders in the ICD facilitates coordination with the classification of other disorders, as well as the search for related mechanisms of etiology, pathophysiology and comorbidity of disease processes. It also provides a basis for parity of mental disorders with other types of health conditions[5]. Mental disorders in the ICD-11 must follow the same set of structural and taxonomic rules as the rest of the classification.

Within the constraints of a categorical system, the ICD-11 has gone to considerable lengths to integrate dimensional constructs into the classification of mental disorders, which has been made possible by specific structural innovations as compared to the ICD-10. One example that is discussed in the paper is the incorporation of a dimensional classification of personality disorders[6,7]. Similarly, the ICD-10 subtypes of schizophrenia (e.g., paranoid, hebephrenic, catatonic) have been replaced by a set of symptom ratings (e.g., positive symptoms, negative symptoms, cognitive symptoms) that may be applied to all primary psychotic disorders[8]. A category for anxious depression based on two correlated but distinct dimensions has been incorporated into the

version of the ICD-11 classification of mental disorders for primary care settings[9]. These innovations will push the ICD-11 in the direction envisioned by HiTOP, but it is possible that they may be experienced as more complex than the purely categorical approach they are replacing, which may stimulate resistance among clinicians and health systems.

While the WHO does appear to be facing this challenge head-on within the structural and taxonomic constraints of the ICD, there is a considerable amount that HiTOP might take on board in order to facilitate further transformations of this nature. Assuming that the correct dimensions have been identified, much work is necessary to translate group-level research results into measures and cutoffs that are predictive at the individual level[3]. Although Krueger et al claim "greater relative utility and empirical accuracy of continuous conceptualizations of psychopathology", very little work has been conducted aimed at developing tools that can be demonstrated as robustly valid as a basis for making individual health care decisions.

Any dimensional system that would seek to replace "authoritative" classifications would need to demonstrate that it is fit for purpose across the range of functions for which the world uses the ICD.

Geoffrey M. Reed
Department of Mental Health and Substance Abuse, World Health Organization, Geneva, Switzerland

The views expressed in this commentary are those of the author and do not necessarily represent the official policies or positions of the WHO.

1. Krueger RF, Kotov R, Watson D et al. World Psychiatry 2018;17:282-93.
2. Insel T. Transforming diagnosis. https://www.nimh.nih.gov.
3. Clark LA, Cuthbert B, Lewis-Fernández R et al. Psychol Sci Public Interest 2017;18:72-145.
4. Cooper R, Blashfield RK. Psychol Med 2016;46:449-56.
5. International Advisory Group for the Revision of ICD-10 Mental and Behavioural Disorders. World Psychiatry 2011;10:86-92.
6. Tyrer P, Reed GM, Crawford MJ. Lancet 2015;385:717-26.
7. Hopwood CJ, Kotov R, Krueger RF et al. Personal Ment Health 2018;12:82-6.
8. Gaebel W. Schizophr Bull 2012;38:895-8.
9. Goldberg DP, Reed GM, Robles R et al. J Affect Disord 2017;213:199-206.

DOI:10.1002/wps.20560

"Throwing out the baby with the bathwater"? Conceptual and methodological limitations of the HiTOP approach

More sophisticated explorations of the higher-order dimensional and hierarchical structure of psychopathology have become an exciting complementary way towards developing an improved classification of mental disorders and reducing artefactual comorbidity.

The impressive work of the Hierarchical Taxonomy of Psychopathology (HiTOP) Consortium with their mission paper[1] provides evidence for considerable advances as compared to previous suggestions, and underscores the potential of such approaches not only for improved future classificatory models with increased utility for research and practice, but also for the development of improved psychometric assessment instruments for psychopathology. However, as impressive such an approach might appear at first sight, there is a need of pointing out several limitations that caution against the use of this model.

On the *conceptual level*, we emphasize first of all that comorbidity is not "a problem", but a clinical characteristic of patients meaningful for treatment and management[2]. The belief that people suffer from only one underlying condition is implausible and misleading. The value of the HiTOP Consortium approach might be in reducing a certain degree of what has been called "artefactual" comorbidity, due to overlapping criteria in our current classification systems.

Second, the suggested hierarchical structural model has a serious limitation: it is based almost exclusively on traditional assessment instruments (dimensional scales, interviews) from cross-sectional studies. Leaving aside the vast array of inherent general psychometric problems, we highlight that such scales merely reflect a subjective-verbal "snapshot" picture of the level of symptom-distress that a person reports at the time of investigation. As essential such a snapshot might be for a first "impressionistic" step of a syndromal diagnosis, it certainly does not allow to decide on a diagnosis relevant for treatment without taking into account the patient's history (e.g., depressive syndromes cannot be equated with diagnoses of major depression or even of any affective or any mental disorder).

Third, the HiTOP approach does not grab appropriately the nature of mental disorders as dysfunctions – up to now insufficiently understood – of basic psychological processes as well as associated "perturbations" in brain functions at the cell and systems level[3]. The former are centrally involved in the behavioral, cognitive-affective and somatic symptom processes currently used to define mental disorders. The latter "perturbations" can be best described as various types of fluctuating dysfunctions in complex structural and functional neural circuits involved in information processing and emotion regulation.

The identification of common causal pathways is of core relevance for an improved diagnostic system. They allow identifying the factors and mechanisms responsible for the onset, progression and maintenance of mental disorders. Proposed models based on such mechanisms provide guidance for improved research strategies and the derivation of improved interventions, targeted to interrupt the causal pathways[3].

Promising examples come from psychosis research. In a clinical staging framework, the at-risk or symptomatic state of a patient can be evaluated to derive tailored interventions spanning from

primary selective prevention in asymptomatic subgroups (stage 0) and high-risk subjects (stage 1), over early treatment in first episode (stage 2) or relapsing psychotic patients (stage 3), to maintenance treatment in unremitting patients (stage 4)[4].

Such frameworks also exist for other facets of psychopathology such as anxiety, depression or substance use, providing specific guidance on early targeted interventions. The "symptom progression - comorbidity development" model[3,5] emphasizes the early signs and symptoms of fear and anxiety in the development of psychopathology and a staging based on "comorbid" escalations from circumscribed manifestations in childhood to more complex diagnostic constellations (multiple anxiety disorders, comorbid depression and substance disorders) later in adolescence or adulthood. Besides a range of vulnerability factors at various levels and in different developmental periods, the initial psychopathology itself entails a causal cascade (e.g., increasing demoralization and inactivity due to avoidance promoting depression)[6]. This model has direct implications for therapeutic and preventive interventions.

Therefore, the first caveat of higher-order taxonomies such as the one suggested by the HiTOP Consortium is that they are at best a complementary piece of descriptive evidence that might prove useful in reducing artefactual comorbidity. But they do not reflect the true dynamic developmental nature of mental disorders and might even be an obstacle for developing improved targeted causal interventions.

Regarding *methodological constraints*, we do not refer here to the numerous mathematical and statistical limitations of the higher-order dimensional and hierarchical approaches that call for caution[7,8]. Beyond these, the strongest evidence against such models comes from prospective-longitudinal investigations, revealing the instability of the assumed higher-order structure and spectra over time[7]. Along the developmental axis, the structure of higher-order dimensions changes significantly, both within factors and across spectra. The assumption that this instability might be due to a limited reliability of assessments is implausible and would actually also argue against such higher-order models in general.

Furthermore, the statement that dimensional measures are advantageous over categorical data is trivial. They simply provide more information and are thus preferable in any approach[9]. Assuming that hierarchical structural models based on dimensional data may lead *per se* to an improved classification of mental disorders and "solve the problem of comorbidity" is like "throwing out the baby with the bathwater" and obscures important issues, given the underlying assumptions and the lack of developmental considerations. This does not invalidate the additional utility and the potential of such approaches, but suggests that these models are at best complementary to other principles and sources of evidence.

Undoubtedly, as compared to previous simpler models, the HiTOP model has increased in breadth and specificity (e.g., spectra for thought disorder and detachment). However, the extensions also cause new inconsistencies, such as enhancing the "distance" between internalizing and externalizing dimensions, although externalizing disorders might involve preceding internalizing pathways (and vice versa). Moreover, as attractive and impressive the visual depiction of a new taxonomy of psychopathology may be, using new words for old ones might increase the risk that already established research findings lack consideration in the future.

Further, "somatoform" diagnoses (dismissed in DSM-5) are reintroduced without explaining the rationale. This particular cluster also serves as an example for the difficulty – even cross-sectionally – to find a coherent general structure of psychopathology. Somatoform syndromes are differentially (i.e., by gender and age group) associated with a broad range of conditions which are spread out in the HiTOP model (anxiety, psychosis, hypomania, post-traumatic stress disorder, and many other diagnoses not mentioned in the framework)[7], which complicates the implementation of the model.

To conclude, higher-order dimensional and hierarchical models of psychopathology such as the ambitious HiTOP model are at best a complementary way towards developing an improved classification of mental disorders for research and practice. Their potential value lies in reducing artefactual comorbidity and deriving improved cross-sectional psychometric assessment instruments.

However, HiTOP provides little specific guidance towards our ultimate goal, namely, a classification of mental disorders based on causal factors and mechanisms involved in the first development of psychopathology and its progression over time. Its inherent weakness remains the overemphasis on cross-sectional psychopathology and the neglect of dynamic developmental pathways and differential diagnostic issues relevant to treatment and management.

Hans-Ulrich Wittchen[1,2], Katja Beesdo-Baum[1,3]
[1]Institute of Clinical Psychology and Psychotherapy, Technische Universität Dresden, Dresden, Germany; [2]Department of Psychiatry and Psychotherapy, Ludwig-Maximilians-Universität München, München, Germany; [3]Behavioral Epidemiology, Technische Universität Dresden, Dresden, Germany

1. Krueger RF, Kotov R, Watson D et al. World Psychiatry 2018;17:282-93.
2. Fava GA, Tossani E, Bech P et al. Int J Methods Psychiatr Res 2014;23:92-101.
3. Wittchen H-U, Knappe S, Andersson G et al. Int J Methods Psychiatr Res 2014;23:28-40.
4. Fusar-Poli P, McGorry PD, Kane JM. World Psychiatry 2017;16:251-65.
5. Shear MK, Bjelland I, Beesdo K et al. Int J Methods Psychiatr Res 2007;16(Suppl. 1):S52-64.
6. Beesdo K, Bittner A, Pine DS et al. Arch Gen Psychiatry 2007;64:903-12.
7. Wittchen H-U, Beesdo-Baum K, Gloster A et al. Int J Methods Psychiatr Res 2009;18:189-203.
8. Wittchen H-U, Beesdo K, Gloster AT. Psychol Med 2009;39:2083-9.
9. Wittchen H-U, Höfler M, Gloster AT et al. In: Regier DA, Narrow WE, Kuhl EA et al. (eds). The conceptual evolution of DSM-5. Washington: American Psychiatric Publishing, 2010:119-43.

DOI:10.1002/wps.20561

The dialectic of quantity and quality in psychopathology

Krueger et al[1] provide a novel and challenging perspective on the perennial divide between the categorical and dimensional approaches to the conceptualization and classification of psychopathological phenomena.

Writing on behalf of the recently established Hierarchical Taxonomy of Psychopathology (HiTOP) Consortium, they address critically the "official nosology", especially as exemplified in the DSM-5. The latter manual is criticized for being "authoritative", guided by psychiatrists, and not immune against "socio-political" considerations in preserving and presenting an *ex cathedra* view of psychopathology as consisting of discrete nomothetic entities or taxa. In contrast, the authors highlight the likely empirical advantages of adopting the alternative position on psychopathology as a continuum of quantitative variation that can be organized hierarchically into several higher-order spectra and dimensions.

Krueger et al claim that recent research, methodologically stronger than its predecessors, overwhelmingly supports the quantitative-dimensional model of psychopathology, and believe that the latter is now fit to be ubiquitously translated into clinical practice. They advocate placing this model of classifying psychopathology on "an empirical playing field" instead of perpetuating the "traditional" nosology, exemplified by the DSM revision process.

Much of the evidence in support of these proposals stems from the comprehensive quantitative review of published taxometric research by Haslam et al[2]. This review was based on a detailed examination and secondary analysis of 177 articles which, when combined, reported a total of 39.9% taxonic results. However, the authors concluded that, after statistically controlling for confounds, the "true" prevalence of taxonic findings was only 14%, mostly involving the domains of schizotypy, autism and substance use disorders. They contended that historical improvements in the methodological quality of taxometric studies, especially the use of simulated comparison data and the linked comparison curve fit index, have contributed to the marked decline of taxonic findings.

There are two possible caveats to this line of reasoning. First, the purely statistical analysis and interpretation of the data is no substitute for a well-designed, real-life comparative study of clinical populations assessed according to both the hierarchical dimensional model illustrated by Krueger et al and the "traditional" categorical nosology of ICD-10 or DSM-5. The outcome criteria in such a hypothetical study should include choice of treatment, prognosis and functional status of the participants. As far as I am aware, no such study has yet been designed or conducted.

My second caveat concerns the applicability of the quantitative dimensional scheme to the bulk of psychotic disorders (marginally mentioned in Krueger et al's paper). Historically, the evolution of the classification of these disorders has taken a path in reverse to that of the common non-psychotic disorders. The theory of the "unitary psychosis" has been dominant in European psychiatry around the middle of the 19th century, being associated with the names of its first proponent A. Zeller and its first critics W. Griesinger and K. Kahlbaum. It postulated a continuum of different stages within a unitary morbid process, terminating ultimately in a complete disintegration of mental life. It was against this background that E. Kraepelin synthesized the three pre-existing entities of hebephrenia, catatonia and paranoid dementia into a single concept, and proposed in 1896 the dichotomy of the unitary spectrum into the discrete entities of dementia praecox and manic-depressive insanity. Renamed as schizophrenia by E. Bleuler in 1908, the former entity was further described as "the group of schizophrenias", to be split further by K. Leonhard into systematic and unsystematic forms, each containing many discrete subtypes[3]. Notably, there has been a recent revival of the continuum model of psychotic disorders[4], which in its turn has been criticized as "scientifically unproven and clinically impractical"[5].

At this point, I shall add my own take on the problem: can a classification of mental disorders be biologically anchored? This is doubtful, at least in the foreseeable future, because: a) the objects classified in psychiatry are explanatory concepts, i.e. abstract entities rather than physical organisms; b) the taxonomic units of "disorders" in DSM-IV, DSM-5 and ICD-10 do not form hierarchies; c) the current psychiatric classifications contain no supraordinate, higher-level organizing concepts. Leaving aside the vexing issue of validity of the categories, the criteria for evaluating psychiatric classifications should at present focus pragmatically on their clinical relevance and utility[6]: capacity of discriminating between syndromes and between degrees of their expression in individual patients; adaptability to different populations and cultural environments; reliability; cognitive ease of use; and reducing stigma. My prediction is that the quantitative/dimensional and the taxonic/discrete approaches to the classification of mental disorders will remain dialectically interconnected as the "yin" and "yang".

A methodological tool eminently suited for empirical research is the grade of membership (GoM) latent structure analysis[7], which enables the aggregation of clinical and/or neurocognitive measures into a parsimonious number of "pure types" (taxons) which represent fuzzy sets, rather than discrete categories, and assigns to each individual a quantitative affinity score indexing the degree to which he/she resembles each one of the taxons. My research group has been using the GoM to split a large cohort of schizophrenia patients into subtypes based on neurocognitive measures and to specify each patient's affinity to any one of the taxons[8].

I am reminded of the Hegelian postulate[9] about the transition ("phase shift") of the accumulation of quantitative changes into a new quality. This sums

up my impression of the stimulating argumentation presented in the paper by Krueger et al.

Assen Jablensky
Division of Psychiatry, University of Western Australia School of Medicine, Perth, WA, Australia

1. Krueger RF, Kotov R, Watson D et al. World Psychiatry 2018;17:282-93.
2. Haslam N, Holland E, Kuppens P. Psychol Med 2012;42:903-20.
3. Leonhard K. Classification of endogenous psychoses and their differentiated etiology. Wien: Springer, 1999.
4. van Os J, Linscott RJ, Myin-Germeys I et al. Psychol Med 2009;39:179-95.
5. Lawrie SM, Hall J, McIntosh AM et al. Br J Psychiatry 2010;197:423-5.
6. Kendell R, Jablensky A. Am J Psychiatry 2003; 160:4-12.
7. Woodbury MA, Manton KC. Soc Methods Res 1978;18:126-63.
8. Hallmayer J, Kalaydjieva L, Badcock J et al. Am J Hum Genet 2005;77:468-76.
9. Hegel GWF. Phenomenology of spirit. Oxford: Oxford University Press, 1977.

DOI:10.1002/wps.20562

After the failure of DSM: clinical research on psychiatric diagnosis

Clinical research on psychiatric diagnosis has failed from 1980 until now. In the DSM-III onwards era, clinical nosology research has been irrelevant. Contrary to the claims made in 1980 with DSM-III, diagnostic reliability did not lead to diagnostic validity, because reliability became an end in itself. The psychiatric profession congratulated itself on agreeing about how to define psychiatric diagnoses, and refused to make any further changes. The process was reified in DSM-III and DSM-IV, such that major changes were infrequent, and when they did occur, they were based on winds of opinion rather than solid, replicated scientific research. Minor changes were fought with passion, despite reasonable scientific data in their support[1].

In short, the greatest obstacle to scientific progress is, and has been, the DSM system of diagnosis. In 1980, DSM-III promised to push psychiatry forward, defining clear criteria for improvement with research. Now, DSM-5 is based on unscientific definitions which the profession's leadership refuses to change based on scientific research.

This perspective can be seen as heretical, as it is still not accepted by the mainstream of the American Psychiatric Association (APA). Yet, not all American psychiatry agrees with the APA. Importantly, the US National Institute of Mental Health (NIMH) leadership strongly criticized DSM-5 upon its publication, and announced it would no longer fund research using DSM criteria. Instead, the NIMH leadership proposed an alternative approach for research: the Research Domain Criteria (RDoC). The main problem with the latter approach is that it gives up on clinical research about diagnosis altogether, claiming that research should begin with brain-based concepts. Both extremes are questionable: the DSM approach is clinical but unscientific; the NIMH approach is scientific but not clinical. The profession still awaits a scientific approach to clinical research on diagnosis.

Krueger et al's paper[2] reflects a positive response to this unfortunate state of affairs. The key leaders of this consortium were involved with the unhappy personality traits vs. disorders controversy in DSM-5[3]. They are researchers who advocated for following scientific data towards a change in personality nosology in favor of traits. They failed. Now they propose a consortium to conduct and promote an empirically-based nosology in psychiatry. This project is long overdue.

Our current dilemma was predictable. We can learn from early critics of DSM, like H. van Praag. In 1993, while the DSM-IV process was in full swing, he wrote[4]: "Today's classification of the major psychiatric disorders is as confusing as it used to be some 30 years ago. All things considered, the present situation is worse. Then, the psychiatrists were at least aware that diagnostic chaos reigned and many of them had not high opinion of diagnosis, anyhow. Now the chaos is codified and thus much more hidden... There is nothing wrong in basing the first draft of an operationalized taxonomy on expert opinion... One should abstain, however, from proceeding further on that route. Yet, this is exactly what happened... I strongly feel that 1) an immediate moratorium should be laid on any further expert-opinion-based alterations in [diagnosis]... and that 2) future changes should be based on research only".

An important feature of the DSM ideology is the rejection of the concept of a hierarchy of diagnosis, on the debatable ground that we cannot have hierarchies in the absence of etiology. If we do not know causes of diseases, we cannot say which ones should be diagnosed preferentially to others. This perspective ignores the importance of differentiating diseases with many symptoms from those with fewer. If a symptom occurs as one of twenty in one illness, and one of two in another, then the first should be ruled out before the second is diagnosed. It is not biologically sound to diagnose "comorbid" panic disorder every time someone has a panic attack in the setting of a depressive or manic episode. The panic symptoms are often caused by mood states, rather than being a separate independent disease. We already take this approach with delusions and hallucinations; if they occur in mood states, we do not diagnose schizophrenia. This is an exception in the DSM system, though, which refuses to use the same logic for other psychopathological states.

Hence two problems result, again as van Praag described decades ago: "nosologomania"[5] (i.e., the creation of many scientifically invalid diagnostic definitions) and many false "comorbidities"[6].

In fact, the concept of "comorbidity" was introduced by Feinstein in 1970 as meaning the simultaneous co-occurrence of two independent, unrelated diseases[7]. The co-occurrence of anxiety and depression does not quality for comorbidity; either they are symptoms of the same condition (like neurotic depression), or they reflect one condition causing another (as in mixed depression, where anxiety is caused by the mixed state).

The hierarchy proposed by this consortium grows out of the personality literature. It includes concepts that may be relevant to personality, but which are less relevant to mood or psychotic diseases. Dimensionality is relevant in both cases, but perhaps in different ways. For instance, the best clinical research supports the dichotomy between schizophrenia and manic-depressive illness. Further, the externalizing/internalizing concepts do not capture many of the features of manic-depressive illness, such as the presence of mixed states. The placement of "mania" as part of an "internalizing" disorder is questionable. The distinction between bipolar illness and "unipolar" depression is assumed in the hierarchical taxonomy, whereas this distinction has questionable validity based on the best available clinical research.

Thus, the proposal of a quantitative hierarchy is welcome, but how it is set up will require more attention to some clinical research that does not appear to have been included in the working taxonomy provided in Krueger et al's paper.

An alternative approach growing out of research on mood and psychotic diseases has been proposed dating back to the 1970s[8]. I have suggested a modernized version of that approach[9]. In this proposal, the hierarchy of psychopathology would involve manic states (bipolar illness) at the top of the pyramid of diagnosis, followed by depressive states (unipolar depression), followed by schizophrenia, then anxiety diagnosis (like obsessive-compulsive disease), then personality "disorders" (such as borderline and antisocial), then attention deficit disorder and narrowly defined diagnoses (such as eating disorders or paraphilias). The general concept is that conditions higher on the hierarchy are polysymptomatic, and cause the symptoms of conditions lower on the hierarchy, and thus the former should be ruled out before the latter are diagnosed.

This is standard medical teaching. Core medical training involves using symptoms to identify diagnoses, and not just converting symptoms into diagnoses, as is the case with DSM-III onwards. Then those diagnoses are organized in a differential diagnosis, where higher order ones are ruled out before lower order ones are made. The opposite approach is taken with the DSM system, which is powerful evidence for an important observation: contrary to what many of the post-modernist and anti-biological critics of DSM claim, the DSM system is not at all representative of the "medical model". In fact, it is quite anti-medical, as shown in its rejection of the hierarchy concept.

In sum, Krueger et al's effort is very worthwhile, but essentially limited to concepts in the personality literature. If expanded to capture affective and psychotic conditions, it could begin to put the profession on the road to a better clinical nosology for the future, leaving DSM in the rearview mirror.

S. Nassir Ghaemi
[1]Tufts University and Harvard Medical School, Boston, MA, USA; [2]Novartis Institutes for Biomedical Research, Cambridge, MA, USA

1. Ghaemi SN. World Psychiatry 2013;12:210-2.
2. Krueger RF, Kotov R, Watson D et al. World Psychiatry 2018;17:282-93.
3. Zachar P, Krueger RF, Kendler KS. Psychol Med 2016;46:1-10.
4. van Praag HM. Make-believes in psychiatry or the perils of progress. New York: Brunner Mazel, 1992.
5. van Praag HM. World J Biol Psychiatry 2000;1: 151-8.
6. van Praag HM. Br J Psychiatry 1996;168(Suppl. 30):129-34.
7. Feinstein AR. J Chronic Dis 1970;23:455-68.
8. Surtees PG, Kendell RE. Br J Psychiatry 1979; 135:438-43.
9. Ghaemi SN. Mood disorders: a practical guide, 2nd ed. Philadelphia: Lippincott Williams & Wilkins, 2007.

DOI:10.1002/wps.20563

Internalizing disorders: the whole is greater than the sum of the parts

The Hierarchical Taxonomy of Psychopathology (HiTOP) consortium is a group of investigators working to advance the empirical classification of psychopathology. In a previous issue of this journal they published a concise account of the work of their consortium[1], and now they put forward a statement of intent and a summary of progress[2].

Practitioners in the mental health field act as though each mental disorder is a discrete category – Mrs. Smith has panic disorder; Mr. Brown has major depressive disorder – and consider that treatment and future developments will naturally follow from the diagnosis. At one level this is appropriate and necessary for the orderly management of treatment for individual patients, but at a higher level this is not correct: the defining symptoms of each mental disorder exist on dimensions that extend from very mild and incomplete sets consistent with wellness to the very severe, complete sets that disable and distress and are incompatible with being well.

The classifications of mental disorders – DSM-5 and ICD-10 – are, at the simplest level, definitions of the threshold at which a set of symptoms becomes sufficiently complete, disabling or distressing to be of clinical concern, and an indicator of the need for treatment. The point on a dimension of increasing severity where a diagnosis is warranted is not indicated by any external measure

such as a sudden change in pathophysiology or of distress or disability. The threshold for a diagnosis in each classification is made by experts convened to define it and hence is somewhat arbitrary. There is broad consensus that mental disorders exist on dimensions, not categories, and in 2008 two members who would later join the HiTOP consortium convened a meeting and edited a seminal book, *Dimensional Approaches in Diagnostic Classification*, as part of work on refining the research agenda for DSM-5[3].

Multivariate research has indicated that a latent general liability – *internalizing* – accounts for higher-than-chance levels of mood and anxiety disorder comorbidity, a finding that has been replicated and extended many times in different data sets and cultures (note that half of people who meet criteria for an anxiety or depressive disorder have a second diagnosis, and a quarter meet criteria for three or more).

For example, within the HiTOP consortium, Eaton et al[4] modelled seven internalizing disorders in a nationally representative sample of 43,093 individuals. The study used a structured diagnostic interview optimized to cover the DSM-IV defining characteristics of these disorders. They found that a two-dimensional (distress-fear) liability structure for internalizing fit best and replicated across gender, assessment waves, and lifetime and 12-month diagnoses. These internalizing liabilities, *not the individual disorders,* predicted future internalizing pathology, suicide attempts, angina, and ulcer.

Waszczuk et al[5] conducted a study based on the Interview for Mood and Anxiety Symptoms that assessed, without the usual skip outs, DSM-IV and ICD-10 emotional disorder symptoms and other manifestations of emotional disorders such as hopelessness, desperation, loss of libido, social withdrawal, and self-harm. In a series of analyses that ranged from symptom components to latent structures, they reported that dimensional components are better predictors of functioning than categorical DSM-IV diagnoses, even though impairment is explicitly included in clinical diagnoses but is not part of those symptom components.

There are two implications from this body of work. First, that considering groups of disorders may be more informative than considering individual diagnoses. Second, that opening up research to include symptoms not presently included in classifications may point to new disorders or new arrangements of existing disorders and reduce the circularity of reanalyzing data from interviews designed to inform existing classifications.

There has been other work on classification independent of the HiTOP consortium that is relevant to the current Forum. As part of the work for DSM-5 and ICD-11, a working group[6], including two members who would later join the HiTOP consortium, explored the feasibility of a meta-structure based on eleven validating criteria comprising both clinical features and risk factors (i.e., shared genetic risk factors; familiarity; shared specific environmental risk factors; shared neural substrates; shared biomarkers; shared temperamental antecedents; shared abnormalities of cognitive or emotional processing; symptom similarity; high rates of comorbidity; course of illness; treatment response). DSM-IV disorders were allocated to one of five clusters as a starting premise. Teams of experts then reviewed the literature to determine within-cluster similarities on the eleven predetermined validating criteria and discovered that those similarities were consistently greater than between-cluster similarities.

The five clusters were neurocognitive (identified principally by neural substrate abnormalities), neurodevelopmental (identified principally by early and continuing cognitive deficits), psychosis (identified principally by clinical features and biomarkers for information processing deficits), emotional/internalizing (identified principally by the temperamental antecedent of negative emotionality), and externalizing (identified principally by the temperamental antecedent of disinhibition). The working group considered that there could be advantages for clinical practice, public administration and principally from the adoption of such an organizing principle. The chapter order in DSM-5 was changed to reflect this.

Computerized cognitive behavioural therapy (CBT) has a long history of focussing on the internalizing disorders as a group. Newby et al[7] identified seventeen randomized controlled trials. Results showed that "transdiagnostic" computerized CBT outperformed control conditions on all outcome measures at post-treatment, with large effect sizes for depression (g=0.84), and medium effect sizes for anxiety (g=0.78) and quality of life (g=0.48), comparable to the benefits seen in diagnosis specific studies[8].

Lastly, and again using "transdiagnostic" computerized CBT, Mewton et al[9] assessed changes in the internalizing construct using a longitudinal latent trait framework that compared internalizing factor means at pre- and post-treatment. The standardized mean reduction in the internalizing construct with treatment was large (effect size 1.23, SE=0.09, $p< 0.001$).

We conclude that treatment aimed at the internalizing construct is to be preferred to disorder specific treatment. In the internalizing disorders, whether one is investigating prognosis, impairment or response to treatment, the whole is greater than the sum of the parts.

Gavin Andrews
School of Psychiatry, University of New South Wales, Sydney, NSW, Australia

1. Kotov R, Krueger RF, Watson D. World Psychiatry 2018;17:24-5.
2. Krueger RF, Kotov R, Watson D et al. World Psychiatry 2018;17:282-93.
3. Helzer JE, Kraemer HC, Krueger RF et al. Dimensional approaches in diagnostic classification. Arlington: American Psychiatric Association, 2008.
4. Eaton NR, Krueger RF, Markon KE et al. J Abnorm Psychol 2013;122:86-92.
5. Waszczuk M, Kotov R, Ruggero C et al. J Abnorm Psychol 2017;126:613-34.
6. Andrews G, Goldberg DP, Krueger RF et al. Psychol Med 2009;39:1993-2000.
7. Newby JM, Twomey C, Yuan Li SS et al. J Affect Disord 2016;199:30-41.
8. Andrews G, Basu A, Cuijpers P et al. J Anxiety Disord (in press).
9. Mewton L, Hobbs MJ, Sunderland M et al. Behav Res Ther 2014;63:132-8.

DOI:10.1002/wps.20564

Categorical and/or continuous? Learning from vascular surgery

R. Krueger and his impressive international team of co-authors offer a clear and comprehensive review of current issues in dimensional approaches to classifying psychopathology[1]. They make a challenging case for the advantages of this approach, as embodied in their Hierarchical Taxonomy of Psychopathology (HiTOP) model, over the categorical classifications widely adopted in psychiatry.

The issues as such are not new. They were widely discussed in the 1960s and 1970s. The British psychiatrist and epidemiologist R. Kendell covered much the same ground in his now classic book *The Role of Diagnosis in Psychiatry*[2]. Then as now the question was whether psychopathology could be "cut at the joints" into distinct categories or whether it was better described quantitatively along one or more dimensions of continuous change. Then as now the answer depended in part on the kind of psychopathology in question and in part on the statistical methods adopted. Then as now opinion remained divided largely along professional lines, with psychiatrists favouring categorical and psychologists favouring dimensional approaches (Krueger, like a majority of his co-authors, is a psychologist).

There are, certainly, as Krueger and his colleagues point out, new factors in play, some positive, others negative. On the positive side, there have been important methodological advances. Statistical methods have progressed dramatically with developments in computing science. Formal logic, too, has a novel role to play: the British philosopher and psychologist P. Koralus' semantic modelling of decision making, for example, offers potentially exciting applications to psychopathology[3]. On the negative side, fifty years of experience with symptom-based psychopathological categories have been disappointingly thin on aetiological insights. The promise of early 20th century advances (with discoveries such as neurosyphilis and Alzheimer's disease) remains, despite all the power of contemporary neuroscience, largely unfulfilled.

Should we then be persuaded by Krueger et al's case that categorical classifications of psychopathology should be abandoned in favour of dimensional description? Experience from other areas of medicine suggests that we should not.

Vascular surgery offers a case in point. As a relatively new specialty (the Vascular Surgical Society of Great Britain and Ireland was founded in 1966), vascular surgery adopted from the start an explicitly evidence-based approach and remains strongly research-led. In this respect, its predominantly categorical classification of disease entities has (as in most other areas of bodily medicine) served it well. Where psychiatry has suffered fifty years of frustration, vascular surgery has made significant and sustained progress in understanding the pathophysiology of a whole range of categorically-defined disorders, ranging from aortic aneurysm to varicose veins, with corresponding advances in both surgical and non-surgical management options.

So far so good then, it would seem, for traditional disease entities. However, closer inspection shows that, while the objects of scientific interest in vascular surgery are indeed categorically defined disorders, the science of vascular surgery has been in many instances dimensional in character. Progress in the treatment of aortic aneurysm, for example, has depended critically on quantitative studies of the relative risk of death respectively from vascular surgery and from aneurysm rupture. The key variable in these studies is the diameter of the aneurysm. The risk of rupture increases as the aneurysm expands. In most people this happens slowly, and international guidelines recommend annual monitoring until the diameter of the patient's aneurysm reaches five and a half centimeters, this being the point at which the risk of rupture within the next twelve months (5%) is sufficient to justify the risks of surgery[4].

Vascular surgical science has thus made progress by combining categorical with dimensional approaches. Similar combined approaches continue to be adopted in ongoing research on the management of aortic aneurysm. The object of interest remains the categorically defined disease entity "aortic aneurysm"; the key variables remain the essentially dimensional variables of relative risk.

Psychopathology, it is true, is different from and in certain respects more complex than vascular pathology. There are, for example, no counterparts in vascular pathology of the conceptual challenges presented by comorbidity in psychopathology (reflected in the difficulties described by Krueger et al in establishing a stable hierarchical structure for their dimensional approach). Comorbidities are, of course, common in vascular pathology, but the requisite divisions and distinctions are largely unproblematic. Similarly unproblematic in vascular pathology are criteria of functioning. Descriptively similar experiences of voice hearing, for example, may be for one person functionally impairing and for another empowering[5,6]. A swelling aorta, by contrast, is a functionally impaired aorta for anyone.

Such differences, though, make the example of vascular pathology more rather than less pertinent for psychopathology. If progress in vascular pathology has been achieved with a combined categorical and dimensional approach, it is at the very least likely that a similar approach will be needed if progress is to be made with the more complex challenges of psychopathology. The point, anyway, is general. All sciences make progress through quantification. But progress through quantification has usually been by way of addition, not substitution. This is evident throughout the medical and biological sciences. It is evident, too, in physics, surely the paradigm of a successful quantitative science (think of wave/particle dualism in quantum mechanics). So why should psychopathology be any different?

Krueger et al might reply: "because this is where the science leads". In the opening paragraphs of their paper, they claim in support of their HiTOP model

the high ground of empirical science, contrasting this with what they describe as the received authority of the DSM. But this is tendentious. The scientific basis specifically of DSM-5 has indeed been widely criticized[7]. But the criticism is precisely that DSM-5 has departed from the explicitly evidence-based principles on which earlier revisions of the DSM (and ICD) were based. Notably, the Research Domain Criteria project, although bracketed by Krueger et al with DSM-5, was in fact inspired by much the same aims as HiTOP for a return to empiricism in psychopathological research[8].

We should thus welcome the advances in quantification of psychopathology described by Krueger et al. But we should welcome these advances as adding to rather than displacing categorical classifications as the basis of psychopathological science. More will be required for effective translation of psychopathological science into practice. In vascular surgery, translation has required teamwork rather than competition between professionals, and attention to values as well as evidence[9]. But, as to the science, the example of vascular surgery suggests that it is time for a change of conjunction. For fifty years the focus of debate in psychopathology has been "categorical *or* continuous". The example of vascular surgery suggests that its time to think instead "categorical *and* continuous".

Kenneth W.M. Fulford[1,2], Ashok Handa[1,3]
[1]St. Catherine's College, Oxford, UK; [2]Faculty of Philosophy, Oxford, UK; [3]Nuffield Department of Surgical Sciences, J. Radcliffe Hospital, Oxford, UK

1. Krueger RF, Kotov R, Watson D et al. World Psychiatry 2018;17:282-93.
2. Kendell RE. The role of diagnosis in psychiatry. Oxford: Blackwell, 1975.
3. Parrott M, Koralus P. J Cogn Neuropsychiatry 2015;20:398-415.
4. The UK Small Aneurysm Trial Participants. Lancet 1998;352:1649-55.
5. Jackson M, Fulford KWM. Philos Psychiatry Psychol 1997;4:41-66.
6. Daalman K, Diederen K, Derks E et al. Psychol Med 2012;42:2475-84.
7. Frances A. Saving normal: an insider's revolt against out-of-control psychiatric diagnosis, DSM-5, big pharma, and the medicalization of ordinary life. New York: Morrow, 2013.
8. Insel TR. Transforming diagnosis. www.nimh.nih.gov.
9. Handa IA, Fulford-Smith L, Barber ZE et al. The importance of seeing things from someone else's point of view. BMJ Careers, August 15, 2016.

DOI:10.1002/wps.20565

Clinical utility of ICD-11 diagnostic guidelines for high-burden mental disorders: results from mental health settings in 13 countries

Geoffrey M. Reed[1,2], Jared W. Keeley[3], Tahilia J. Rebello[1,4], Michael B. First[1,4], Oye Gureje[5], José Luis Ayuso-Mateos[6], Shigenobu Kanba[7], Brigitte Khoury[8], Cary S. Kogan[9], Valery N. Krasnov[10], Mario Maj[11], Jair de Jesus Mari[12], Pratap Sharan[13], Dan J. Stein[14], Min Zhao[15], Tsuyoshi Akiyama[16], Howard F. Andrews[1,4,17], Elson Asevedo[12], Majda Cheour[18], Tecelli Domínguez-Martínez[2,19], Joseph El-Khoury[8], Andrea Fiorillo[11], Jean Grenier[20], Nitin Gupta[21], Lola Kola[5], Maya Kulygina[10], Itziar Leal-Leturia[6], Mario Luciano[11], Bulumko Lusu[13], J. Nicolás I. Martínez-López[2], Chihiro Matsumoto[22], Mayokun Odunleye[23], Lucky Umukoro Onofa[24], Sabrina Paterniti[25], Shivani Purnima[13], Rebeca Robles[2], Manoj K. Sahu[26], Goodman Sibeko[13], Na Zhong[15], Wolfgang Gaebel[27], Anne M. Lovell[28], Toshimasa Maruta[29], Kathleen M. Pike[1], Michael C. Roberts[30], María Elena Medina-Mora[2]

[1]Department of Psychiatry, Columbia University College of Physicians and Surgeons, New York, NY, USA; [2]National Institute of Psychiatry Ramón de la Fuente Muñiz, Mexico City, Mexico; [3]Department of Psychology, Virginia Commonwealth University, Richmond, VA, USA; [4]New York State Psychiatric Institute, New York, NY, USA; [5]Department of Psychiatry, University of Ibadan, Ibadan, Nigeria; [6]Department of Psychiatry, Universidad Autonoma de Madrid; Instituto de Salud Carlos III, Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM); Instituto de Investigación Sanitaria La Princesa, Madrid, Spain; [7]Department of Neuropsychiatry, Kyushu University, Fukuoka City, Japan; [8]Department of Psychiatry, American University of Beirut Medical Center, Beirut, Lebanon; [9]School of Psychology, University of Ottawa, Ottawa, ON, Canada; [10]Moscow Research Institute of Psychiatry, National Medical Research Centre for Psychiatry and Narcology, Moscow, Russian Federation; [11]Department of Psychiatry, University of Campania "L. Vanvitelli", Naples, Italy; [12]Department of Psychiatry, Universidade Federal de São Paulo, São Paulo, Brazil; [13]Department of Psychiatry, All India Institute of Medical Sciences, New Delhi, India; [14]Department of Psychiatry, University of Cape Town and South African Medical Research Council Unit on Risk and Resilience in Mental Disorders, Cape Town, South Africa; [15]Shanghai Mental Health Center and Department of Psychiatry, Shanghai Jiao Tong University School of Medicine, Shanghai, People's Republic of China; [16]NTT Medical Center Tokyo, Tokyo, Japan; [17]Department of Biostatistics, Columbia University College of Physicians and Surgeons, New York, NY, USA; [18]Department of Psychiatry, Tunis Al Manar University and Al Razi Hospital, Tunis, Tunisia; [19]Cátedras CONACYT, National Council for Science and Technology, Mexico City, Mexico; [20]Institut du Savoir Montfort - Hôpital Montfort & Université d'Ottawa, Ottawa, ON, Canada; [21]Department of Psychiatry, Government Medical College and Hospital, Chandigarh, India; [22]Japanese Society of Psychiatry and Neurology, Tokyo, Japan; [23]Department of Psychiatry, University College Hospital, Ibadan, Nigeria; [24]Federal Neuropsychiatric Hospital Aro, Abeokuta, Nigeria; [25]Institute of Mental Health Research, Royal Ottawa Mental Health Centre, and Department of Psychiatry, University of Ottawa, Ottawa, ON, Canada; [26]Pt. Jawahar Lal Nehru Memorial Medical College, Raipur, Chhattisgarh, India; [27]Department of Psychiatry and Psychotherapy, Medical Faculty, Heinrich-Heine University, Düsseldorf, Germany; [28]Institut National de la Santé et de la Recherche Médicale U988, Paris, France; [29]Health Management Center, Seitoku University, Matsudo City, Japan; [30]Office of Graduate Studies and Clinical Child Psychology Program, University of Kansas, Lawrence, KS, USA

In this paper we report the clinical utility of the diagnostic guidelines for ICD-11 mental, behavioural and neurodevelopmental disorders as assessed by 339 clinicians in 1,806 patients in 28 mental health settings in 13 countries. Clinician raters applied the guidelines for schizophrenia and other primary psychotic disorders, mood disorders (depressive and bipolar disorders), anxiety and fear-related disorders, and disorders specifically associated with stress. Clinician ratings of the clinical utility of the proposed ICD-11 diagnostic guidelines were very positive overall. The guidelines were perceived as easy to use, corresponding accurately to patients' presentations (i.e., goodness of fit), clear and understandable, providing an appropriate level of detail, taking about the same or less time than clinicians' usual practice, and providing useful guidance about distinguishing disorder from normality and from other disorders. Clinicians evaluated the guidelines as less useful for treatment selection and assessing prognosis than for communicating with other health professionals, though the former ratings were still positive overall. Field studies that assess perceived clinical utility of the proposed ICD-11 diagnostic guidelines among their intended users have very important implications. Classification is the interface between health encounters and health information; if clinicians do not find that a new diagnostic system provides clinically useful information, they are unlikely to apply it consistently and faithfully. This would have a major impact on the validity of aggregated health encounter data used for health policy and decision making. Overall, the results of this study provide considerable reason to be optimistic about the perceived clinical utility of the ICD-11 among global clinicians.

Key words: International Classification of Diseases, ICD-11, diagnosis, mental disorders, clinical utility, ease of use, goodness of fit, treatment selection, assessing prognosis

(World Psychiatry 2018;17:306–315)

The World Health Organization (WHO) has released the 11th revision of the International Classification of Diseases and Related Health Problems to its member states to prepare for implementation[1]. The new classification will be presented for approval by the World Health Assembly, the WHO's governing body, in May 2019.

As we have previously described[2-5], an important focus in the development of the ICD-11 chapter on Mental, Behavioural and Neurodevelopmental Disorders by the WHO Department of Mental Health and Substance Abuse has been to improve its clinical utility.

For the purpose of developing the ICD classification of mental disorders, the WHO has defined the clinical utility of a classification construct, category, or system as depending on: a) its value in communicating (e.g., among practitioners, patients, families, administrators); b) its implementation characteristics in clinical practice, including its goodness of fit (i.e., accuracy of description), its ease of use, and the time required to use it (i.e., feasibility); and c) its usefulness in selecting interventions and in making clinical management decisions[2]. This definition is based in part on those proposed by M. First and colleagues[6,7].

Similar concepts had also been included in the ICD-10 field trails[8,9], which asked clinicians to provide ratings of goodness of fit, confidence in their selected diagnosis, ease or difficulty of making a diagnosis, and adequacy of the diagnostic guidelines for cases evaluated as a part of the study.

In a recent study[10], we expanded the operationalization of clinical utility considerably to include an assessment of utility in relation to specific components of the diagnostic guidelines as well as to specific uses of the guidelines (e.g., meeting administrative requirements, assigning a diagnosis, treatment selection, communication, teaching).

Moreover, the WHO Department of Mental Health and Substance Abuse has conducted a major programmatic field studies effort for ICD-11 focusing on clinical utility[3]. This program of research extends the concept of clinical utility to include diagnostic accuracy and diagnostic consistency, as diagnoses that are neither accurate nor reliable are unlikely to be useful.

Thus, there are both subjective and objective components to clinical utility, and these overlap to some extent with both reliability and validity[2]. Clinical utility is not simply a matter of clinician preferences. Nonetheless, the subjective components are important because clinicians who do not feel that a classification system provides them with useful and valuable information are unlikely to apply it carefully, with major implications for the quality of health encounter data related to diagnosis.

Finally, with the goal of improving clinical utility, the Department of Mental Health and Substance Abuse has made a series of substantive changes in the Clinical Descriptions and Diagnostic Guidelines (CDDG) for ICD-11 Mental, Behavioural and Neurodevelopmental Disorders as compared to the ICD-10 CDDG[11]. The CDDG is the version that is intended to be used by mental health professionals in clinical settings. Many of these changes have involved ensuring that the ICD-11 CDDG provide consistent and relatively uniform diagnostic information across the various categories[4], something that has been identified as a shortcoming of the ICD-10 CDDG. Diagnostic guidelines have been drafted so as to allow for the appropriate exercise of clinical judgment, minimizing the use of arbitrary or pseudo-precise symptom counts and cutoffs when these are not strongly supported by evidence. The new structure of groupings and categories for ICD-11 is also intended to be more logical and more consistent with how clinicians conceptualize mental disorders[12,13].

The data presented in this paper were collected as a part of the ICD-11 developmental field study of reliability of diagnoses of high-burden mental disorders, undertaken in 13 countries around the world. The initial reliability data have been published in this journal[14], indicating that the joint-rater reliability of the ICD-11 diagnostic guidelines ranged from moderate to almost perfect (.45 to .88)[15], and was generally superior to results obtained for ICD-10[8]. The current paper focuses on clinicians' evaluations of the clinical utility of the diagnostic guidelines, using a scale that is based in part on clinical utility concepts from the ICD-10 field trial, but that more fully operationalizes the WHO's definition of clinical utility for ICD-11.

METHODS

Study design and procedures

Two study protocols were implemented to assess the clinical utility and the reliability of the proposed ICD-11 diagnostic guidelines. Protocol 1 tested the utility and reliability of the guidelines for schizophrenia and other primary psychotic disorders and for mood disorders, while Protocol 2 tested the guidelines for mood disorders, anxiety and fear-related disorders, and disorders specifically associated with stress.

Adult (≥18 years of age) patients exhibiting any psychotic symptoms and presenting for care at a participating study site were eligible to participate in Protocol 1, while adult patients exhibiting mood symptoms, anxiety symptoms, or stress-related symptoms but no psychotic symptoms and presenting for care at the participating field study center were eligible to participate in Protocol 2. Prospective participants who met these criteria were excluded only if they could not reasonably be expected to participate in the diagnostic assessment (e.g., for reasons of language or cognitive impairment).

These relatively loose criteria were in part intended to more closely approximate the natural circumstances under which the ICD-11 will be implemented in mental health settings.

Study protocols were implemented at 28 sites in 13 countries[14]. The local language was always used for the diagnostic assessments. The ICD-11 guidelines, training materials, and all material for the study were developed in English and then translated into four other languages: Chinese, Japanese, Russian and Spanish. For Tunisia, the guidelines, but not all of the other training materials, were translated into French. In other sites where English was not the local language (e.g., Brazil, Italy), the English guidelines and training materials were used even though the interviews were conducted in the local language, again replicating the circumstances under which the ICD-11 will be implemented in many settings. Details on clinician recruitment and training, study implementation processes, data collection, and ethical clearance have been provided previously[14].

Following informed consent, patients were interviewed by two clinicians with whom they had not had any prior clinical contact. One clinician rater served as the primary interviewer and the second as an observer. The observer was allowed to ask additional follow-up questions at the end of the interview. Based on the interview, clinician raters independently arrived at a diagnostic formulation consisting of up to three diagnoses. Diagnoses were non-hierarchical (i.e., not specified as primary, secondary or tertiary) and could fall within any mental, behavioural or neurodevelopmental disorder diagnostic grouping in the ICD-11. Participating clinicians could also specify a non-mental or behavioural disorder diagnosis, or no diagnosis.

After finalizing their selected diagnostic formulation, clinicians were asked twelve detailed questions about the clinical utility of the diagnostic guidelines as applied to that particular patient. These included: core clinical utility questions (ease of use, goodness of fit, clarity and understandability), questions on implementation characteristics of the guidelines (level of detail, feasibility of assessment requirements, time required), questions about the utility of specific sections of the guidelines (boundary with normality and differential diagnosis), and questions about the utility of the guidelines for specific purposes (selecting a treatment, predicting prognosis, communicating with other professionals, educating patients and family members). Specific wording of the questions and the Likert-type response options for each question are shown in Table 1.

Clinicians provided clinical utility ratings for the specific categories that were part of diagnostic groupings which were the focus of Protocols 1 and 2, i.e., schizophrenia and other primary psychotic disorders, mood disorders (including depressive disorders and bipolar disorders), anxiety and fear-related disorders, and disorders specifically associated with stress. If more than one diagnosis from these groupings was applied to a particular patient, clinical utility ratings were made for all such selected diagnoses taken together rather than for each diagnosis separately.

Participants

A total of 339 clinicians from the 28 study sites in 13 countries served as clinician raters for Protocol 1 and/or Protocol 2. The mean age of clinician raters was 37.2±8.3 years, and their ages were comparable across countries. There was a slight majority of male clinician raters in the global sample (56.6%). The overwhelming majority of clinician raters in the study were psychiatrists (93.2%), with a small representation of psychologists (3.8%), nurses (1.5%) and other health care professionals (1.5%). Clinicians had an average of 7.6±7.5 years of professional clinical experience following completion of their clinical training (including post-graduate training).

As shown in Table 2, 1,806 patients participated in the study for Protocol 1 (N=1,041) or Protocol 2 (N=765). The average age of participating patients was 39.9±13.7 years, and was comparable across countries. The global patient sample had an equal gender distribution. The marital status of the majority of patients across countries was single (54.9%); 33.1% were married/cohabitating, 9.8% were separated or divorced, and 2.2% were widowed. More than half of the patients in the global sample were unemployed (55.9%) and only 22.3% had full time employment. A slight majority of patients who participated in the study were inpatients (55.0%) and the remainder were mostly outpatients (44.4%). The small remaining proportion (0.6%) were enrolled in other types of programs such as partial day hospitalization.

Data collection, management and processing

Clinician interviewers entered interview data using the Electronic Field Study System (EFSS), a secure web-based data collection system developed using Qualtrics™ (Provo, UT, USA) survey software, made available in all five study languages. Data from the sites were stored and managed centrally by the Data Coordinating Center (DCC) at Columbia University.

Data quality was established through continuous monitoring of the data collection procedures by local research staff at each site and through use of programmed functions within Qualtrics™, such as forced response and content validation options. This provided a mechanism for collecting data in a standardized, uniform format from all sites. Site-based research teams kept records of any errors in data entry that were discovered during the review process and these were passed on to the DCC for correction.

Data analysis

A total of 3,608 sets of clinical utility ratings were made by the 339 clinicians. Because there were two raters for each patient, the N for each analysis should be double that of the number of patients (N=1,806; see Table 1), but in four cases only one set of ratings was available for a particular patient.

Clinician raters' responses to each of the 12 clinical utility variables were summarized using frequency counts for each response. To provide a metric of overall favorable responses, ratings of "Quite" and "Extremely" were combined for questions where this was appropriate (see Table 1). Responses to the clinical utility variables by country were also calculated (not all reported; available from the authors by request), as were responses to clinical utility for the five most commonly used diagnoses.

For reliability analyses, intraclass kappa coefficients were calculated with bootstrapped 95% confidence intervals, based on 1,000 resamples, for each country. Reliability coefficients were calculated for only the most common diagnoses within the study (i.e., N≥130), to maximize the chance of having a sufficient number of diagnoses within a country to estimate kappa. Per-diagnosis ratings of clinical utility were also calculated for these same diagnoses.

RESULTS

Clinical utility ratings across countries are shown in Table 1. Evaluations were overwhelmingly positive, though with some differences between items.

For the three core clinical utility questions (ease of use, goodness of fit, clarity and understandability), the overwhelming majority of participants (82.5 to 83.9%) provided ratings of "Quite" or "Extremely", indicating favourable clinical utility.

Table 1 Clinical utility questions and responses across countries (N=3,608)

Core clinical utility questions

Please rate the overall ease of use of the diagnostic guidelines with respect to this patient:

Not at all:	Somewhat:	Quite:	Extremely:	Quite + extremely:
32 (0.9%)	556 (15.4%)	2,471 (68.5%)	549 (15.2%)	3,020 (83.7%)

Please rate the overall goodness of fit or accuracy of the diagnostic guidelines with respect to this patient:

Not at all:	Somewhat:	Quite:	Extremely:	Quite + extremely:
28 (0.8%)	604 (16.7%)	2,497 (69.2%)	479 (13.3%)	2,976 (82.5%)

Please rate the extent to which the diagnostic guidelines were clear and understandable overall as applied to this patient:

Not at all:	Somewhat:	Quite:	Extremely:	Quite + extremely:
14 (0.4%)	567 (15.7%)	2,473 (68.5%)	554 (15.4%)	3,027 (83.9%)

Implementation characteristics

Which of the following statements best describes your evaluation of the level of detail and specificity of the essential features for the diagnosis or diagnoses that you applied to this patient?

Insufficient:	About the right amount:	Too much:
148 (4.1%)	3,275 (90.8%)	185 (5.1%)

Please rate the extent to which the guidelines imposed assessment requirements that were difficult to apply to this patient (e.g., requirements that rely too much on the patient's memory of remote events or the patient's ability to report temporal relationships between symptoms):

Very difficult:	Somewhat difficult:	Quite easy:	Extremely easy:	Quite + extremely easy:
35 (1.0%)	518 (14.4%)	2,752 (76.3%)	303 (8.4%)	3,055 (84.7%)

How would you describe the amount of time that it took you to apply all of the Essential Features to this patient for the diagnosis or diagnoses that you selected, in comparison to your usual clinical practice?

Much longer:	Somewhat longer:	About the same:	Shorter:
30 (0.8%)	472 (13.1%)	2,669 (74.0%)	437 (12.1%)

Specific sections

Please rate the extent to which the description of the boundary between disorder and normality contained in the guidelines was useful as applied to this patient:

Not at all:	Somewhat:	Quite:	Extremely:	Quite + extremely:
78 (2.2%)	770 (21.3%)	2,304 (63.9%)	456 (12.6%)	2,760 (76.5%)

Please rate the extent to which the description of the boundary between this patient's disorder and other disorders (section on differential diagnosis) was useful as applied to this patient:

Not at all:	Somewhat:	Quite:	Extremely:	Quite + extremely:
49 (1.4%)	762 (21.1%)	2,322 (64.4%)	475 (13.2%)	2,797 (77.5%)

Specific uses

How useful would the diagnostic guidelines be in helping you to select a treatment for this patient?

Not at all:	Somewhat:	Quite:	Extremely:	Quite + extremely:
70 (1.9%)	887 (24.6%)	2,223 (61.6%)	428 (11.9%)	2,651 (73.5%)

How useful would the diagnostic guidelines be in helping you to assess this patient's prognosis?

Not at all:	Somewhat:	Quite:	Extremely:	Quite + extremely:
83 (2.3%)	1,055 (29.2%)	2,104 (58.3%)	366 (10.1%)	2,470 (68.5%)

How useful would the diagnostic guidelines be in helping you to communicate about this patient with a colleague or other health care professional?

Not at all:	Somewhat:	Quite:	Extremely:	Quite + extremely:
49 (1.4%)	746 (20.7%)	2,216 (61.4%)	597 (16.5%)	2,813 (78.0%)

How useful would the diagnostic guidelines be in helping you to educate this patient and/or family about his or her condition?

Not at all:	Somewhat:	Quite:	Extremely:	Quite + extremely:
52 (1.4%)	884 (24.5%)	2,236 (62.0%)	436 (12.1%)	2,672 (74.1%)

For implementation characteristics, a large majority indicated that the guidelines did not impose assessment requirements that were difficult to apply (84.7%), provided about the right level of detail (90.4%), and took about the same amount of time or less time than their usual practice (86.1%).

Regarding specific sections, the manner in which the guidelines provided guidance about differentiating disorders from

Table 2 Patient demographics by country

	Total (N=1,806)	Brazil (N=100)	Canada (N=53)	China (N=203)	India (N=209)	Italy (N=100)	Japan (N=168)	Lebanon (N=103)	Mexico (N=153)	Nigeria (N=132)	Russia (N=104)	South Africa (N=208)	Spain (N=70)	Tunisia (N=203)
Age, years (mean ± SD)	39.9 ± 13.7	32.9 ± 9.6	39.8 ± 14.2	43.9 ± 15.6	36.5 ±11.4	41.4 ±11.2	47.0 ±15.1	36.4 ±12.5	38.1 ±13.0	37.5 ±12.2	36.3 ±11.7	35.1 ±11.0	52.0 ± 16.2	43.2 ± 12.6
Gender, N (%)														
Male	908 (50.3)	62 (62.0)	19 (35.8)	123 (60.6)	120 (57.4)	50 (50.0)	72 (42.9)	38 (36.9)	48 (31.4)	65 (49.2)	44 (42.3)	133 (63.9)	26 (37.1)	108 (53.2)
Female	897 (49.7)	38 (38.0)	33 (62.3)	80 (39.4)	89 (42.6)	50 (50.0)	96 (57.1)	65 (63.1)	105 (68.6)	67 (50.8)	60 (57.7)	75 (36.1)	44 (62.9)	95 (46.8)
Relationship status, N (%)														
Single	992 (54.9)	81 (81.0)	22 (41.5)	110 (54.2)	66 (31.6)	71 (71.0)	77 (45.8)	68 (66.0)	91 (59.5)	68 (51.5)	65 (62.5)	167 (80.3)	28 (40.0)	78 (38.4)
Married/ cohabitating	597 (33.1)	12 (12.0)	17 (32.1)	75 (36.9)	133 (63.6)	19 (19.0)	64 (38.1)	20 (19.4)	42 (27.5)	41 (31.1)	22 (21.2)	25 (12.0)	28 (40.0)	99 (48.8)
Separated/divorced	177 (9.8)	6 (6.0)	13 (24.5)	15 (7.4)	4 (1.9)	7 (7.0)	21 (12.5)	15 (14.6)	20 (13.1)	18 (13.6)	13 (12.5)	13 (6.3)	9 (12.9)	23 (11.3)
Widowed	40 (2.2)	1 (1.0)	1 (1.9)	3 (1.5)	6 (2.9)	3 (3.0)	6 (3.6)	0	0	5 (3.8)	4 (3.8)	3 (1.4)	5 (7.1)	3 (1.5)
Employment, N (%)														
Full time	403 (22.3)	4 (4.0)	14 (26.4)	47 (23.2)	69 (33.0)	11 (11.0)	26 (15.5)	16 (15.5)	17 (11.1)	41 (31.1)	22 (21.2)	22 (10.6)	26 (37.1)	88 (43.3)
Part time	142 (7.9)	5 (5.0)	6 (11.3)	3 (1.5)	12 (5.7)	9 (9.0)	14 (8.3)	11 (10.7)	31 (20.3)	11 (8.3)	6 (5.8)	8 (3.8)	3 (4.3)	23 (11.3)
Unemployed	1009 (55.9)	76 (76.0)	30 (56.6)	80 (39.4)	110 (52.6)	74 (74.0)	109 (64.9)	66 (64.1)	79 (51.6)	64 (48.5)	53 (51.0)	167 (80.3)	20 (28.6)	81 (39.9)
Student	136 (7.5)	6 (6.0)	4 (7.5)	15 (7.4)	15 (7.2)	4 (4.0)	10 (6.0)	15 (14.6)	30 (19.6)	10 (7.6)	7 (6.7)	12 (5.8)	2 (2.9)	6 (3.0)
Retired	152 (8.4)	10 (10.0)	1 (1.9)	62 (30.5)	3 (1.4)	2 (2.0)	15 (8.9)	0	5 (3.3)	8 (6.1)	18 (17.3)	0	22 (31.4)	6 (3.0)
Treatment setting, N (%)														
Outpatient	801 (44.4)	82 (82.0)	53 (100)	0	122 (58.4)	67 (67.0)	48 (28.6)	14 (13.6)	135 (88.2)	84 (63.6)	4 (3.8)	0	49 (70.0)	143 (70.4)
Inpatient	994 (55.0)	18 (18.0)	0	203 (100)	87 (41.6)	33 (33.0)	120 (71.4)	89 (86.4)	17 (11.1)	48 (36.4)	91 (87.5)	207 (99.5)	21 (30.0)	60 (29.6)
Other	11 (0.6)	0	0	0	0	0	0	0	1 (0.7)	0	9 (8.7)	1 (0.5)	0	0

normality and from other disorders was also rated very positively, with 76.5% and 77.5% of participants, respectively, indicating that these sections were quite or extremely useful.

Regarding the clinical utility of the guidelines for specific purposes, 78.0% of participants indicated that they would be quite or extremely useful for communicating with colleagues or other professionals. The lowest, though still positive overall, ratings were provided for potential usefulness in selecting a treatment (73.5%) and assessing prognosis (68.5%).

We also examined variations in clinical utility ratings across countries. Table 3 shows ratings by country for the three core clinical utility questions. Ratings by country for other clinical utility variables (see Table 1) are not reported here, but are available upon request. The most apparent variation across these three questions is that the ratings shown are substantially lower for Japan (47.9 to 49.7% answering "Quite" or "Extremely") and somewhat lower for Tunisia (69.0 to 70.4%) as compared to the proportion of participants answering "Quite" or "Extremely" for other countries (81.5 to 97.9%).

If variability in perceived clinical utility were directly related to the adequacy of the guidelines, it might be expected that perceived clinical utility and inter-rater reliability would vary together. Table 4 shows concurrent reliability or joint rater agreement, represented by interclass kappa with bootstrapped 95% confidence intervals, for the five most common diagnoses among the sample: schizophrenia, schizoaffective disorder, bipolar type I disorder, single episode depressive disorder and recurrent depressive disorder. While there is clearly variability in reliability by country, there is not a discernible relationship between lower ratings of clinical utility by Japanese and Tunisian participants and the reliability of their diagnostic assignments. Conversely, lower reliability coefficients (e.g., for the Russian Federation) did not correspond with low perceptions of clinical utility.

Clinical utility ratings by diagnosis are shown for these same five diagnoses in Table 5. Across the three core overall clinical utility questions, depressive disorders had slightly lower ratings than schizophrenia, schizoaffective disorder, and bipolar disorder. Slightly lower reliability estimates for single episode depressive disorder and recurrent depressive disorder appear to correspond to slightly lower clinical utility ratings for these categories, but schizoaffective disorder had very high clinical utility ratings in spite of having similarly lower reliability.

DISCUSSION

In the current analyses, clinician ratings of clinical utility of the proposed ICD-11 diagnostic guidelines proved to be very positive overall. This was likely in part related to the attention to clinical utility in the construction of the guidelines[4], as well as the fact that they had already been tested in Internet-based studies in global, multilingual studies via the Global Clinical Practice Network (https://gcp.network) and refined on that basis[16,17].

The guidelines were perceived as easy to use, corresponding accurately to patients' presentations (i.e., goodness of fit), clear and understandable, providing an appropriate level of detail, taking about the same or less time than clinicians' usual practice, and providing useful guidance about distinguishing disorder from normality and from other disorders. Clinicians evaluated the guidelines as relatively less useful for treatment selection and assessing prognosis than for communicating with other health professionals, though the former ratings were still positive overall.

As described, two of the core clinical utility questions used in this study were based on questions used in the ICD-10 field study[8,9]. In that study, 82.5% of participating global clinicians rated the goodness of fit of ICD-10 guidelines as good or very good, and 85.0% said that they were moderately or very easy to use[18]. These percentages are nearly identical to the ones obtained in this study for the ICD-11 guidelines, but differences in the scaling (see Table 1) suggest that the current results could be viewed as more positive.

It should be noted that participating clinicians would likely have been disposed to view the guidelines positively, given that they were participating in a WHO field study about the new global classification system in which their institutions were specifically involved. There may have been both a positive cognitive bias and a social desirability element to their responses. It is possible that clinicians not participating in this type of study will greet the ICD-11 guidelines with less enthusiasm when asked to implement them within their clinical settings. However, this would be true of any parallel assessment of clinical utility such as those for ICD-10[8,9,18] and DSM-5[19], and does not change the overall interpretation of the results.

The pattern of results related to the usefulness of guidelines for specific functions (e.g., treatment selection, prognosis, communicating with other professionals) is entirely consistent with the pattern of results from a separate survey regarding clinicians' current use of the ICD-10, DSM-IV, and DSM-5[10]. It is expected that ratings of the utility of treatment selection and prognosis might not be as high as other uses of the ICD-11, as many treatments are not specific to a single diagnostic label[20], nor is the ICD-11 intended to be a treatment guide.

It is nonetheless reassuring that, although following the same pattern, clinicians' ratings of the usefulness of the ICD-11 diagnostic guidelines they had just used for treatment selection, assessing prognosis, and educating patients and families were substantially higher than the ratings clinicians participating in the other study made about the ICD-10 or the DSM-IV or the DSM-5[10]. Even so, this may be an inherent limitation of current categorical classification systems (i.e., ICD-11, ICD-10, and DSM-5), which are not organized around the most meaningful typologies for selecting treatment or establishing prognosis[20,21]. Future efforts at creating a closer link between

Table 3 Clinical utility ratings by country for three core questions

Ease of use

	Not at all	Somewhat	Quite	Extremely	Quite + extremely
Brazil (N=200)	4 (2.0%)	30 (15.0%)	125 (62.5%)	41 (20.5%)	166 (83.0%)
Canada (N=106)	0	19 (17.9%)	71 (67.0%)	16 (15.1%)	87 (82.1%)
PR China (N=405)	3 (0.7%)	62 (15.3%)	306 (75.6%)	34 (8.4%)	340 (84.0%)
India (N=418)	3 (0.7%)	46 (11.0%)	291 (69.6%)	78 (18.7%)	369 (88.3%)
Italy (N=200)	0	13 (6.5%)	125 (62.5%)	62 (31.0%)	187 (93.5%)
Japan (N=336)	13 (3.9%)	161 (47.9%)	147 (43.8%)	15 (4.5%)	162 (48.2%)
Lebanon (N=206)	1 (0.5%)	15 (7.3%)	147 (71.4%)	43 (20.9%)	190 (92.2%)
Mexico (N=306)	1 (0.3%)	25 (8.2%)	213 (69.6%)	67 (21.9%)	280 (91.5%)
Nigeria (N=264)	0	13 (4.9%)	185 (70.1%)	66 (25.0%)	251 (95.1%)
Russian Fed. (N=208)	0	25 (12.0%)	166 (79.8%)	17 (8.2%)	183 (88.0%)
Spain (N=140)	0	3 (2.1%)	133 (95.0%)	4 (28.6%)	137 (97.9%)
South Africa (N=413)	3 (0.7%)	25 (6.1%)	303 (73.4%)	82 (19.9%)	385 (93.2%)
Tunisia (N=406)	4 (1.0%)	119 (29.3%)	259 (63.8%)	24 (5.9%)	283 (69.7%)

Goodness of fit

	Not at all	Somewhat	Quite	Extremely	Quite + extremely
Brazil (N=200)	6 (3.0%)	31 (15.5%)	120 (60.0%)	43 (21.5%)	163 (81.5%)
Canada (N=106)	1 (0.9%)	28 (26.4%)	63 (59.4%)	14 (13.2%)	77 (72.6%)
PR China (N=405)	4 (1.0%)	58 (14.3%)	293 (72.3%)	50 (12.3%)	343 (84.6%)
India (N=418)	3 (0.7%)	49 (11.7%)	293 (70.1%)	73 (17.5%)	366 (87.6%)
Italy (N=200)	0	11 (5.5%)	123 (61.5%)	66 (33.0%)	189 (94.5%)
Japan (N=336)	7 (2.1%)	168 (50.0%)	149 (44.3%)	12 (3.6%)	161 (47.9%)
Lebanon (N=206)	1 (0.5%)	20 (9.7%)	139 (67.5%)	46 (22.3%)	185 (89.8%)
Mexico (N=306)	2 (0.7%)	37 (12.1%)	209 (68.3%)	58 (19.0%)	267 (87.3%)
Nigeria (N=264)	0	22 (8.3%)	195 (73.9%)	47 (17.8%)	242 (91.7%)
Russian Fed. (N=208)	0	28 (13.5%)	162 (77.9%)	18 (8.7%)	180 (86.5%)
Spain (N=140)	0	7 (5.0%)	127 (90.7%)	6 (4.3%)	133 (95.0%)
South Africa (N=413)	2 (0.5%)	27 (6.5%)	360 (87.2%)	24 (5.8%)	384 (93.0%)
Tunisia (N=406)	2 (0.5%)	118 (29.1%)	264 (65.0%)	22 (5.4%)	286 (70.4%)

Clarity and understandability

	Not at all	Somewhat	Quite	Extremely	Quite + extremely
Brazil (N=200)	1 (0.5%)	20 (10.0%)	141 (70.5%)	38 (19.0%)	179 (89.5%)
Canada (N=106)	0	18 (17.0%)	65 (61.3%)	23 (21.7%)	88 (83.0%)
PR China (N=405)	2 (0.5%)	55 (13.6%)	296 (73.1%)	52 (12.8%)	348 (85.9%)
India (N=418)	2 (0.5%)	51 (12.2%)	281 (67.2%)	84 (20.1%)	365 (87.3%)
Italy (N=200)	0	7 (3.5%)	115 (57.5%)	78 (39.0%)	193 (96.5%)
Japan (N=336)	5 (1.5%)	164 (48.8%)	154 (45.8%)	13 (3.9%)	167 (49.7%)
Lebanon (N=206)	0	22 (10.7%)	147 (71.4%)	37 (18.0%)	184 (89.3%)
Mexico (N=306)	1 (0.3%)	25 (8.2%)	214 (69.9%)	66 (21.6%)	280 (91.5%)
Nigeria (N=264)	0	17 (6.4%)	191 (72.3%)	56 (21.2%)	247 (93.6%)
Russian Fed. (N=208)	0	26 (12.5%)	159 (76.4%)	23 (11.1%)	182 (87.5%)
Spain (N=140)	0	6 (4.3%)	127 (90.7%)	7 (5.0%)	134 (95.7%)
South Africa (N=413)	1 (0.2%)	32 (7.7%)	328 (79.4%)	52 (12.6%)	380 (92.1%)
Tunisia (N=406)	2 (0.5%)	124 (30.5%)	255 (62.8%)	25 (6.2%)	280 (69.0%)

Table 4 Concurrent reliability (joint rater agreement, represented by interclass kappa) and bootstrapped 95% confidence interval (CI) for five most common diagnoses by country

	Kappa (95% CI)				
Country	**Schizophrenia**	**Schizoaffective disorder**	**Bipolar type I disorder**	**Single episode depressive disorder**	**Recurrent depressive disorder**
Brazil (N=100)	.61 (.39 to .79)	.45 (.14 to .73)	.85 (.56 to 1.00)	.43 (−.03 to .78)	-
Canada (N=53)	-	-	-	.65 (.30 to .90)	.85 (.68 to .96)
PR China (N=203)	.96 (.92 to .99)	-	.87 (.78 to .95)	.32 (−.02 to .66)	.71 (.55 to .84)
India (N=209)	.90 (.82 to .96)	.59 (−.01 to .91)	.88 (.78 to .96)	.76 (.61 to .87)	.85 (.70 to .97)
Italy (N=100)	.85 (.74 to .96)	.79 (.59 to .93)	.95 (.84 to 1.00)	-	-
Japan (N=168)	.90 (.82 to .97)	-	.77 (.53 to .94)	.77 (.61 to .90)	.75 (.61 to .87)
Lebanon (N=103)	.95 (.86 to 1.00)	.82 (.64 to .95)	.82 (.67 to .93)	-	.64 (.29 to .88)
Mexico (N=153)	.87 (.76 to .96)	.38 (−.02 to .74)	-	.46 (.27 to .62)	.64 (.52 to .76)
Nigeria (N=132)	.93 (.86 to .98)	.71 (.45 to .89)	.83 (.68 to .94)	.93 (.72 to 1.00)	-
Russian Fed. (N=104)	.54 (.33 to .73)	.45 (.20 to .66)	.52 (−.02 to .88)	-	-
South Africa (N=208)	.71 (.60 to .81)	.68 (.55 to .80)	.80 (.71 to .88)	-	.76 (.40 to 1.00)
Spain (N=70)	.84 (.51 to 1.00)	-	.86 (.70 to .97)	.58 (.24 to .84)	.83 (.58 to 1.00)
Tunisia (N=203)	.84 (.75 to .92)	.59 (.30 to .80)	.69 (.52 to .84)	.63 (.41 to .80)	.50 (.24 to .71)
Overall	.87 (.84 to .89)	.66 (.58 to .72)	.84 (.81 to .87)	.64 (.57 to .77)	.74 (.69 to .79)

Cells without values are those with an insufficient number of observations to calculate kappa

Table 5 Clinical utility ratings for three core questions for five most common diagnoses

Ease of use					
	Not at all	**Somewhat**	**Quite**	**Extremely**	**Quite + extremely**
Schizophrenia	4 (0.3%)	127 (10.0%)	896 (70.9%)	237 (18.8%)	1133 (89.6%)
Schizoaffective disorder	0	24 (11.1%)	166 (76.5%)	27 (12.4%)	193 (88.9%)
Bipolar type I disorder	1 (0.2%)	64 (10.8%)	412 (69.8%)	113 (19.2%)	525 (89.0%)
Single episode depressive disorder	1 (0.4%)	56 (21.5%)	165 (63.5%)	38 (14.6%)	203 (78.1%)
Recurrent depressive disorder	4 (0.9%)	78 (18.4%)	290 (68.6%)	51 (12.1%)	341 (80.6%)
Goodness of fit					
	Not at all	**Somewhat**	**Quite**	**Extremely**	**Quite + extremely**
Schizophrenia	3 (0.2%)	141 (11.2%)	897 (71.0%)	223 (17.6%)	1120 (88.6%)
Schizoaffective disorder	0	33 (15.2%)	163 (75.1%)	21 (9.7%)	184 (84.8%)
Bipolar type I disorder	1 (0.2%)	65 (11.0%)	446 (75.6%)	78 (13.2%)	524 (88.8%)
Single episode depressive disorder	1 (0.4%)	58 (22.3%)	173 (66.5%)	29 (11.2%)	202 (77.7%)
Recurrent depressive disorder	3 (0.7%)	81 (19.1%)	284 (67.1%)	55 (13.0%)	339 (80.1%)
Clarity and understandability					
	Not at all	**Somewhat**	**Quite**	**Extremely**	**Quite + extremely**
Schizophrenia	1 (0.1%)	134 (10.6%)	890 (70.4%)	239 (18.9%)	1129 (89.3%)
Schizoaffective disorder	0	26 (12.0%)	161 (74.2%)	30 (13.8%)	191 (88.0%)
Bipolar type I disorder	0	61 (10.3%)	434 (73.6%)	95 (16.1%)	529 (89.7%)
Single episode depressive disorder	0	48 (18.5%)	174 (66.9%)	39 (15.0%)	213 (81.9%)
Recurrent depressive disorder	0	82 (19.4%)	283 (66.9%)	58 (13.7%)	341 (80.6%)

This analysis excluded diagnostic formulations in which more than one of the five index diagnoses included in the table had been assigned (N=853)

mental health diagnosis and treatment planning would be a worthwhile endeavor from the perspective of enhancing public health, but would need to take a variety of other factors into account (e.g., functional status, treatment availability and acceptability).

Looking at country-level clinical utility ratings, it is clear that clinicians' perceptions of the utility of the diagnostic guidelines were similarly positive across a very diverse set of countries: Brazil, Canada, China, India, Italy, Lebanon, Mexico, Nigeria, Russia, Spain, and South Africa. This may reflect the substantial international participation in the development of the guidelines, with all WHO regions represented and a substantial number of experts from low- and middle-income countries included in all ICD-11 Working Groups, as well as prior international multilingual testing via the Global Clinical Practice Network.

It is encouraging that conducting the clinical assessment in a wide range of local languages did not seem to impact the perceived utility of the diagnostic guidelines. The main deviation from this was the substantially lower ratings of clinical utility made by Japanese participants and the somewhat lower (though still positive) ratings made by Tunisian participants. For Japan, it is possible that these differences are partly related to a cultural tendency not to make extreme ratings, either positive or negative[22], and for both countries this may have been affected by the particular characteristics of the clinician raters involved. For Tunisia, not having all of the training materials available in French may have affected the outcome. However, it is also possible that the proposed ICD-11 diagnostic guidelines specifically correspond less well to presentations of mental disorders more characteristic of Japanese and Tunisian patients as compared to patients from other countries. Further research will be necessary to understand more about global variation in the perceived clinical utility of diagnostic guidelines.

It is important to note, however, that the observed variations in perceived clinical utility, either by country or by diagnosis, had no discernible relationship to variations in reliability. In particular, the lower ratings by Japanese participants of clinical utility did not seem to impact their ability to apply the guidelines consistently. Similarly, instances of lower reliability did not result in correspondingly poorer ratings of clinical utility. This finding highlights the importance of taking into account multiple characteristics of the classification system when evaluating its performance. Neither clinical utility ratings nor reliability estimates provide the whole story.

This paper adds to our previous finding that inter-diagnostician reliability using the proposed ICD-11 diagnostic guidelines was moderate to almost perfect (.45 to .88)[15] for mental disorders accounting for the greatest proportion of global disease burden and the highest levels of service utilization among adult patients presenting for treatment at 28 participating centers in 13 countries[14]. Reliability was superior overall to that previously reported for equivalent ICD-10 guidelines.

WHO's model for ICD-11[2] does not consider clinical utility as defined solely by preference ratings. Instead, it is a dynamic construct that is directly integrated with the actual use of the manual as intended. As such, adequate reliability or consistency of application across the globe is also evidence of the clinical utility of the new ICD-11 guidelines.

CONCLUSIONS

The 11th revision of the Mental, Behavioural and Neurodevelopmental Disorders chapter of the ICD has made substantive changes to the conceptualization of many disorders, which may impact their clinical utility, in addition to their reliability and validity. This study is part of a program of field studies focused on clinical utility adopted by WHO in revising the Mental and Behavioural Disorders chapter of ICD-10[3].

In clinical settings, the ICD functions partly as an interface between health encounters and health information[13], and diagnostic guidelines that are experienced by their intended users as lacking in clinical utility have little chance of being implemented faithfully and consistently. In this event, the validity of the diagnostic components of health encounter data would be seriously compromised, with downstream implications for the quality of decision-making regarding health policy and programmes and resource allocation based on those data.

Therefore, field studies that assess perceived clinical utility of the proposed ICD-11 diagnostic guidelines among its intended users have very important implications. For this reason, the study was conducted in a broad spectrum of secondary and tertiary mental health care settings across countries with varied languages, cultures, and resource levels.

Overall, the results provide considerable reason to be optimistic about the perceived clinical utility of the ICD-11 among global clinicians.

ACKNOWLEDGEMENTS

The opinions contained in the paper are those of its authors and, except as specifically stated, are not intended to represent the official policies or positions of the WHO. Funding was received for national activities related to this project in the following countries: Brazil – Conselho Nacional de Desenvolvimento Científico e Tecnológico; Canada – University Medical Research Fund, Royal's University of Ottawa Institute of Mental Health Research; Japan – Japanese Society of Psychiatry and Neurology, and Japan Agency for Medical Research and Development; Mexico – National Council of Science and Technology (project no. 234473). Additional support for data collection in Brazil, Lebanon, Nigeria, South Africa and Tunisia was provided by the Columbia University Global Mental Health Program. Otherwise, this project was funded by in-kind contributions of the participating institutions. The authors express their gratitude to the following individuals who contributed substantially to the conduct of this research: Gustavo M. Barros, Ary Gadelha, Michel Haddad, Nuno H.P. Santos (Brazil); Huajian Ma, Zhen Wang, Jingjing Huang (China); Huma Kamal, Nidhi Malhotra (India); Gaia Sampogna, Lucia Del Gaudio, Giuseppe Piegari, Francesco Perris, Luca Steardo (Italy); Tomofumi Miura, Itta Namamura, Kiyokazu Atake, Ayako Endo, Yuki Kako, Shinichi Kishi, Michihiko Koeda, Shinsuke Kondo, Akeo Kurumaji, Shusuke Numata, Naoya Oribe, Futoshi Suzuki, Masashi Yagi (Japan); Sariah Daouk, Chadia Haddad, François Kazour, Nicole Khauli (Lebanon); Francisco Juárez, Alejandra González, Omar Hernández, Carolina Muñoz (Mexico); Tatiana Kiska, Oleg Limankin, Pavel Ponizovsky (Russian

Federation); Roxanne James, Christine Lochner, Adele Pretorius (South Africa); Carolina Ávila, Cora Fernández; Julián Gómez, Ana Izquierdo, Beatriz Vicario, Rubén Vicente (Spain); Rahma Damak (Tunisia).

REFERENCES

1. World Health Organization. WHO releases new International Classification of Diseases (ICD 11). http://www.who.int/news-room/detail/18-06-2018-who-releases-new-international-classification-of-diseases-(icd-11).
2. Reed GM. Toward ICD-11: improving the clinical utility of WHO's international classification of mental disorders. Prof Psychol Res Pr 2010;41:457-64.
3. Keeley JW, Reed GM, Roberts MC et al. Developing a science of clinical utility in diagnostic classification systems: field study strategies for ICD-11 mental and behavioural disorders. Am Psychol 2016;71:3-16.
4. First MB, Reed GM, Hyman SE et al. The development of the ICD-11 clinical descriptions and diagnostic guidelines for mental and behavioural disorders. World Psychiatry 2015;14:82-90.
5. Reed GM, First MB, Medina-Mora ME et al. Draft diagnostic guidelines for ICD-11 mental and behavioural disorders available for review and comment. World Psychiatry 2016;15:112-3.
6. First MB, Pincus HA, Levine JB et al. Clinical utility as a criterion for revising psychiatric diagnoses. Am J Psychiatry 2004;161:946-54.
7. First MB. Clinical utility in the revision of the Diagnostic and Statistical Manual of Mental Disorders (DSM). Prof Psychol Res Pr 2010;41:465-73.
8. Sartorius N, Kaelber CT, Cooper JE et al. Progress toward achieving a common language in psychiatry. Results from the field trial of the clinical guidelines accompanying the WHO classification of mental and behavioral disorders in ICD-10. Arch Gen Psychiatry 1993;50:115-24.
9. Sartorius N, Ustün TB, Korten A et al. Progress toward achieving a common language in psychiatry, II: Results from the international field trials of the ICD-10 diagnostic criteria for research for mental and behavioral disorders. Am J Psychiatry 1995;152:1427-37.
10. First MB, Rebello TJ, Keeley JW et al. Do mental health professionals use diagnostic classifications the way we think they do? A global survey. World Psychiatry 2018;17:187-95.
11. World Health Organization. The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. Geneva: World Health Organization, 1992.
12. Reed GM, Roberts MC, Keeley J et al. Mental health professionals' natural taxonomies of mental disorders: implications for the clinical utility of the ICD-11 and the DSM-5. J Clin Psychol 2013;69:1191-212.
13. Roberts MC, Reed GM, Medina-Mora ME et al. A global clinicians' map of mental disorders to improve ICD-11: analysing meta-structure to enhance clinical utility. Int Rev Psychiatry 2012;24:578-90.
14. Reed GM, Sharan P, Rebello TJ et al. The ICD-11 developmental field study of reliability of diagnoses of high-burden mental disorders: results among adult patients in mental health settings of 13 countries. World Psychiatry 2018;17:174-86.
15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
16. Keeley JW, Reed GM, Roberts MC et al. Disorders specifically associated with stress: a case-controlled field study for ICD-11 mental and behavioural disorders. Int J Clin Health Psychol 2016;16:109-27.
17. Keeley JW, Gaebel W, First MB et al. Psychotic disorder symptom rating scales: are dichotomous or multi-point scales more clinically useful? An ICD-11 field study. J Schizophr Res (in press).
18. Regier DA, Kaelber CT, Roper MT et al. The ICD-10 clinical field trial for mental and behavioral disorders: results in Canada and the United States. Am J Psychiatry 1994;151:1340-50.
19. Mościcki E, Clarke DE, Kuramoto SJ et al. Testing DSM-5 in routine clinical practice settings: feasibility and clinical utility. Psychiatr Serv 2013;64:952-60.
20. Maj M. Why the clinical utility of diagnostic categories in psychiatry is intrinsically limited and how we can use new approaches to complement them. World Psychiatry 2018;17:121-2.
21. Clark LA, Cuthbert B, Lewis-Fernandez R et al. ICD-11, DSM-5, and RDoC: three approaches to understanding and classifying mental disorder. Psychol Sci Public Interest 2017;18:72-145.
22. Dolnicar S, Grun B. Cross-cultural differences in survey response patterns. Int Market Rev 2007;24:127-43.

DOI:10.1002/wps.20581

Psychological interventions to reduce positive symptoms in schizophrenia: systematic review and network meta-analysis

Irene Bighelli[1], Georgia Salanti[2], Maximilian Huhn[1], Johannes Schneider-Thoma[1], Marc Krause[1], Cornelia Reitmeir[1], Sofia Wallis[1], Felicitas Schwermann[1], Gabi Pitschel-Walz[1], Corrado Barbui[3], Toshi A. Furukawa[4], Stefan Leucht[1]

[1]Department of Psychiatry and Psychotherapy, Klinikum rechts der Isar, Technische Universität München, Munich, Germany; [2]Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland; [3]Department of Neuroscience, Biomedicine and Movement Sciences, Section of Psychiatry, University of Verona, Verona, Italy; [4]Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine, Kyoto, Japan and School of Public Health, Japan

Psychological treatments are increasingly regarded as useful interventions for schizophrenia. However, a comprehensive evaluation of the available evidence is lacking and the benefit of psychological interventions for patients with current positive symptoms is still debated. The present study aimed to evaluate the efficacy, acceptability and tolerability of psychological treatments for positive symptoms of schizophrenia by applying a network meta-analysis approach, that can integrate direct and indirect comparisons. We searched EMBASE, MEDLINE, PsycINFO, PubMed, BIOSIS, Cochrane Library, World Health Organization's International Clinical Trials Registry Platform and ClinicalTrials.gov for randomized controlled trials of psychological treatments for positive symptoms of schizophrenia, published up to January 10, 2018. We included studies on adults with a diagnosis of schizophrenia or a related disorder presenting positive symptoms. The primary outcome was change in positive symptoms measured with validated rating scales. We included 53 randomized controlled trials of seven psychological interventions, for a total of 4,068 participants receiving the psychological treatment as add-on to antipsychotics. On average, patients were moderately ill at baseline. The network meta-analysis showed that cognitive behavioural therapy (40 studies) reduced positive symptoms more than inactive control (standardized mean difference, SMD=–0.29; 95% CI: –0.55 to –0.03), treatment as usual (SMD=–0.30; 95% CI: –0.45 to –0.14) and supportive therapy (SMD=–0.47; 95% CI: –0.91 to –0.03). Cognitive behavioural therapy was associated with a higher dropout rate compared with treatment as usual (risk ratio, RR=0.74; 95% CI: 0.58 to 0.95). Confidence in the estimates ranged from moderate to very low. The other treatments contributed to the network with a lower number of studies. Results were overall consistent in sensitivity analyses controlling for several factors, including the role of researchers' allegiance and blinding of outcome assessor. Cognitive behavior therapy seems to be effective on positive symptoms in moderately ill patients with schizophrenia, with effect sizes in the lower to medium range, depending on the control condition.

Key words: Schizophrenia, positive symptoms, psychological interventions, cognitive behavioural therapy, network meta-analysis

(World Psychiatry 2018;17:316–329)

Psychological interventions for schizophrenia have been developed to address many aspects of the disorder and, according to guidelines from the National Institute for Health and Care Excellence (NICE)[1] in the UK and the Schizophrenia Patient Outcomes Research Team (PORT)[2] in the US, are regarded as useful interventions.

A number of systematic reviews of randomized studies have been conducted on these treatments[3]. However, findings are unclear and often contradictory. For example, while some reviews[4,5] have found a superiority of cognitive behavioural therapy (CBT) compared to usual care, other authors could not replicate this finding when non-blinded randomized controlled trials (RCTs) were excluded[6]. A Cochrane review found CBT to be effective in the long term, but not in the short or medium term[7], while another meta-analysis did not find a benefit for CBT[8].

Moreover, the current evidence presents several shortcomings. First, all the existing reviews have compared two interventions at a time using pairwise meta-analysis. This method summarizes results only when two treatments have already been compared in existing studies, leaving open questions for all the other possible comparisons. Even in the review by Turner et al[9], which included only studies comparing two "active psychological interventions", pairwise meta-analysis was applied to compare each intervention with the pooled others, again not providing information on the comparisons that were not already considered in a trial.

Furthermore, the existing reviews have included heterogeneous samples, pooling patients with different sets of symptoms. No review focused specifically on patients with current positive symptoms, which are – at least in the acute phase – at the core of the disorder. Also the review by Zimmermann et al[5], aiming at evaluating the effect of CBT on positive symptoms, did not restrict its selection to studies on patients presenting these symptoms.

As a result of these limitations in the current evidence, it is still unclear whether there are efficacious and acceptable psychological interventions for treating positive symptoms in schizophrenia.

The aim of the present study was to overcome these limitations by conducting a network meta-analysis, which integrates direct and indirect comparisons of interventions[10], and informs about differences between treatments, even when direct comparisons are not available. Such a meta-analysis requires a certain degree of homogeneity in the population, settings and methods across the studies. A careful definition of the target population of the intervention is therefore essential in order to produce information that is useful for clinical practice.

Our network meta-analysis covered psychological interventions addressing positive symptoms of schizophrenia, in pa-

tients currently experiencing such symptoms, in order to generate results that will be relevant for this specific population.

METHODS

Study design and participants

The detailed methodology for this systematic review and network meta-analysis is described in the study protocol, that was registered *a priori* at PROSPERO (no. CRD42017067795) and published[3]. In reporting results, we followed the PRISMA extension statement for network meta-analyses[11,12].

We included studies in adult individuals with a diagnosis of schizophrenia or a related disorder (such as schizophreniform or schizoaffective disorder), presenting active positive symptoms, or in the phase of acute exacerbation, as defined by inclusion criteria of the trial, without restrictions on setting, gender or ethnicity. We optimized homogeneity of studies within and across treatment comparisons by excluding studies on patients with predominant negative symptoms or concomitant medical or psychiatric illness, and patients in their first psychotic episode or at risk of psychosis. Studies were included if at least 80% of the patients had schizophrenia or related disorders. In case of a mixed population, data about patients with schizophrenia were extracted, if available. We included the trials irrespective of the diagnostic criteria used.

Interventions and comparators

As defined *a priori* in our protocol[3], interventions were any psychological treatments that occur through an interaction between therapist and patient, either face-to-face individually or in group, with the primary aim to reduce positive symptoms.

Comparators were classified as follows: a) interventions (e.g., cognitive remediation, psychoeducation) with a primary target different from improving positive symptoms (e.g., cognition, knowledge of the illness, adherence to medication, functioning), which were primarily analyzed as separate nodes, then combined in a sensitivity analysis; b) inactive controls, defined as interventions intended to control for non-specific aspects of the therapy (befriending, recreation and support, social activity therapy, supportive counselling), also sometimes referred to as "psychological placebos"; c) treatment as usual (i.e., patients continue to receive standard psychiatric care); d) waiting list.

Outcomes

The primary outcome was the change in positive symptoms of schizophrenia, as measured by a rating scale such as the positive subscale of the Positive and Negative Syndrome Scale (PANSS)[13], the positive subscale of the Brief Psychiatric Rating Scale (BPRS)[14], or any other published scale.

Secondary outcomes were: study dropout for any reason (all-cause discontinuation), effects on overall symptoms of schizophrenia, effects on negative symptoms, response (as defined in the study), relapse (operationalized by rating scales or, if not available, rehospitalization due to psychopathology), adherence and insight, changes in depressive symptoms, quality of life, functioning, adverse events that might be related to psychological treatment (according to Linden et al[15]), and mortality (measured as death for any reason, death due to natural causes, death due to suicide). All outcomes were measured at study endpoint, as defined in each study.

Search strategy and selection criteria

We searched EMBASE, MEDLINE, PsycINFO, PubMed, BIOSIS, Cochrane Library, World Health Organization's International Clinical Trials Registry Platform and ClinicalTrials.gov for RCTs published up to January 10, 2018, comparing psychological interventions with each other or with a non-pharmacological control condition in people with schizophrenia who presented active positive symptoms. Additionally, we searched the reference lists of previous reviews.

We applied no language restrictions, with the exception that we did not search Chinese databases. We contacted authors of included studies published in the last 30 years for missing or additional information about their studies.

Data extraction and risk of bias assessment

All abstracts identified by the search were reviewed independently by two researchers of the group. Disagreements were resolved by discussion, and in case of doubts the full paper was retrieved for further inspection. Full reports were obtained for all eligible papers, and again assessed by two independent reviewers. Disagreements were discussed with the senior author and, in case of need, study authors were contacted for further information.

Two researchers independently extracted data from the selected studies, considering main reports and supplementary materials, entered the relevant information into a Microsoft Access database especially created for this study, and assessed risk of bias using the Cochrane risk of bias tool[16]. The following domains of possible bias were considered: sequence generation, allocation concealment, blinding of participants, blinding of outcome assessors, incomplete outcome data, selective reporting, researchers' allegiance[17,18], other bias. We also made a global risk of bias rating for each study based on criteria applied in a network meta-analysis of antidepressants[19].

Statistical analysis

We performed random effects pairwise meta-analyses and network meta-analysis in a frequentist framework using the

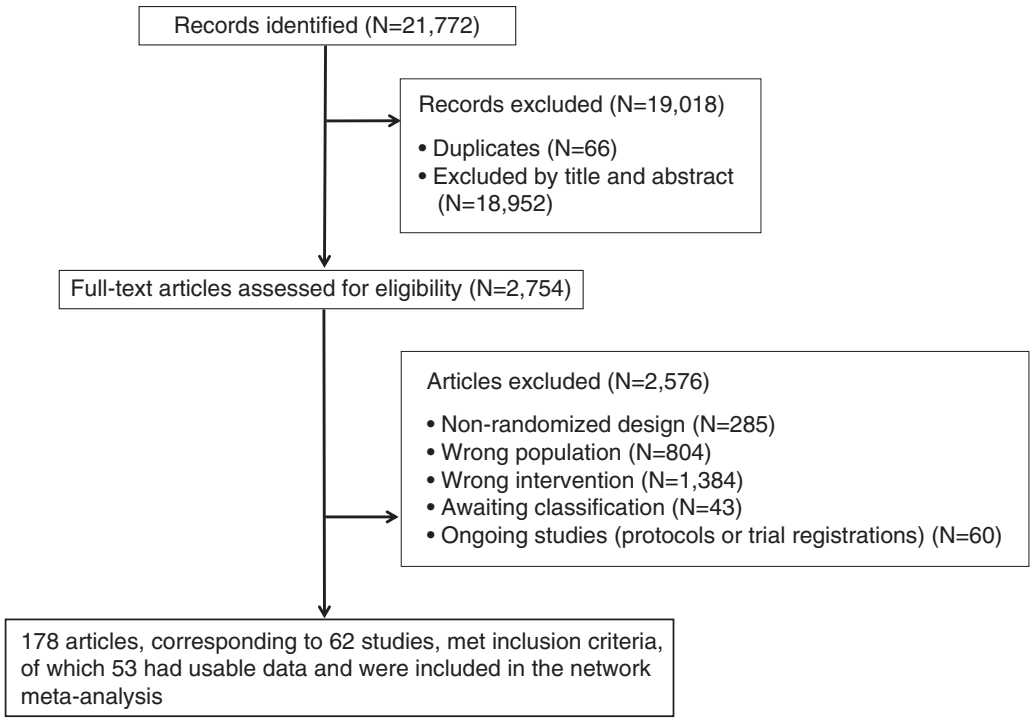

Figure 1 PRISMA flow chart of the study selection process

netmeta package in R (version 3.4.3)[20,21]. We calculated standardized mean differences (SMDs) for continuous outcomes, and risk ratios (RRs) for binary outcomes, both presented with their 95% confidence intervals (CIs). We also calculated the relative ranking for each intervention using the Surface Under the Cumulative Ranking curve (SUCRA), estimated within the frequentist framework (as P scores)[22].

Before running the network meta-analysis, we attempted to assess the transitivity assumption. This assumption implies that studies comparing different sets of interventions are sufficiently similar to provide valid indirect inferences, which we tried to ensure by applying narrow inclusion criteria and making populations as similar as possible within and across treatment comparisons. We also considered whether the potential effect modifiers (listed below) were distributed similarly across the available direct comparisons.

We assumed a common heterogeneity parameter across the various treatment comparisons, and presented the between study variance (tau^2) for each outcome. We characterized the amount of heterogeneity as low, moderate or high, using the first and third quantiles of their empirical distributions[23]. Statistical inconsistency was evaluated separating direct evidence from indirect evidence provided by the entire network, and then testing the agreement of these two pieces of evidence[24]. The magnitude of inconsistency factors (the difference in direct and indirect SMD) and their respective p values were used to identify the presence of inconsistency. We also applied the design-by-treatment interaction model, that evaluates inconsistency in the network jointly[25].

To explore potential sources of heterogeneity or inconsistency, we planned *a priori* subgroup analyses for the primary outcome on the following potential effect modifiers: number of sessions, study duration, setting (individual vs. group), expertise of the therapist, baseline severity. Sensitivity analyses were performed excluding open label studies, studies that presented only completer analyses, studies at overall high risk of bias[19], studies with high risk of researchers' allegiance, studies focused on treatment-resistant patients, and studies with a non-active comparison group. We also assessed small trial effects (potentially associated with publication bias) by examining funnel plots of pairwise meta-analyses and comparison-adjusted funnel plots, if ten or more studies were included[26]. Additionally, we assessed the confidence in estimates of the main outcome with Confidence in Network Meta-Analysis (CINeMA), an adaptation of the Grading of Recommendations Assessment, Development and Evaluation framework (GRADE) specifically developed for network meta-analysis[27].

RESULTS

Characteristics of included studies

21,772 references were identified by the search (last update January 10, 2018), and 2,754 articles were retrieved in full text (Figure 1). We included 62 randomized controlled trials, of which 53 had usable data and were included in the network meta-analysis (involving 4,068 participants) (Table 1).

Table 1 Characteristics of studies

Study	Country	Treatments (N. patients)	Trial duration (weeks)	N. sessions	Diagnosis	Study design	Risk of bias (overall)
Barrowclough et al[28]	UK	Cognitive behavioural therapy (N=57), TAU (N=56)	26	10.4	Schizophrenia or schizoaffective disorder (DSM-IV)	SB	Moderate
Bechdolf et al[29]	Germany	Cognitive behavioural therapy (N=40), psychoeducation (N=48)	8	11.9 (cognitive behavioural therapy), 6.4 (psychoeducation)	Episode of a schizophrenic or related disorder (ICD-10)	SB	High
Birchwood et al[30]	UK	Cognitive behavioural therapy (N=98), TAU (N=99)	39	19	Schizophrenia or schizoaffective disorder (ICD-10)	SB	Moderate
Drury et al[31]	UK	Cognitive therapy (N=30), recreation and support (N=32)	12	NA	Functional psychosis (DSM-IV)	OL	High
Durham et al[32]	UK	Cognitive behavioural therapy (N=22), supportive therapy (N=23), TAU (N=21)	39	20	Schizophrenia, schizoaffective disorder or delusional disorder (ICD-10 and DSM-IV)	SB	High
England[33]	NA (author's affiliation in Canada)	Cognitive nursing intervention (N=44), TAU (N=21)	18	12	Schizophrenia or schizoaffective disorder (DSM-IV)	SB	Moderate
Foster et al[34]	UK	Cognitive behavioural therapy (N=12), TAU (N=12)	4	4	Schizophrenia, schizoaffective disorder or delusional disorder (clinical diagnosis)	OL	High
Freeman et al[35]	UK	Cognitive behavioural therapy (N=15), TAU (N=15)	8	6	Schizophrenia, schizoaffective disorder or delusional disorder (clinical diagnosis)	SB	Low
Freeman et al[36]	UK	Cognitive behavioural therapy (N=73), TAU (N=77)	8	5.5	Schizophrenia, schizoaffective disorder or delusional disorder (clinical diagnosis)	SB	High
Freeman et al[37]	UK	Cognitive behavioural therapy (N=24), TAU (N=26)	12	7.3	Schizophrenia, schizoaffective disorder or delusional disorder (clinical diagnosis)	SB	Moderate
Garety et al[38]	UK	Cognitive behavioural therapy (N=27), family intervention (N=28), TAU (N=28)	39	13.9	Non-affective psychosis (DSM-IV and ICD–10)	SB	Moderate
Garety et al[38]	UK	Cognitive behavioural therapy (N=106), TAU (N=112)	39	14.3	Non-affective psychosis (DSM-IV and ICD–10)	SB	Moderate
Gottlieb et al[39]	US	Cognitive behavioural therapy (N=19), TAU (N=18)	24	10	Schizophrenia, schizoaffective disorder or psychosis not otherwise specified (NA)	SB	Moderate
Habib et al[40]	Pakistan	Cognitive behavioural therapy (N=21), TAU (N=21)	21	13	Schizophrenia (DSM-IV-TR)	SB	High
Haddock et al[41]	UK	Cognitive behavioural therapy (N=10), supportive counselling (N=11)	5	10.2	Schizophrenia or schizoaffective disorder (DSM-IV)	SB	Moderate

Table 1 Characteristics of studies (*continued*)

Study	Country	Treatments (N. patients)	Trial duration (weeks)	N. sessions	Diagnosis	Study design	Risk of bias (overall)
Haddock et al[42]	UK	Cognitive behavioural therapy (N=38), social activity therapy (N=39)	26	17 (cognitive behavioural therapy), 17.4 (social activity therapy)	Schizophrenia or schizoaffective disorder (DSM-IV)	SB	Moderate
Hazell et al[43]	UK	Cognitive behavioural therapy (N=15), waitlist (N=15)	12	8	Schizophrenia and related disorders (NA)	SB	Moderate
Krakvik et al[44]	Norway	Cognitive behavioural therapy (N=23), waitlist (N=22)	26	20	Schizophrenia, schizoaffective disorder or persistent delusional disorder (ICD-10)	OL	Moderate
Kuipers et al[45]	UK	Cognitive behavioural therapy (N=28), TAU (N=32)	39	18.6	Paranoid schizophrenia (DSM-III-R)	OL	High
Lecomte et al[46]	Canada	Cognitive behavioural therapy (N=48), social skills training (N=54), waitlist (N=27)	13	24	Schizophrenia spectrum disorder (NA)	SB	Moderate
Lee et al[47]	South Korea	Cognitive behavioural social skills training (N=12), TAU (N=13)	7	12	Schizophrenia (DSM-IV-TR)	SB	Moderate
Lee et al[48]	South Korea	Cognitive behavioural therapy (N=25), supportive therapy (N=25)	32	20.1	Schizophrenia (DSM-IV)	SB	Moderate
Levine et al[49]	NA (author's affiliation in Israel)	Cognitive therapy (N=6), supportive therapy (N=6)	6	6	Paranoid schizophrenia (DSM-III-R)	NA	High
Li et al[50]	China	Cognitive behavioural therapy (N=96), supportive therapy (N=96)	24	15	Schizophrenia (DSM-IV)	SB	Moderate
McLeod et al[51]	UK	Cognitive behavioural therapy (N=10), waitlist (N=10)	12	8	Schizophrenia (DSM-IV)	NA	High
Morrison et al[52]	UK	Cognitive therapy (N=37), TAU (N=37)	39	13.3	Schizophrenia, schizoaffective disorder or delusional disorder (ICD-10 or PANSS)	SB	Moderate
Penn et al[53]	US	Cognitive behavioural therapy (N=32), supportive therapy (N=33)	12	8.3	Schizophrenia or schizoaffective disorder (DSM-IV)	SB	Low
Pinninti et al[54]	US	Cognitive behavioural therapy (N=18), TAU (N=15)	12	11.93	Schizophrenia or schizoaffective disorder (DSM-IV)	SB	Moderate
Pot-Kolder et al[55]	The Netherlands	Virtual reality based cognitive behavioural therapy (N=58), waitlist (N=58)	12	16	Psychotic disorder (DSM-IV)	SB	Low
Rector et al[56]	Canada	Cognitive behavioural therapy (N=24), TAU (N=21)	26	20	Schizophrenia or schizoaffective disorder (DSM-IV)	SB	Moderate

Table 1 Characteristics of studies (*continued*)

Study	Country	Treatments (N. patients)	Trial duration (weeks)	N. sessions	Diagnosis	Study design	Risk of bias (overall)
Sensky et al[57]	UK	Cognitive behavioural therapy (N=46), befriending (N=44)	39	19	Schizophrenia (ICD-10 Research Criteria and DSM-IV)	SB	Moderate
Startup et al[58]	UK	Cognitive behavioural therapy (N=47), TAU (N=43)	26	12.9	Schizophrenia, schizophreniform disorder or schizoaffective disorder (DSM-IV)	OL	High
Tarrier et al[59]	UK	Cognitive behavioural therapy (N=33), supportive Counselling (N=26), TAU (N=28)	10	20	Schizophrenia, schizoaffective psychosis or delusional disorder (DSM-III-R)	SB	Moderate
Trower et al[60]	UK	Cognitive behavioural therapy (N=18), TAU (N=20)	26	16	Schizophrenia or related disorder (ICD-10)	SB	High
Turkington et al[61]	UK	Cognitive behavioural therapy (N=13), befriending (N=6)	8	6	Schizophrenia (ICD-10 Research Criteria)	SB	Moderate
Valmaggia et al[62]	The Netherlands, Belgium	Cognitive behavioural therapy (N=36), supportive counselling (N=26)	23	16	Schizophrenia (DSM-IV)	SB	Moderate
van der Gaag et al[63]	The Netherlands	Cognitive behavioural therapy (N=110), TAU (N=106)	26	13	Schizophrenia or schizoaffective disorder (DSM-IV-TR)	SB	High
Velligan et al[64]	US	Cognitive behavioural therapy (N=43), cognitive adaptation training (N=41), cognitive behavioural therapy + cognitive adaptation training (N=40), TAU (N=42)	39	26.6 (cognitive behavioural therapy), 27.5 (cognitive adaptation training), 27.5 (cognitive behavioural therapy + cognitive adaptation training)	Schizophrenia or schizoaffective disorder (DSM-IV)	SB	High
Wahass & Kent[65]	Saudi Arabia	Cognitive behavioural therapy (N=3), TAU (N=3)	9	25	Schizophrenia (ICD-10)	OL	Moderate
Wittorf et al[66]	Germany	Cognitive behavioural therapy (N=50), supportive therapy (N=50)	33	20	Schizophrenia, schizophreniform disorder, schizoaffective disorder or delusional disorder (DSM-IV)	SB	High
Wykes et al[67]	UK	Cognitive behavioural therapy (N=45), TAU (N=40)	10	7	Schizophrenia (DSM-IV)	OL	High
ACTRN12616000976482[68]	Australia	Metacognitive training (N=28), cognitive remediation (N=28)	4	4	Schizophrenia spectrum disorder (DSM-V)	SB	Moderate
Briki et al[69]	France	Metacognitive training (N=35), supportive therapy (N=33)	8	14.6	Schizophrenia or schizoaffective disorders (DSM-IV-TR)	SB	High
Favrod et al[70]	Switzerland	Metacognitive training (N=26), TAU (N=26)	8	7	Schizophrenia spectrum disorder (ICD-10)	SB	Moderate
Kumar et al[71]	India	Metacognitive training (N=8), TAU (N=8)	4	8	Paranoid schizophrenia (ICD-10)	NA	High

Table 1 Characteristics of studies (*continued*)

Study	Country	Treatments (N. patients)	Trial duration (weeks)	N. sessions	Diagnosis	Study design	Risk of bias (overall)
So et al[72]	Hong Kong	Metacognitive training (N=23), waitlist (N=21)	4	3.15	Schizophrenia spectrum disorder (clinical diagnosis)	SB	Moderate
van Oosterhout et al[73]	The Netherlands	Metacognitive training (N=75), TAU (N=79)	8	8	Psychotic disorder in the DSM-IV schizophrenia spectrum (DSM-IV-TR)	SB	Moderate
Chadwick et al[74]	UK	Mindfulness (N=11), waitlist (N=11)	10	10	Psychotic disorder (NA)	OL	High
Chadwick et al[75]	UK	Mindfulness (N=54), TAU (N=54)	16	12	Schizophrenia or schizoaffective disorder (ICD-10)	SB	Low
Bach & Hayes[76]	US	Acceptance and Commitment therapy (N=40), TAU (N=40)	16	4	Auditory hallucinations or delusions (clinical diagnosis) (81.25% diagnosed with schizophrenia, schizoaffective disorder or delusional disorder)	OL	High
Shawyer et al[77]	Australia	Acceptance and commitment therapy (N=49), befriending (N=47)	13	7	Schizophrenia or schizoaffective disorder (DSM-IV-TR)	SB	Low
Schnackenberg et al[78]	Germany	Experienced focused counselling (N=12), TAU (N=10)	44	NA	Schizophrenia and schizoaffective disorder (NA)	OL	High
Jenner et al[79]	The Netherlands	Hallucination focused integrative treatment (N=39), TAU (N=39)	39	11	Non-affective psychosis, including schizophrenia, schizoaffective or psychotic disorder not otherwise specified (DSM-IV)	OL	High
Craig et al[80]	UK	AVATAR therapy (N=75), supportive counselling (N=75)	12	5.6 (AVATAR therapy), 5.1 (supportive counselling)	Schizophrenia spectrum disorder or affective disorder with psychotic symptoms (ICD-10)	SB	Low

TAU – treatment as usual, OL – open label, SB – single blind, NA – not available, PANSS – Positive and Negative Syndrome Scale

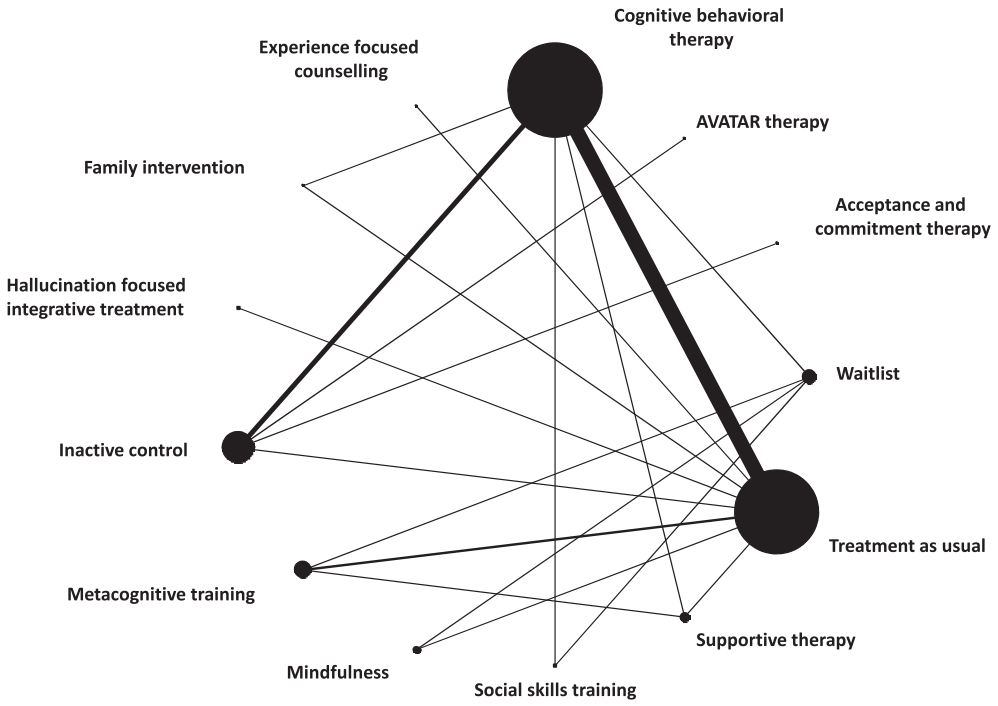

Figure 2 Network meta-analysis of eligible comparisons for positive symptoms. Line width is proportional to the number of trials comparing every pair of treatments. Node size is proportional to the number of studies providing data for each treatment.

These trials provided comparisons of the following psychological treatments: CBT (N=40)[28-67], metacognitive training (N=6)[68-73], mindfulness (N=2)[74,75], acceptance and commitment therapy (N=2)[76,77], experience focused counselling (N=1)[78], hallucination focused integrative treatment (N=1)[79], and AVATAR therapy (N=1)[80].

The mean sample size was 76.5 participants (range 6-218), and the median trial duration was 13 weeks (range 4-44 weeks). Of 3,941 participants whose gender was reported, 2,361 were men (59.9%). The mean duration of illness was 12.4 years, and the mean age of participants was 37.4 years. Nine studies included only inpatients, 15 only outpatients and 14 both, while 15 did not provide information on patients' status. On average, patients had moderate schizophrenic symptoms, with a mean reported PANSS baseline score of 68.26[81,82]. Thanks to collaboration of the authors, we were able to include unpublished data for some studies[36,37,41-43,57,61,68,72].

Risk of bias assessment

Six, 27 and 21 of the included studies were considered to be at low, moderate and high overall risk of bias, respectively (see Table 1). The risk of bias was low in 26 studies (50%) concerning random sequence generation; in 13 studies (25%) concerning allocation concealment; in no study concerning blinding of participants and personnel; in 18 studies (34.6%) concerning blinding of outcome assessment; in seven studies (13.5%) concerning attrition bias; in 11 studies (21.1%) concerning selective reporting; in six studies (11.5%) concerning researchers' alliance; and in 41 studies (78.8%) concerning other bias.

Primary outcome: positive symptoms

Figure 2 shows the network of treatments for the primary outcome. Two studies were not considered in the analyses, because they were not connected to the rest of the network, contributing neither direct nor indirect evidence[29,68].

Network meta-analysis results show that, for the primary outcome, CBT was associated with a higher decrease in positive symptoms than inactive control (SMD=–0.29; 95% CI: –0.55 to –0.03, seven RCTs contributing direct evidence to the network meta-analysis, low confidence in the estimates), treatment as usual (SMD=–0.30; 95% CI: –0.45 to –0.14, 18 RCTs contributing direct evidence, moderate confidence in the estimates) and supportive therapy (SMD=–0.47; 95% CI: –0.91 to –0.03, two RCTs contributing direct evidence, low confidence in the estimates). The difference was not significant for the comparison with waitlist (SMD=–0.24; 95% CI: –0.65 to 0.16), but only two small trials (with 30 and 45 participants respectively[43,44]) contributed direct evidence to this comparison (Figure 3).

One study on hallucination focused integrative treatment showed a decrease in symptoms in comparison to treatment as usual and supportive therapy (moderate and low confi-

HFIT	NMA 0.60 (0.10; 3.58)	NMA 0.53 (0.09; 3.01)	NMA 0.48 (0.08; 2.97)	NMA 0.48 (0.07; 3.20)	NMA 0.40 (0.08; 1.94); PWA 0.40 (0.08; 1.94)	NMA 0.39 (0.07; 2.29)	NMA 0.32 (0.06; 1.88)	NMA 0.31 (0.05; 1.82)	NMA 0.30 (0.06; 1.47)	NMA 0.28 (0.05; 1.39)	NMA 0.27 (0.05; 1.40)	**NMA 0.11 (0.02; 0.66)**
HFIT	MF	NMA 0.88 (0.30; 2.61)	NMA 0.80 (0.23; 2.79)	NMA 0.80 (0.21; 3.09)	NMA 0.67 (0.29; 1.56); PWA 0.67 (0.25; 1.74)	NMA 0.66 (0.24; 1.79); PWA 0.67 (0.14; 3.24)	NMA 0.54 (0.17; 1.70)	NMA 0.51 (0.16; 1.67)	NMA 0.50 (0.21; 1.19)	NMA 0.46 (0.19; 1.15)	NMA 0.45 (0.17; 1.19)	**NMA 0.19 (0.07; 0.56)**
NMA -0.31 (-1.23; 0.61)	AVATAR	MT	NMA 0.91 (0.28; 2.93)	NMA 0.91 (0.25; 3.27)	NMA 0.76 (0.36; 1.58); PWA 0.78 (0.34; 1.78)	NMA 0.75 (0.30; 1.85); PWA 0.68 (0.17; 2.71)	NMA 0.61 (0.21; 1.78)	NMA 0.58 (0.19; 1.75)	NMA 0.56 (0.26; 1.20)	NMA 0.52 (0.23; 1.17)	NMA 0.51 (0.21; 1.23)	**NMA 0.22 (0.08; 0.58)**
NMA -0.40 (-1·07; 0·28)	NMA -0.09 (-0.71; 0.54)	CBT	EFT	NMA 1.00 (0.25; 4.03)	NMA 0.83 (0.33; 2.08); PWA 0.83 (0.33; 2.08)	NMA 0.82 (0.25; 2.72)	NMA 0.67 (0.20; 2.24)	NMA 0.64 (0.19; 2.19)	NMA 0.62 (0.24; 1.59)	NMA 0.57 (0.22; 1.53)	NMA 0.56 (0.20; 1.58)	**NMA 0.24 (0.07; 0.78)**
NMA -0.47 (-1.21; 0.28)	NMA -0.15 (-0.88; 0.57)	NMA -0.07 (-0.44; 0.30)	MT	FI	NMA 0.84 (0.29; 2.41); PWA 0.80 (0.24; 2.67)	NMA 0.82 (0.22; 3.02)	NMA 0.67 (0.18; 2.46)	NMA 0.64 (0.17; 2.43)	NMA 0.62 (0.22; 1.78); PWA 0.64 (0.20; 2.03)	NMA 0.58 (0.19; 1.72)	NMA 0.56 (0.18; 1.76)	**NMA 0.24 (0.07; 0.86)**
NMA -0.46 (-1.86; 0.94)	NMA -0.15 (-1.54; 1.25)	NMA -0.06 (-1.31; 1.19)	NMA 0.01 (-1.28; 1.29)	EFC	TAU	NMA 0.98 (0.45; 2.14)	NMA 0.80 (0.37; 1.77)	NMA 0.77 (0.34; 1.75); PWA 1.00 (0.31; 3.19)	**NMA 0.74 (0.58; 0.95); PWA 0.76 (0.58; 0.98)**	**NMA 0.69 (0.48; 0.99); PWA 0.49 (0.27; 0.91)**	NMA 0.67 (0.40; 1.11); PWA 0.44 (0.09; 2.02)	**NMA 0.29 (0.13; 0.62)**
NMA -0.55 (-1.46; 0.36)	NMA -0.24 (-1.13; 0.65)	NMA -0.15 (-0.79; 0.48); PWA -0.11 (-0.84; 0.63)	NMA -0.09 (-0.81; 0.64)	NMA -0.09 (-1.48; 1.29)	FI	WL	NMA 0.82 (0.28; 2.41)	NMA 0.78 (0.25; 2.39)	NMA 0.75 (0.35; 1.63); PWA 0.73 (0.27; 1.95)	NMA 0.70 (0.31; 1.59)	NMA 0.68 (0.28; 1.66)	**NMA 0.29 (0.13; 0.65); PWA 0.38 (0.15; 1.00)**
NMA -0.57 (-1.41; 0.28)	NMA -0.25 (-1.08; 0.57)	NMA -0.17 (-0.71; 0.38)	NMA -0.10 (-0.72; 0.52)	NMA -0.11 (-1.45; 1.24)	NMA -0.02 (-0.84; 0.81)	MF	AVATAR	NMA 0.95 (0.32; 2.81)	NMA 0.92 (0.43; 1.97)	NMA 0.86 (0.43; 1.73); PWA 0.86 (0.43; 1.73)	NMA 0.83 (0.34; 2.02)	NMA 0.36 (0.12; 1.03)
NMA -0.60 (-1.47; 0.27)	NMA -0.29 (-1.12; 0.54)	NMA -0.21 (-0.75; 0.34); PWA -0.10 (-0.69; 0.49)	NMA -0.14 (-0.77; 0.50)	NMA -0.15 (-1.50; 1.21)	NMA -0.05 (-0.89; 0.78)	NMA -0.04 (-0.78; 0.71)	SST	ACT	NMA 0.97 (0.42; 2.22)	NMA 0.90 (0.39; 205); PWA 1.15 (0.38; 3.52)	NMA 0.87 (0.34; 2.24)	NMA 0.37 (0.12; 1.13)
NMA -0.64 (-1.42; 0.14)	NMA -0.33 (-1.07; 0.41)	NMA -0.24 (-0.65; 0.16); PWA -0.40 (-0.90; 0.09)	NMA -0.18 (-0.64; 0.29); PWA 0.28 (-0.46; 1.02)	NMA -0.18 (-1.49; 1.12)	NMA -0.09 (-0.84; 0.66)	NMA -0.07 (-0.68; 0.53); PWA -0.26 (-1.29; 0.77)	NMA -0.04 (-0.61; 0.53); PWA 0.10 (-0.54; 0.74)	WL	CBT	NMA 0.93 (0.69; 1.25); PWA 0.94 (0.69; 1.27)	NMA 0.90 (0.57; 1.42); PWA 0.92 (0.58; 1.47)	**NMA 0.39 (0.18; 0.81); PWA 0.32 (0.14; 0.73)**
NMA -0.69 (-1.41; 0.03)	NMA -0.38 (-0.95; 0.19); PWA -0.38 (-0.95; 0.19)	**NMA -0.29 (-0.55; -0.03); PWA -0.34 (-0.60; -0.07)**	NMA -0.22 (-0.67; 0.22)	NMA -0.23 (-1.50; 1.04)	NMA -0.14 (-0.82; 0.54)	NMA -0.12 (-0.72; 0.48)	NMA -0.09 (-0.69; 0.52)	NMA -0.05 (-0.53; 0.43)	IC	IC	NMA 0.97 (0.57; 1.66)	**NMA 0.42 (0.19; 0.92)**
NMA -0.69 (-1.35; -0.04); PWA -0.69 (-1.35; -0.04)	NMA -0.38 (-1.02; 0.26)	**NMA -0.30 (-0.45; -0.14); PWA -0.28 (-0.44; -0.12)**	NMA -0.23 (-0.58; 0.12); PWA -0.30 (-0.72; 0.13)	NMA -0.24 (-1.47; 1.00); PWA -0.24 (-1.47; 1.00)	NMA -0.14 (-0.78; 0.49); PWA -0.10 (-0.82; 0.62)	NMA -0.13 (-0.66; 0.40); PWA -0.06 (-0.67; 0.54)	NMA -0.09 (-0.66; 0.47)	NMA -0.05 (-0.47; 0.36)	NMA -0.00 (-0.30; 0.29); PWA -0.41 (-1.11; 0.28)	TAU	ST	NMA 0.43 (0.18; 1.02)
NMA -0.91 (-1.85; 0.02)	NMA -0.60 (-1.43; 0.22)	NMA -0.52 (-1.17; 0.14)	NMA -0.45 (-1.19; 0.30)	NMA -0.46 (-1.86; 0.95)	NMA -0.36 (-1.27; 0.55)	NMA -0.35 (-1.19; 0.50)	NMA -0.31 (-1.16; 0.54)	NMA -0.27 (-1.04; 0.49)	NMA -0.22 (-0.82; 0.37); PWA -0.22 (-0.82; 0.37)	NMA -0.22 (-0.89; 0.45)	ACT	SST
NMA -0.87 (-1.66; -0.07)	NMA -0.55 (-1.32; 0.21)	**NMA -0.47 (-0.91; -0.03);** PWA -0.29 (-0.84; 0.26)	NMA -0.40 (-0.88; 0.08); PWA -0.64 (-1.36; 0.08)	NMA -0.41 (-1.72; 0.90)	NMA -0.32 (-1.08; 0.45)	NMA -0.30 (-0.98; 0.38)	NMA -0.26 (-0.96; 0.43)	NMA -0.23 (-0.80; 0.35)	NMA -0.18 (-0.68; 0.33)	NMA -0.17 (-0.61; 0.27); PWA -0.11 (-0.88; 0.67)	NMA 0.05 (-0.74; 0.83)	ST

Positive symptoms | Treatments | Dropout | Treatments

Figure 3 Comparisons between psychological treatments for positive symptoms and study dropouts. Results for positive symptoms are presented in the lower triangle; results for dropout are presented in the upper triangle. Significant results are presented in bold. Relative treatments effects are measured by standardized mean difference (SMD) for positive symptoms and risk ratio (RR) for study dropout along with their 95% confidence intervals (95% CIs). SMDs lower than 0 and RRs lower than 1 favour the column defining treatment. SMDs of –0.2 can be considered small, –0.5 medium, and –0.8 large. To obtain SMDs for comparisons in the opposite direction, negative values should be converted into positive values, and vice versa. To obtain RRs for comparisons in the opposite direction, reciprocals should be taken. ACT – acceptance and commitment therapy, CBT – cognitive behavioral therapy, EFC – experience focused counselling, FI – family intervention, HFIT – hallucination focused integrative treatment, IC – inactive control, MT – metacognitive training, MF – mindfulness, SST – social skills training, ST – supportive therapy, TAU – treatment as usual, WL – waitlist, NMA – network meta-analysis, PWA – pairwise meta-analysis.

CBT	NMA 0.05 (-0.40; 0.49)	NMA 0.01 (-0.37; 0.39); PWA 0.08 (-0.32; 0.48)	NMA -0.00 (-0.65; 0.65); PWA 0.00 (-0.65; 0.65)	NMA -0.09 (-0.59; 0.41)	NMA -0.09 (-0.26; 0.08); PWA -0.09 (-0.26; 0.08)	NMA -0.12 (-0.48; 0.24); PWA -0.15 (-0.53; 0.22)	NMA -0.16 (-0.80; 0.48)	NMA -0.14 (-0.55; 0.26)	NMA -0.18 (-0.69; 0.33); PWA -0.19 (-0.78; 0.41)	**NMA -0.16 (-0.29; -0.03); PWA -0.15 (-0.29; -0.02)**
HFIT	ACT	NMA -0.04 (-0.63; 0.55)	NMA -0.05 (-0.84; 0.74)	NMA -0.14 (-0.81; 0.53)	NMA -0.14 (-0.55; 0.28); PWA -0.14 (-0.55; 0.28)	NMA -0.16 (-0.74; 0.41)	NMA -0.21 (-0.99; 0.57)	NMA -0.19 (-0.74; 0.36)	NMA -0.23 (-0.91; 0.44)	NMA -0.21 (-0.67; 0.26)
NMA -0.26 (-0.98; 0.46)	CBT	SST	NMA -0.01 (-0.77; 0.75)	NMA -0.10 (-0.73; 0.53)	NMA -0.10 (-0.52; 0.32)	NMA -0.13 (-0.55; 0.30); PWA 0.00 (-0.47; 0.47)	NMA -0.17 (-0.88; 0.54)	NMA -0.16 (-0.71; 0.40)	NMA -0.19 (-0.83; 0.44)	NMA -0.17 (-0.57; 0.23)
NMA -0.10 (-1.55; 1.36)	NMA 0.16 (-1.13; 1.45)	EFC	ST	NMA -0.09 (-0.91; 0.73)	NMA -0.09 (-0.76; 0.59)	NMA -0.12 (-0.86; 0.63)	NMA -0.16 (-1.08; 0.75)	NMA -0.14 (-0.91; 0.62)	NMA -0.18 (-1.01; 0.64)	NMA -0.16 (-0.82; 0.51)
NMA -0.42 (-1.35; 0.51)	NMA -0.16 (-0.75; 0.43); PWA -0.08 (-0.71; 0.55)	NMA -0.32 (-1.74; 1.10)	SST	HFIT	NMA 0.00 (-0.52; 0.53)	NMA -0.02 (-0.64; 0.59)	NMA -0.07 (-0.87; 0.74)	NMA -0.05 (-0.69; 0.59)	NMA -0.09 (-0.79; 0.61)	NMA -0.07 (-0.55; 0.41); PWA -0.07 (-0.55; 0.41)
NMA -0.43 (-1.21; 0.34)	NMA -0.17 (-0.46; 0.11); PWA -0.17 (-0.46; 0.11)	NMA -0.34 (-1.66; 0.99)	NMA -0.02 (-0.67; 0.64)	IC	IC	NMA -0.03 (-0.43; 0.37)	NMA -0.07 (-0.74; 0.59)	NMA -0.06 (-0.42; 0.31); PWA -0.06 (-0.42; 0.31)	NMA -0.10 (-0.63; 0.44)	NMA -0.07 (-0.28; 0.14); PWA -0.12 (-0.70; 0.47)
NMA -0.54 (-1.50; 0.43)	NMA -0.28 (-0.94; 0.39); PWA -0·23 (-1.01; 0.54)	NMA -0.44 (-1.88; 1.00)	NMA -0.12 (-1.01; 0.77)	NMA -0.10 (-0.83; 0.62)	FI	WL	NMA -0.04 (-0.66; 0.57); PWA -0.19 (-0.94; 0.55)	NMA -0.03 (-0.57; 0.51)	NMA -0.07 (-0.69; 0.55)	NMA -0.04 (-0.42; 0.34)
NMA -0.59 (-1.60; 0.41)	NMA -0.33 (-1.03; 0.37)	NMA -0.49 (-1.96; 0.98)	NMA -0.17 (-1.09; 0.74)	NMA -0.16 (-0.80; 0.48); PWA -0.16 (-0.80; 0.48)	NMA -0.06 (-1.02; 0.91)	ACT	MT	NMA 0.02 (-0.74; 0.77)	NMA -0.02 (-0.84; 0.79)	NMA 0.00 (-0.64; 0.65) PWA -0.26 (-1.25; 0.73)
NMA -0.59 (-1.47; 0.29)	NMA -0.33 (-0.86; 0.20); PWA -0.28 (-0.85; 0.29)	NMA -0.49 (-1.88; 0.90)	NMA -0.17 (-0.96; 0.62)	NMA -0.16 (-0.76; 0.45)	NMA -0.05 (-0.90; 0.79)	NMA 0.00 (-0.88; 0.88)	ST	AVATAR	NMA -0.04 (-0.69; 0.61)	NMA -0.01 (-0.44; 0.41)
NMA -0.62 (-1.46; 0.21)	NMA -0.36 (-0.80; 0.07); **PWA -0.50 (-0.96; -0.05)**	NMA -0.53 (-1.89; 0.83)	NMA -0.20 (-0.81; 0.40); PWA -0.10 (-0.78; 0.58)	NMA -0.19 (-0.71; 0.33)	NMA -0.09 (-0.88; 0.70)	NMA -0.03 (-0.86; 0.79)	NMA -0.03 (-0.72; 0.65)	WL	FI	NMA 0.03 (-0.48; 0.53); PWA 0.02 (-0.56; 0.60)
NMA -0.64 (-1.33; 0.06) PWA -0.64 (-1.33; 0.06)	**NMA -0.38 (-0.56; -0.20); PWA -0.36 (-0.54; -0.17)**	NMA -0.54 (-1.82; 0.74); PWA -0.54 (-1.82; 0.74)	NMA -0.22 (-0.83; 0.40)	NMA -0.20 (-0.54; 0.13)	NMA -0.10 (-0.77; 0.56); PWA -0.06 (-0.82; 0.70)	NMA -0.05 (-0.77; 0.68)	NMA -0.05 (-0.59; 0.50); PWA 0.05 (-0.76; 0.86)	NMA -0.01 (-0.48; 0.45)	TAU	TAU
NMA -0.75 (-1.79; 0.29)	NMA -0.49 (-1.26; 0.28)	NMA -0.65 (-2.14; 0.84)	NMA -0.33 (-1.24; 0.58)	NMA -0.32 (-1.13; 0.50)	NMA -0.21 (-1.22; 0.79)	NMA -0.16 (-1.20; 0.88)	NMA -0.16 (-1.09; 0.77)	NMA -0.13 (-0.86; 0.61); PWA -0.69 (-1.61; 0.22)	NMA -0.11 (-0.88; 0.66); PWA 0.77 (-0.37; 1.92)	MT

Overall symptoms | Treatments | Negative symptoms | Treatments

Figure 4 Results for overall symptoms are presented in the lower triangle; results for negative symptoms are presented in the upper triangle. Significant results are presented in bold. Relative treatments effects are measured by standardized mean difference (SMD) along with its 95% confidence intervals (95% CIs). SMDs lower than 0 favour the column defining treatment. SMDs of –0.2 can be considered small, –0.5 medium, and –0.8 large. To obtain SMDs for comparisons in the opposite direction, negative values should be converted into positive values, and vice versa. ACT – acceptance and commitment therapy, CBT – cognitive behavioral therapy, EFC – experience focused counselling, FI – family intervention, HFIT – hallucination focused integrative treatment, IC – inactive control, MT – metacognitive training, MF – mindfulness, SST – social skills training, ST – supportive therapy, TAU – treatment as usual, WL – waitlist, NMA – network meta-analysis, PWA – pairwise meta-analysis.

dence in the estimate, respectively). All other relative treatment effects were very imprecise, but on average they favored the active psychological treatment over the inactive control interventions.

The heterogeneity variance (tau^2) was 0.0514, hence considered to be low to moderate[23]. The design-by-treatment interaction test did not reveal significant inconsistency (p= 0.35). By splitting direct and indirect evidence for each comparison, we found no evidence for disagreement between these two pieces of evidence for any of the comparisons. None of the methods we used suggested important inconsistency but, given the low number of studies for most of the comparisons, the power of these tests is low. The assessments of confidence in the estimates using CINeMA highlighted moderate to very low confidence, primarily due to study limitations (high risk of bias) and imprecision.

The interpretation of subgroup analyses is limited due to restricted number of studies available for the different subgroups. We did not detect any important indication that the advantage of CBT over treatment as usual is moderated by number of sessions, study duration, setting (individual vs. group), therapist's expertise and severity at baseline.

Similarly, exclusion of studies for the different sensitivity analyses left a low number of trials for most of the treatments. When excluding open label studies, results of CBT compared to treatment as usual and supportive therapy were consistent with the main analysis (SMD=–0.27; 95% CI: –0.41 to –0.13 and SMD=–0.47; 95% CI: –0.86 to –0.08, respectively), while the difference between CBT and inactive control was not significant anymore (SMD=–0.14; 95% CI: –0.37 to 0.09).

Sensitivity analyses excluding studies presenting only completer analyses, studies with high risk of bias, studies at high risk of bias for researchers' allegiance, or studies focused on treatment resistant patients were overall consistent with the main analyses.

The results of a *post-hoc* sensitivity analysis pooling the "active control" comparators did not differ from the main analysis.

Investigation of small study effect and publication bias with conventional funnel plot did not reveal any association between study precision and effect size (only possible for CBT versus treatment as usual). However, the comparison-adjusted funnel plot suggests that small studies that did not show a benefit for the newer psychological treatment over the older treatment are underrepresented in our data (i.e., they possibly remain unpublished).

Secondary outcomes

CBT and inactive control were less acceptable than treatment as usual in terms of all-cause discontinuation. All treatments had fewer dropouts than social skills training (with the exception of AVATAR therapy, acceptance and commitment therapy, and supportive therapy) (Figure 3).

CBT was associated with a higher reduction of overall symptoms compared to waitlist and treatment as usual, and with higher reduction in negative symptoms compared with treatment as usual (Figure 4). Hallucination focused integrative treatment and CBT were associated with larger probability of response compared with treatment as usual and inactive control.

When looking at adherence and insight, metacognitive training, social skills training, CBT and treatment as usual produced a higher improvement in comparison to supportive therapy. For quality of life and functioning, CBT was more efficacious than treatment as usual. No significant differences were observed for depression. Mortality was in general a rare event, and did not differ between treatments. Very few data were available for relapse, adverse events and other mortality outcomes.

Heterogeneity variance assessed with tau^2 ranged from 0 to 0.0649, being evaluated from none to low-to-moderate. The design-by-treatment interaction model revealed some inconsistency for the secondary outcome of depression (p=0.03).

DISCUSSION

To our knowledge, this is the first network meta-analysis on psychological treatments for patients with positive symptoms of schizophrenia.

With 40 studies, CBT was the most represented among the included treatments. We found significant efficacy for CBT in comparison with treatment as usual in many outcomes (positive, overall and negative symptoms, response to treatment, quality of life and functioning), higher efficacy in comparison with inactive control for positive symptoms and response to treatment, and in comparison with supportive therapy for adherence. There was no convincing proof of efficacy of other treatments, probably due to the small number of studies.

CBT was also associated with higher dropout rates than treatment as usual (18.8% versus 12%). CBT might actually be less acceptable, and not all patients might be willing to engage in such a demanding treatment; however, we argue that to compare the dropout rates with those in treatment as usual could be misleading. Patients in this latter arm – by definition – continue their usual care, and they might have less reason to leave in comparison with patients assigned to a new intervention, that they could find demanding or challenging, or about which they may have high expectations, being discouraged if they do not see results in a few sessions. As a confirmation to this hypothesis, the inactive control condition (where patients participate to sessions like befriending and recreation activities) also had a higher dropout rate than treatment as usual.

Patients in the included studies were only moderately ill on the average, compared with those in a meta-analysis of studies testing antipsychotic drugs vs. placebo, where they were markedly ill[82]. It seems that severely ill patients are usually not enrolled in psychotherapy studies. But this finding just reflects clinical practice: psychotherapy requires a minimum ability of

patients to collaborate, and many patients do not have this ability when they are very acutely ill.

Interpretation of subgroup and sensitivity analyses was limited by the low number of studies available. However, results on CBT remained stable after all pre-planned sensitivity analyses, corroborating the robustness of the results for this intervention. We also tested the potential role of researchers' allegiance[18], by excluding the studies in which the authors tested the efficacy of an intervention that was developed by themselves, and did not find significantly different results from the main analysis.

One open and increasingly relevant issue is whether psychological interventions might cause harm[15]. We collected all the available data about adverse events potentially connected with the psychological intervention, but we found this aspect very poorly reported in the trials. We believe that future studies should collect and report this information, in order to address this still unclear question[83].

Our results are in agreement with findings from some previous pairwise meta-analyses, where CBT was found to be efficacious for overall, positive and negative symptoms of schizophrenia in comparison with control conditions[4-6], but not when compared with other psychological therapies[7]. However, the results of previous studies and reviews regarding the efficacy of CBT for schizophrenia have been conflicting.

In this context, the role of blinded studies may be particularly critical[8]. Here, our results are in contrast with the findings of Jauhar et al[6]: when excluding studies with a non-blind outcome assessor, they found no differences between CBT and any control condition. On the contrary, we found that the superiority for CBT over treatment as usual and inactive control was maintained also in blinded studies. It was not maintained over supportive therapy and waiting list, but only very few studies (two and one, respectively) contributed direct evidence for these comparators.

However, our work cannot be directly compared with that of Jauhar et al[6], because they included any patients with schizophrenia without a restriction to positive symptoms, they used somewhat different criteria for risk of bias, and they lumped all comparators together in their pairwise meta-analysis.

Our findings have the following limitations. First, available data for other treatments than CBT and for CBT versus other nodes than treatment as usual are based on few studies only, leading to low power to detect possible differences. Therefore, results should be interpreted with caution, in particular when looking at sensitivity and subgroup analyses. For this reason we did not focus our interpretation on hierarchies (SUCRA rankings), that could be misleading when there are no statistically significant differences among active treatments.

Second, our focus was on the treatment of positive symptoms, and the findings observed for other outcomes might be secondary to the effect of the treatment on these symptoms. For example, a patient might experience withdrawal, lack of spontaneity, depressive symptoms or a lower functioning due to the difficulties connected with delusions or hallucinations. When these are treated, the quality of life and the other symptoms may benefit as well. For this reason, we focus our interpretations mainly on positive symptoms.

Third, patients in the included trials were also receiving antipsychotic medication. We collected the available information on the use of antipsychotics. However, this was rarely given and never provided for experimental and control arm separately. The only exception is the study of Morrison et al[52], that included patients not receiving antipsychotic medication (a *post-hoc* sensitivity analysis excluding this study did not materially change the results). As a result, it was not possible to assess the role of pharmacological treatment as a moderator. However, we assume that the intake of medications can be considered similar across study arms, due to randomization. Furthermore, we argue that the situation in the included studies resembles what happens in real-life clinical practice, where psychological interventions are intended to be used as add-on to pharmacological therapy, and participants usually continue their previous medication.

On the other hand, this work presents outstanding strengths. First, the study was carefully planned in agreements with PRISMA guidelines, and followed a sound methodology that was *a priori* published in the protocol[3]. This included comprehensive outcome measures and the evaluation of quality at study level (risk of bias) and confidence in results at outcome level (CINeMA). Second, the consideration of control conditions such as treatment as usual and waiting list as separate allowed to ascertain their relative efficacy. This is particularly important, as waitlist has been found to be connected with a nocebo effect[83]. Third, the strict selection criteria led to a homogenous population, as confirmed by very low heterogeneity, coherence across direct and indirect comparisons, and by side-splitting test and design-by-treatment interaction test. This makes us confident that the results of this study are robust.

In conclusion, cognitive behavior therapy seems to be effective on positive symptoms in moderately ill patients with schizophrenia, with effect sizes in the lower to medium range, depending on the control condition.

ACKNOWLEDGEMENTS

This study was funded by European Union's Horizon 2020 Research and Innovation Programme, Marie Skłodowska-Curie grant agreement no. 701717. The funder had no role in study design, data collection, analysis or interpretation, writing of the report, or decision to submit the paper for publication. The authors thank S. Roberts for help in the literature search, P. Kratochwill for assistance in full text acquisition and proof reading, Y. Zhu for help with screening and data extraction from Chinese studies, and C. Carmi for assistance in data extraction. They thank all authors of the included studies, especially the ones who provided additional information and data. On the website http://www.psykl.mri.tum.de/node/69, extensive information can be found about search strategy, included treatments, risk of bias assessment, results for secondary outcomes, SUCRA rankings, evaluation of heterogeneity and inconsistency, subgroup and sensitivity analyses, evaluation of confidence in the estimates. This information is also available upon request from the authors.

REFERENCES

1. National Collaborating Centre for Mental Health. Core interventions in the treatment and management of schizophrenia in adults in primary and sec-

ondary care (Clinical Guideline CG82). London: National Collaborating Centre for Mental Health, 2009.
2. Buchanan RW, Kreyenbuhl J, Kelly DL et al. The 2009 schizophrenia PORT psychopharmacological treatment recommendations and summary statements. Schizophr Bull 2010;36:71-93.
3. Bighelli I, Salanti G, Reitmeir C et al. Psychological interventions for positive symptoms in schizophrenia: protocol for a network meta-analysis of randomised controlled trials. BMJ Open 2018;8:e019280.
4. Wykes T, Steel C, Everitt B et al. Cognitive behavior therapy for schizophrenia: effect sizes, clinical models, and methodological rigor. Schizophr Bull 2008;34:523-37.
5. Zimmermann G, Favrod J, Trieu VH et al. The effect of cognitive behavioral treatment on the positive symptoms of schizophrenia spectrum disorders: a meta-analysis. Schizophr Res 2005;77:1-9.
6. Jauhar S, McKenna PJ, Radua J et al. Cognitive-behavioural therapy for the symptoms of schizophrenia: systematic review and meta-analysis with examination of potential bias. Br J Psychiatry 2014;204:20-9.
7. Jones C, Hacker D, Cormac I et al. Cognitive behaviour therapy versus other psychosocial treatments for schizophrenia. Cochrane Database Syst Rev 2012;4:CD008712.
8. Lynch D, Laws KR, McKenna PJ. Cognitive behavioural therapy for major psychiatric disorder: does it really work? A meta-analytical review of well-controlled trials. Psychol Med 2010;40:9-24.
9. Turner DT, van der Gaag M, Karyotaki E et al. Psychological interventions for psychosis: a meta-analysis of comparative outcome studies. Am J Psychiatry 2014;171:523-38.
10. Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. Res Synth Methods 2012;3:80-97.
11. Hutton B, Salanti G, Caldwell DM et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. Ann Intern Med 2015;162:777-84.
12. Shamseer L, Moher D, Clarke M et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ 2015;349:g7647.
13. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr Bull 1987;13:261-76.
14. Overall JE, Gorham DR. The brief psychiatric rating scale. Psychol Rep 1962; 10:799-812.
15. Linden M, Schermuly-Haupt M-L. Definition, assessment and rate of psychotherapy side effects. World Psychiatry 2014;13:306-9.
16. Higgins JPT, Churchill R, Chandler J et al (eds). Cochrane handbook for systematic reviews of interventions, Version 5.2.0, Cochrane, 2017.
17. Munder T, Brütsch O, Leonhart R et al. Researcher allegiance in psychotherapy outcome research: an overview of reviews. Clin Psychol Rev 2013; 33:501-11.
18. Lieb K. von der Osten-Sacken J, Stoffers-Winterling J et al. Conflicts of interest and spin in reviews of psychological therapies: a systematic review. BMJ Open 2016;6:e010606.
19. Furukawa TA, Salanti G, Atkinson LZ et al. Comparative efficacy and acceptability of first-generation and second-generation antidepressants in the acute treatment of major depression: protocol for a network meta-analysis. BMJ Open 2016;6:e010919.
20. Schwarzer G. meta: an R package for meta-analysis. R News 2007;7:40-5.
21. Schwarzer G, Carpenter JR, Rücker G. Meta-analysis with R (Use-R!). Basel: Springer, 2015.
22. Rücker G, Schwarzer G. Ranking treatments in frequentist network meta-analysis works without resampling methods. BMC Med Res Methodol 2015; 15:58.
23. 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.
24. Dias S, Welton NJ, Caldwell DM et al. Checking consistency in mixed treatment comparison meta-analysis. Stat Med 2010;29:932-44.
25. 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.
26. 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.
27. Salanti G, Del Giovane C, Chaimani A et al. Evaluating the quality of evidence from a network meta-analysis. PLoS One 2014;9:e99682.
28. Barrowclough C, Haddock G, Lobban F et al. Group cognitive-behavioural therapy for schizophrenia. Randomised controlled trial. Br J Psychiatry 2006;189:527-32.
29. Bechdolf A, Knost B, Kuntermann C et al. A randomized comparison of group cognitive-behavioural therapy and group psychoeducation in patients with schizophrenia. Acta Psychiatr Scand 2004;110:21-8.
30. Birchwood M, Michail M, Meaden A et al. Cognitive behaviour therapy to prevent harmful compliance with command hallucinations (COMMAND): a randomized controlled trial. Lancet Psychiatry 2014;1:23-33.
31. Drury V, Birchwood M, Cochrane R et al. Cognitive therapy and recovery from acute psychosis: a controlled trial. I. Impact on psychotic symptoms. Br J Psychiatry 1996;169:593-601.
32. Durham RC, Guthrie M, Morton RV et al. Tayside-Fife clinical trial of cognitive-behavioural therapy for medication-resistant psychotic symptoms. Results to 3-month follow-up. Br J Psychiatry 2003;182:303-11.
33. England M. Efficacy of cognitive nursing intervention for voice hearing. Perspect Psychiatr Care 2007;43:69-76.
34. Foster C, Startup H, Potts L et al. A randomised controlled trial of a worry intervention for individuals with persistent persecutory delusions. J Behav Ther Exper Psychiatry 2010;41:45-51.
35. Freeman D, Pugh K, Dunn G et al. An early Phase II randomised controlled trial testing the effect on persecutory delusions of using CBT to reduce negative cognitions about the self: the potential benefits of enhancing self confidence. Schizophr Res 2014;160:186-92.
36. Freeman D, Dunn G, Startup H et al. Effects of cognitive behaviour therapy for worry on persecutory delusions in patients with psychosis (WIT): a parallel, single-blind, randomised controlled trial with a mediation analysis. Lancet Psychiatry 2015;2:305-13.
37. Freeman D, Waite F, Startup H et al. Efficacy of cognitive behavioural therapy for sleep improvement in patients with persistent delusions and hallucinations (BEST): a prospective, assessor-blind, randomised controlled pilot trial. Lancet Psychiatry 2015;2:975-83.
38. Garety PA, Fowler DG, Freeman D et al. Cognitive-behavioural therapy and family intervention for relapse prevention and symptom reduction in psychosis: randomised controlled trial. Br J Psychiatry 2008;192:412-23.
39. Gottlieb JD, Gidugu V, Maru M et al. Randomized controlled trial of an internet cognitive behavioral skills-based program for auditory hallucinations in persons with psychosis. Psychiatr Rehabil J 2017;40:283-92.
40. Habib N, Dawood S, Kingdon D et al. Preliminary evaluation of culturally adapted CBT for psychosis (CA-CBTp): findings from developing culturally-sensitive CBT project (DCCP). Behav Cogn Psychother 2015;43:200-8.
41. Haddock G, Tarrier N, Morrison AP et al. A pilot study evaluating the effectiveness of individual inpatient cognitive-behavioural therapy in early psychosis. Soc Psychiatry Psychiatr Epidemiol 1999;34:254-8.
42. Haddock G, Barrowclough C, Shaw JJ et al. Cognitive-behavioural therapy v. social activity therapy for people with psychosis and a history of violence: randomised controlled trial. Br J Psychiatry 2009;194:152-7.
43. Hazell CM, Hayward M, Cavanagh K et al. Guided self-help cognitive behavioral intervention for VoicEs (GiVE): study protocol for a pilot randomized controlled trial. Trials 2016;17:351.
44. Krakvik B, Grawe RW, Hagen R et al. Cognitive behaviour therapy for psychotic symptoms: a randomized controlled effectiveness trial. Behav Cogn Psychother 2013;41:511-24.
45. Kuipers E, Garety P, Fowler D et al. London-East Anglia randomised controlled trial of cognitive-behavioural therapy for psychosis. I: Effects of the treatment phase. Br J Psychiatry 1997;171:319-27.
46. Lecomte T, Leclerc C, Corbiere M et al. Group cognitive behavior therapy or social skills training for individuals with a recent onset of psychosis? Results of a randomized controlled trial. J Nerv Ment Dis 2008;196:866-75.
47. Lee DH, Ko SM, Choi YS et al. A randomized controlled pilot study of cognitive behavioral social skills training (Korean version) for middle- or older-aged patients with schizophrenia: a pilot study. J Korean Neuropsychiatr Assoc 2012;51:192-201.
48. Lee DE, Lee HJ, Yoon OS et al. The effect of cognitive behavioral therapy in drug-resistant patients with schizophrenia. J Korean Neuropsychiatr Assoc 2013;52:26-32.
49. Levine J, Barak Y, Granek I. Cognitive group therapy for paranoid schizophrenics: applying cognitive dissonance. J Cogn Psychother 1998;12:3-12.
50. Li ZJ, Guo ZH, Wang N et al. Cognitive-behavioural therapy for patients with schizophrenia: a multicentre randomized controlled trial in Beijing, China. Psychol Med 2015;45:1893-905.
51. McLeod T, Morris M, Birchwood M et al. Cognitive behavioural therapy group work with voice hearers. Part 1. Br J Nurs 2007;16:248-52.

52. Morrison AP, Turkington D, Pyle M et al. Cognitive therapy for people with schizophrenia spectrum disorders not taking antipsychotic drugs: a single-blind randomised controlled trial. Lancet 2014;383:1395-403.
53. Penn DL, Meyer PS, Evans E et al. A randomized controlled trial of group cognitive-behavioral therapy vs. enhanced supportive therapy for auditory hallucinations. Schizophr Res 2009;109:52-9.
54. Pinninti NR, Rissmiller DJ, Steer RA. Cognitive-behavioral therapy as an adjunct to second-generation antipsychotics in the treatment of schizophrenia. Psychiatr Serv 2010;61:940-3.
55. Pot-Kolder RMCA, Geraets CNW, Veling W et al. Virtual-reality-based cognitive behavioural therapy versus waiting list control for paranoid ideation and social avoidance in patients with psychotic disorders: a single-blind randomised controlled trial. Lancet Psychiatry 2018;5:217-26.
56. Rector NA, Seeman MV, Segal ZV. Cognitive therapy for schizophrenia: a preliminary randomized controlled trial. Schizophr Res 2003;63:1-11.
57. Sensky T, Turkington D, Kingdon D et al. A randomized controlled trial of cognitive-behavioral therapy for persistent symptoms in schizophrenia resistant to medication. Arch Gen Psychiatry 2000;57:165-72.
58. Startup M, Jackson MC, Bendix S. North Wales randomized controlled trial of cognitive behaviour therapy for acute schizophrenia spectrum disorders: outcomes at 6 and 12 months. Psychol Med 2004;34:413-22.
59. Tarrier N, Yusupoff L, Kinney C et al. Randomised controlled trial of intensive cognitive behaviour therapy for patients with chronic schizophrenia. BMJ 1998;317:303-7.
60. Trower P, Birchwood M, Meaden A et al. Cognitive therapy for command hallucinations: randomised controlled trial. Br J Psychiatry 2004;184:312-20.
61. Turkington D, Kingdon D. Cognitive-behavioural techniques for general psychiatrists in the management of patients with psychoses. Br J Psychiatry 2000;177:101-6.
62. Valmaggia LR, van der Gaag M, Tarrier N et al. Cognitive-behavioural therapy for refractory psychotic symptoms of schizophrenia resistant to atypical antipsychotic medication. Randomised controlled trial. Br J Psychiatry 2005;186:324-30.
63. van der Gaag M, Stant AD, Wolters KJ et al. Cognitive-behavioural therapy for persistent and recurrent psychosis in people with schizophrenia-spectrum disorder: cost-effectiveness analysis. Br J Psychiatry 2011;198:59-65.
64. Velligan DI, Tai S, Roberts DL et al. A randomized controlled trial comparing cognitive behavior therapy, cognitive adaptation training, their combination and treatment as usual in chronic schizophrenia. Schizophr Bull 2015;41:597-603.
65. Wahass S, Kent G. The modification of psychological interventions for persistent auditory hallucinations to an Islamic culture. Behav Cogn Psychother 1997;25:351.
66. Wittorf A, Jakobi UE, Bannert KK et al. Does the cognitive dispute of psychotic symptoms do harm to the therapeutic alliance? J Nerv Ment Dis 2010;198:478-85.
67. Wykes T, Hayward P, Thomas N et al. What are the effects of group cognitive behaviour therapy for voices? A randomised control trial. Schizophr Res 2005;77:201-10.
68. ACTRN12616000976482. Efficacy of individualised metacognitive therapy (MCT+) for delusions in psychosis. Australian New Zealand Clinical Trial Registry, 2016.
69. Briki M, Monnin J, Haffen E et al. Metacognitive training for schizophrenia: a multicentre randomised controlled trial. Schizophr Res 2014;157:99-106.
70. Favrod J, Rexhaj S, Bardy S et al. Sustained antipsychotic effect of metacognitive training in psychosis: a randomized-controlled study. Eur Psychiatry 2014;29:275-81.
71. Kumar D, Zia Ul Haq M, Dubey I et al. Effect of meta-cognitive training in the reduction of positive symptoms in schizophrenia. Eur J Psychother Couns 2010;12:149-58.
72. So SH, Chan AP, Chong CS et al. Metacognitive training for delusions (MCTd): effectiveness on data-gathering and belief flexibility in a Chinese sample. Front Psychol 2015;6:730.
73. van Oosterhout B, Krabbendam L, de Boer K et al. Metacognitive group training for schizophrenia spectrum patients with delusions: a randomized controlled trial. Psychol Med 2014;44:3025-35.
74. Chadwick P, Hughes S, Russell D et al. Mindfulness groups for distressing voices and paranoia: a replication and randomized feasibility trial. Behav Cogn Psychother 2009;37:403-12.
75. Chadwick P, Strauss C, Jones AM et al. Group mindfulness-based intervention for distressing voices: a pragmatic randomised controlled trial. Schizophr Res 2016;175:168-73.
76. Bach P, Hayes SC. The use of acceptance and commitment therapy to prevent the rehospitalization of psychotic patients: a randomized controlled trial. J Consult Clin Psychol 2002;70:1129-39.
77. Shawyer F, Farhall J, Thomas N et al. Acceptance and commitment therapy for psychosis: randomised controlled trial. Br J Psychiatry 2016;210:140-8.
78. Schnackenberg J, Fleming M, Martin CR. A randomised controlled pilot study of Experience Focused Counselling with voice hearers. Psychosis 2017;9:12-24.
79. Jenner JA, Nienhuis FJ, Wiersma D et al. Hallucination focused integrative treatment: a randomized controlled trial. Schizophr Bull 2004;30:133-45.
80. Craig TK, Rus-Calafell M, Ward T et al. AVATAR therapy for auditory verbal hallucinations in people with psychosis: a single-blind, randomised controlled trial. Lancet Psychiatry 2018;5:31-40.
81. Adams CE, Coutinho E, Davis JM et al. Cochrane Schizophrenia Group. The Cochrane library. Chichester: Wiley, 2011.
82. Leucht S, Kane JM, Kissling W et al. What does the PANSS mean? Schizophr Res 2005;79:231-8.
83. Hutton P. Should people with psychosis be supported in choosing cognitive therapy as an alternative to antipsychotic medication: a commentary on a commentary. Schizophr Res (in press).

DOI:10.1002/wps.20577

Tardive dyskinesia risk with first- and second-generation antipsychotics in comparative randomized controlled trials: a meta-analysis

Maren Carbon[1], John M. Kane[1-4], Stefan Leucht[4], Christoph U. Correll[1-3,5]

[1]Department of Psychiatry, Zucker Hillside Hospital, Glen Oaks, NY, USA; [2]Department of Psychiatry and Molecular Medicine, Hofstra Northwell School of Medicine, Hempstead, NY, USA; [3]Center for Psychiatric Neuroscience, Feinstein Institute for Medical Research, Manhasset, NY, USA; [4]Department of Psychiatry and Psychotherapy, Klinikum rechts der Isar der Technischen Universität München, Munich, Germany; [5]Campus Virchow-Klinikum, Charité-Universitätsmedizin Berlin, and Department of Child and Adolescent Psychiatry, Berlin Institute of Health, Berlin, Germany

Tardive dyskinesia (TD) risk with D2/serotonin receptor antagonists or D2 receptor partial agonists (second-generation antipsychotics, SGAs) is considered significantly lower than with D2 antagonists (first-generation antipsychotics, FGAs). As some reports questioned this notion, we meta-analyzed randomized controlled studies (RCTs) to estimate the risk ratio (RR) and annualized rate ratio (RaR) of TD comparing SGAs vs. FGAs and SGAs vs. SGAs. Additionally, we calculated raw and annualized pooled TD rates for each antipsychotic. Data from 57 head-to-head RCTs, including 32 FGA and 86 SGA arms, were meta-analyzed, yielding 32 FGA-SGA pairs and 35 SGA-SGA pairs. The annualized TD incidence across FGA arms was 6.5% (95% CI: 5.3-7.8%) vs. 2.6% (95% CI: 2.0-3.1%) across SGA arms. TD risk and annualized rates were lower with SGAs vs. FGAs (RR=0.47, 95% CI: 0.39-0.57, p<0.0001, k=28; RaR=0.35, 95% CI: 0.28-0.45, p<0.0001, number-needed-to-treat, NNT=20). Meta-regression showed no FGA dose effect on FGA-SGA comparisons (Z=–1.03, p=0.30). FGA-SGA TD RaRs differed by SGA comparator (Q=21.8, df=7, p=0.003), with a significant advantage of olanzapine and aripiprazole over other non-clozapine SGAs in exploratory pairwise comparisons. SGA-SGA comparisons confirmed the olanzapine advantage vs. non-clozapine SGAs (RaR=0.66, 95% CI: 0.49-0.88, p=0.006, k=17, NNT=100). This meta-analysis confirms a clinically meaningfully lower TD risk with SGAs vs. FGAs, which is not driven by high dose FGA comparators, and documents significant differences with respect to this risk between individual SGAs.

Key words: Tardive dyskinesia, first-generation antipsychotics, second-generation antipsychotics, randomized controlled studies, schizophrenia, meta-analysis, annualized incidence, clozapine, aripiprazole

(World Psychiatry 2018;17:330–340)

Can tardive dyskinesia (TD), a condition of potentially irreversible abnormal involuntary movements associated to treatment with D2 receptor antagonists (first-generation antipsychotics, FGAs), and producing a significant impairment of functioning and quality of life[1,2], be considered relatively irrelevant for treatment with second-generation antipsychotics (SGAs)?

Based on studies conducted until 2004, the annual TD incidence during SGA treatment was estimated as 0.8% in non-elderly adults[3], one fifth of the rate (5.4%) with FGAs. Surprisingly, however, equal rates of TD during SGA or select FGA treatment were reported in two large randomized controlled trials (RCTs) in schizophrenia, the UK-based Cost Utility of the Latest Antipsychotic Drugs in Schizophrenia Study (CUtLASS-1)[4] and the US-based Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) study[5]. As a pre-emptive risk reduction, the CATIE study[5] did not include subjects with a history of TD into the FGA arms, but only in the SGA arms, limiting the interpretation of reported TD rates. Moreover, most meta-analyses of SGA treatment do not comment on TD risk, as most RCTs are shorter than 3 months (the minimum duration required to diagnose TD[6,7] and/or report continuous outcome measures for dyskinesia), being not suited to identify TD cases[8-10]. Thus, information on incident TD during RCTs is very scarce.

We recently summarized prevalence rates of TD[11], finding that 20% of subjects with current SGA treatment presented with at least mild screening-based probable TD. This observation contrasts with the clinical perception of vanishing TD during SGA treatment, possibly due to the overrating of mild TD cases detected only with screening, but not perceived as clinically meaningful. Importantly, however, prevalence rates are inappropriate indicators of treatment-associated risk for TD, as the antipsychotic given at the time of TD development may not be the one prescribed at the time of TD assessment.

To determine the incidence and the relative risk of treatment-emergent TD with FGA or SGA treatment, we searched for RCTs with a duration of ≥3 months comparing ≥2 antipsychotics and reporting incident TD cases. We aimed to describe TD risk by: a) class-wise comparisons of pooled FGAs to each specific SGA, b) drug-wise comparisons of a specific FGA to one or more specific SGAs, and c) comparisons between individual SGAs.

Based on our earlier study[3], we hypothesized that the incidence and the relative TD risk would be lower with SGAs vs. FGAs. Moreover, based on the association between TD incidence and FGA dose[12], we hypothesized that studies using high dose FGA comparators would drive the estimated risk reduction of SGAs vs. FGAs. Lastly, based on the association between TD incidence and early parkinsonian side effects[12], we hypothesized that SGAs with a higher propensity for extrapyramidal side effects (EPS) would be associated to a higher risk for TD vs. SGAs with lower EPS potential.

METHODS

Literature search

Two authors (CC, MC) independently conducted a PubMed/Web of Science search without language or time restriction for comparative randomized antipsychotic studies (last search update: January 31, 2018). The following terms were used in the advanced search option: all fields (= any field): (antipsychotic*) AND (amisulprid* OR aripiprazol* OR asenapin* OR clozapine OR olanzapin* OR paliperidon* OR quetiapin* OR risperidon* OR sertindole OR ziprasidon* OR zotepin* OR iloperidon* OR cariprazin* OR brexpiprazol* OR lurasidone OR blonanserin OR melperone) AND (randomized controlled trial); publication type: NOT review.

Additionally, we hand-searched references from Cochrane meta-analyses on SGAs. Whenever data needed for the meta-analysis were missing, we repeatedly contacted the authors for additional information and for access to data repositories for unpublished data.

Study inclusion criteria

We included all head-to-head comparisons of one of the above listed antipsychotics in any (oral or i.m.) form of administration to any FGA or other SGA without restriction on age or gender of participants. Requiring that at least one arm consisted of an SGA, we ensured that any FGA comparator was studied at the same time as SGAs were available, avoiding potential biases due to time effects regarding different patient populations and dosing schemas in trials before availability of SGAs.

Included studies had to provide information on the rate of probable treatment-emergent TD in subjects free of TD at study baseline. The diagnosis of probable TD could be clinical or scale-based, as long as diagnostic criteria were clearly defined and identical for all treatment arms. Fixed and flexible dose studies were eligible, as long as baseline randomization was present. Any psychiatric or medical diagnosis was allowed, except for movement disorders. Hence, antipsychotic trials in schizophrenia patients with a serious concomitant medical illness as an inclusion criterion were not excluded. Studies allowing concomitant or prophylactic anticholinergic medication were similarly included, and anticholinergic medication was assessed as moderator. Trials that allowed switching of treatments between groups were excluded. For trials which had a crossover design, only results from the first randomization period were considered, to avoid carry-over effects.

A minimum trial duration of three months was considered for the identification of cases of probable dyskinesia. However, longer observation periods are more appropriate, and the duration of exposure was explored as moderator of TD incidence. Trials in which treatment consisted of concomitant use of ≥2 antipsychotics per individual were excluded, as we aimed to differentiate effects of specific antipsychotics. We did not exclude randomized, open-label studies, but excluded these in sensitivity analyses to address the lack of blinding, which has been shown to be a substantial source of bias[13].

Data extraction

Two authors (MC, CC) independently checked eligibility and extracted data. Any disagreement was resolved by discussion/consensus.

In addition to antipsychotic-specific and class-wise TD rates, we extracted: chlorpromazine equivalent dose for FGAs and olanzapine equivalent dose for SGAs (using the methods described by Leucht et al[14], where 300 mg of chlorpromazine equals 10 mg of olanzapine according to the daily dose method), publication year, study design, geographic region, patient gender, age, ethnicity, clinical diagnoses, illness duration, disease severity, comorbid parkinsonism, TD rating scale and scores, diagnostic TD criteria. Data on TD severity were provided in a minority of studies only. Whenever possible, rates of persistent TD were used, as this measure is clinically more meaningful and less prone to variability than probable TD.

Statistical analysis

We conducted two parallel sets of pairwise, head-to-head analyses: FGA-SGA across-class comparisons and SGA-SGA within-class comparisons. In both sets of analyses, the effect size calculation was based on the pairing of treatment arms as in the original RCT to maintain control of confounding variables. Thus, we calculated an effect size for each available head-to-head combination. From all studies we extracted the number of TD cases per treatment arm (N = intent-to-treat, ITT) and the duration of antipsychotic exposure within this group. Then, we calculated the raw TD incidence risk, as the ratio of TD cases per total exposed subject number, and the annualized TD incidence risk, as cases per exposed subject number per time of antipsychotic exposure (i.e., person years), each with their respective 95% confidence intervals (CIs). Numbers-needed-to-treat (NNTs) reflecting reduced annualized risk for TD were calculated dividing 1 by the rate difference.

We next used subgroup comparisons and meta-regression to explore the relative risk change by antipsychotic class and for specific moderators: mean age, male gender percentage, Caucasian ethnicity percentage, dosing (dichotomizing mean/median study dose – i.e., <500 mg vs. ≥500 mg chlorpromazine equivalent, and <3 mg vs. ≥3 mg haloperidol in studies using haloperidol – or dichotomizing the maximum allowed haloperidol dose range: <10 mg vs. ≥10 mg), illness duration and stage, disease severity, comorbid parkinsonism, information on prior FGA exposure, publication year, study design, duration and sponsorship, data source, case definition, geographic region.

A multivariable analysis was performed, including significant moderators which had been identified in univariate analysis. Since TD risk is cumulative and higher the longer patients are followed, we conducted these moderator analyses only for annualized rate ratios (RaRs) that correct for any differences in observation time. All analyses used a random effects model and were two-sided, with alpha=0.05. Data were analyzed with Comprehensive Meta-Analysis Version 3.

RESULTS

Search results

Of 3,438 hits in PubMed and Web of Science, 273 full-text articles were screened, resulting in 57 articles that fulfilled all inclusion criteria (Figure 1). Notably, a lack of TD reporting resulted in the exclusion of 75 studies of eligible cohorts.

Sample characteristics

FGA-SGA studies

Thirty-two randomized studies comparing FGAs to SGAs and providing data on TD incidence in 10,706 subjects were included[5,15-45]. Twenty-two of them were double-blind studies[5,15,17,19,21-25,27-29,31,33,36,38-44] and ten were randomized open studies[16,18,20,26,30,32,34,37,39,45] (partly with blinded ratings). Haloperidol was the FGA comparator in the majority of studies. All but one study[43] included subjects with schizophrenia-spectrum disorders, and one third of the studies were industry independent[5,22,23,25,28,30,34-36,39,42,45].

Overall, the studies included 65.6% males, 63.2% Caucasians, with a mean age of 37.7±12.4 years and a mean illness duration of 13.2±12.6 years. The mean chlorpromazine equivalent dose was 423.9±252.4 mg/day (range: 69-733.5 mg/day) for FGA treatment, and 522.6±199.3 mg/day (range: 118-905.7 mg/day) for SGA treatment[14]. The mean PANSS total (from the fifteen studies reporting this measure[5,21-24,27-30,33,34,36,38,40,42]) was 76.3±20.2 at baseline. The median study duration (by design) was one year (interquartile range, IQR: 0.44-2.0). The median observed follow-up lasted 108 person years (IQR: 34.5-254.3) for SGAs and 73.1 person years (IQR: 25.0-117.6) for FGAs.

Few studies included only subjects with first-episode or early-phase schizophrenia. The reported mean illness duration was 14 years (IQR: 10.6-17.1, k=21) for patients on SGAs and 13.7 years (IQR: 10.2-17.4, k=21) for patients on FGAs. Data on prior FGA exposure were very sparse; sixteen studies[5,16,24,27,29,30,32,34-37,39,41,42,45,46] reported that included subjects had prior FGA exposure. The mean percentage of patients with pre-baseline FGA exposure was 76.4±30.5% (IQR: 40.0-100%, k=12).

The majority (81.3%) of FGA-SGA studies used standardized screening instruments to detect and report dyskinesia. The Abnormal Involuntary Movement Scale (AIMS) was mostly used in conjunction with Schooler-Kane criteria for probable TD. Four studies[5,21,23,32] reported both probable and persistent TD, and five studies[15,22,24,38,42] reported persistent TD only. Systematic screening was also performed using the Extrapyramidal Symptom Rating Scale (ESRS)[19,23,24,41,47], the St. Hans Rat-

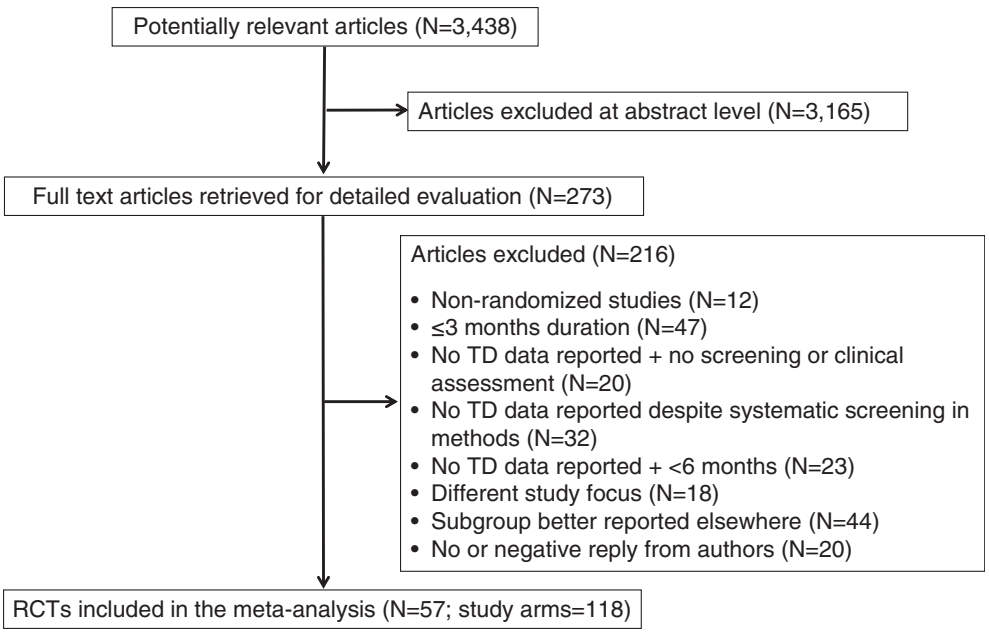

Figure 1 PRISMA flow chart. TD – tardive dyskinesia, RCTs – randomized controlled trials

ing Scale[30] and the Dyskinesia Rating Scale[25]. Only six studies assessed TD based on clinical reporting[16,17,26,28,39,45].

SGA-SGA studies

Twenty-three randomized studies comparing one SGA to another SGA and providing data on TD incidence in 9,153 subjects were included in the analysis (comparators: olanzapine[31,47-59], clozapine[60-62], or risperidone[63-68]).

SGA-SGA comparisons also included treatment arms from FGA-SGA studies with randomization into multiple SGA arms and with TD data reporting by specific antipsychotic, thus adding data on SGA treatment from five studies listed also above[5,30,31,34,41].

Based on the frequency of included studies and their use of specific antipsychotics (but not *a priori*), the SGA-SGA studies were grouped into the following three subsets for analyses: a) 19 comparisons with olanzapine as one comparator; b) three comparisons with clozapine as one comparator; c) six comparisons with risperidone or paliperidone (which is metabolized from risperidone) as one comparator. This latter group excluded olanzapine as the other comparator, as these head-to-head comparisons had been included in the first subset. We used sensitivity analyses to separate the effect of risperidone-olanzapine head-to-head comparisons in the studies of the first and third group.

The majority of studies included adults with schizophrenia spectrum disorders; two studies included subjects with bipolar disorder[57,64]; two subjects with acute psychosis[51,52]; one patients with first-episode psychosis[48]; and one patients with pediatric schizophrenia or schizoaffective disorder[22].

Overall, these studies included 62.2% males, 67.0% Caucasian, with a mean age of 37.0±10.7 years and a mean illness duration of 11.8±6.8 years. Four studies included only subjects with first-episode or early-phase schizophrenia[22,30,48,51]. Data on prior FGA exposure were reported only in six studies (being 0-33% in three studies[30,47,51], and >50% in two[5,50]).

SGA doses converted into mean olanzapine equivalent doses[14] were: a) for studies with olanzapine, 13.5±3.4 mg/day (range: 4.8-20.1 mg/day) for olanzapine and 10.6±2.9 mg/day (range: 6.8-13.6 mg/day) for the comparators; b) for studies with clozapine, 10.7 mg/day for clozapine and 21.0±4.6 mg/day (range: 19.0-23.4 mg/day) for the comparators; c) for studies with risperidone or paliperidone, 12.1±3.2 mg/day (range: 8.6-16.0 mg/day) for risperidone/paliperidone and 12.3±2.7 mg/day (range: 8.6-15.0 mg/day) for the comparators. The mean PANSS total (from studies reporting this measure[22,47-49,52-56,60-62,64-67]) was 79.5±16.6 at baseline.

In contrast to FGA-SGA studies, the use of standardized TD screening instruments was slightly lower in SGA-SGA studies (64.3%), with a third of studies relying on clinical or patient reporting[48,50,52,54,58,63,67,68]. The majority of studies reported probable TD, except for CATIE[5].

Full sample analyses: mean TD incidence in randomized studies

The estimated weighted mean incidence of TD across all FGA treatment groups was 6.5% (95% CI: 4.6-9.0%) for 3,763 subjects in 32 treatment arms. Similarly, the annualized incidence was 6.5% (95% CI: 5.3-7.8%; see Table 1 for specific FGAs).

The estimated weighted mean incidence of TD across all SGA treatment groups was 3.0% (95% CI: 2.4-3.8%) for 15,092 subjects in 86 treatment arms. The annualized incidence was 2.6% (95% CI: 2.0-3.1%; see Table 1 for specific SGAs).

As randomization with regard to TD was unbalanced in the CATIE study (subjects with a history of TD were barred from randomization to FGA treatment), a sensitivity analysis was performed after excluding the CATIE data, yielding a weighted mean annualized incidence of TD for FGAs of 6.8% (95% CI: 5.5-8.1%) and for SGAs of 2.6% (95% CI: 2.1-3.2%).

Further sensitivity analyses corroborated the range of observations above, but also showed slightly lower rates for SGAs after exclusion of clozapine studies (weighted mean annualized incidence of 2.4%, 95% CI: 1.9-3.0).

TD incidence in studies using only clinical reporting (but not systematic screening) was 3.8% (95% CI: 2.1-6.7%) for FGA arms, and 0.9% (95% CI: 0.3-2.4%) for SGA arms.

FGA-SGA sample analyses: TD risk and moderators of TD risk differences

Of the 32 FGA-SGA included studies, only 28 contributed analytically to the comparative TD risk analysis, as meta-analytic risk calculations exclude studies with zero events in both treatment arms.

Primary analysis

The estimated TD risk ratio (RR) was significantly lower with SGAs relative to FGAs (RR=0.47, 95% CI: 0.39-0.57, p<0.0001, k=28). Similarly, the estimated TD RaR, reflecting the annualized incidence, was significantly lower in SGAs relative to FGAs (RaR=0.35, 95% CI: 0.28-0.45, p<0.0001, k=28; NNT=20, 95% CI: 15-31) (Figure 2).

The estimated TD RR after exclusion of CATIE data was 0.35 (95% CI: 0.27-0.44, p<0.0001, k=27). The TD RaR after CATIE exclusion was 0.34 (95% CI: 0.27-0.44, p<0.0001, k=27).

Moderator analyses of TD RaRs

The mean TD RaR varied significantly by SGA comparator (Q=21.8, df=7, p=0.003). The advantage of SGAs vs. FGAs was most prominent with aripiprazole and smallest with quetiapine (Table 2). Exploratory pairwise comparisons showed significantly lower RaRs with aripiprazole relative to all other

Table 1 Incidence of tardive dyskinesia (TD) for specific antipsychotics

Antipsychotic	N. studies/treatment arms	N. subjects	Mean raw TD incidence		Mean annualized TD incidence	
			%	95% CI	%	95% CI
FGAs						
Perphenazine	1	853	3.3	0.5-17.4	3.7	0.1-6.7
Molindone	1	20	2.4	0.1-38.5	4.2	–0.8 to 16.5
Haloperidol	22	2,975	6.6	4.5-9.6	7.5	5.9-9.2
Chlorpromazine	1	80	21.3	4.4-60.1	11.2	4.8-17.7
Fluphenazine	1	28	3.6	3.0-32.8	12.5	12.3-37.3
SGAs						
Aripiprazole	3	1,215	0.9	0.3-3.1	1.7	–0.8 to 4.1
Amisulpride	3	558	1.5	0.4-5.4	2.4	–0.4 to 5.2
Asenapine	4	1,472	1.2	0.5-3.3	2.4	–0.1 to 4.8
Risperidone	20	853	4.2	2.6-6.6	2.4	1.2-3.5
Quetiapine	6	221	2.8	0.2-1.1	2.5	0.2-4.8
Olanzapine	29	5,686	2.7	1.9-4.0	2.9	1.8-3.9
Ziprasidone	7	918	3.5	1.6-7.5	3.5	1.3-5.7
Clozapine	6	348	8.2	3.9-16.6	4.2	1.7-6.7
Lurasidone	1	427	2.6	0.5-13.1	4.8	–0.2 to 9.3

The TD rates for the individual medications are not directly comparable, as they do not originate from randomized trials in which those medications were compared head-to-head, but are pooled irrespective of study comparators. FGAs – first-generation antipsychotics, SGAs – second-generation antipsychotics

SGAs. Moreover, RaRs with olanzapine were also significantly lower relative to risperidone and quetiapine (Table 3).

The mean TD RaR did not vary significantly by FGA comparator (Q=0.23, df=1, p=0.63). The advantage of SGAs over FGAs persisted after exclusion of all studies in which haloperidol was used as the FGA comparator (RaR=0.39, 95% CI: 0.25-0.61, p<0.0001, k=8). In those studies with haloperidol as the comparator, similar RaRs were obtained (RaR=0.34, 95% CI: 0.26-0.45, p<0.0001, k=20). In the studies with specified, non-haloperidol FGAs[5,21,25,34], a somewhat weaker, but still significant advantage of SGAs over FGAs emerged (RaR=0.47, 95% CI: 0.27-0.82, p=0.007, k=4). The comparison with all other studies (i.e., those that included haloperidol as the only, or as one possible FGA) was non-significant (Q=1.27, df=1, p=0.26).

TD RaRs were similar in first episode/early-phase psychosis cohorts[22-25,27,30] vs. the remaining unrestricted cohorts (Q=0.04, df=1, p=0.85). Prior FGA exposure was not consistently reported, and no cohort explicitly included only subjects without prior FGA exposure.

TD RaRs varied significantly by sponsorship (Q=10.0, df=1, p=0.003). The reduction in TD incidence associated with SGA treatment was accentuated in industry-sponsored relative to academic studies, but persisted independently in both of them (academic studies: RaR=0.59, 95% CI: 0.39-0.87, p=0.008, k=11; industry-sponsored studies: RaR=0.28, 95% CI: 0.21-0.35, p< 0.001, k=20).

A mixed regression model including SGA comparator and sponsorship (Q=20.74, df=6, p=0.002) confirmed the independent effect of SGA comparator (Q=10.78, df=5, p=0.05), while no significant effect of sponsorship was found (Z=–1.2, df=1, p=0.23).

Moderator and meta-regression analyses were non-significant for age, gender, illness duration (both as years of disease duration and categorically as first episode/early-phase psychosis vs. other cohorts), disease severity, study region, study duration, anticholinergic use, study design (open-label extension vs. blinded trial), year of study start, and case definition. There was no effect of FGA dose on TD rate ratios, neither when data were dichotomized (below vs. above 500 mg chlorpromazine equivalents, Q=0.19, df=1, p=0.66), nor when mean FGA dose in chlorpromazine equivalents was used in a meta-regression model (Z=–1.03, p=0.30). Similarly, there was no effect of the FGA-SGA ratio on TD RaRs (Z=1.56; p=0.12).

Data on TD severity were provided in a minority of studies (31.3%). TD severity was described in very heterogeneous formats, which did not allow for the use of this variable in the meta-analysis. Discontinuation due to TD was reported rarely, but slightly more frequently in FGAs (N=1 in FGA vs. N=0 in SGA[21]; 2.7% in FGA vs. 0.7% with paliperidone[36]). Greater severity of TD with FGAs was reported repeatedly[5,23,32,35,38]. Reports of severe cases were rare in general, and found only in FGA-treated subjects in four studies[27,38,40,41].

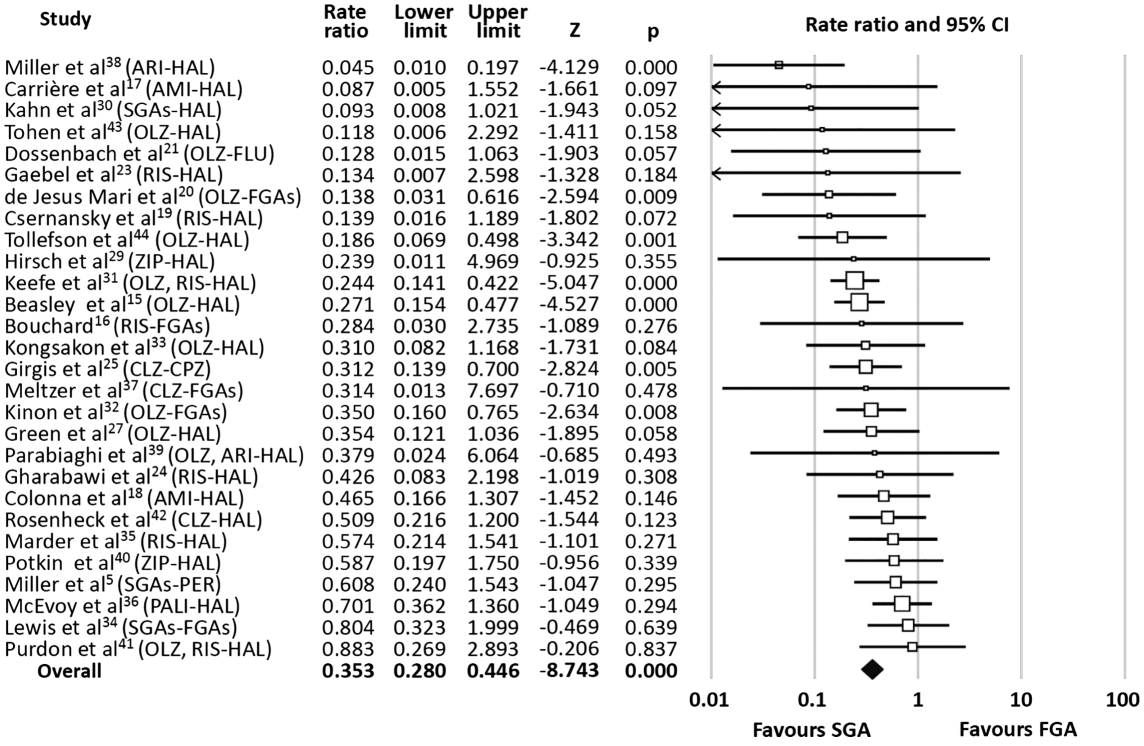

Figure 2 Forest plot of tardive dyskinesia rate ratios of randomized studies comparing first-generation antipsychotics (FGAs) to second generation antipsychotics (SGAs). ARI – aripiprazole, AMI – amisulpride, CLZ – clozapine, CPZ – chlorpromazine, HAL – haloperidol, OLZ – olanzapine, PALI – paliperidone, PER – perphenazine, RIS – risperidone, ZIP – ziprasidone

SGA-SGA sample analyses: TD risk and moderators of TD RaR differences

Olanzapine vs. all other non-clozapine SGAs

Of the 28 SGA-SGA included studies, the most frequently studied SGA was olanzapine, which was a SGA comparator in 17 studies against other non-clozapine SGAs. As CATIE[5] and EUFEST[30] randomized patients also to SGA treatment arms, 19 comparative olanzapine/non-clozapine SGA studies were available. In these 19 studies, comparators were risperidone (k=9), asenapine (k=5), quetiapine (k=4), ziprasidone (k=3), amisulpiride (k=1) and aripiprazole (k=1). The mean dose in the olanzapine treatment arms was 13.5±3.4 mg/day, being 10.6±2.9 mg/day in the comparators, recalculated via mean olanzapine equivalent for all[14]. Three studies[22,52,68] reported no cases of TD in all arms and did thus not contribute to risk calculations, as the risk of non-occurrence is formally inestimable.

Table 2 Mean annualized tardive dyskinesia (TD) rate ratios in SGAs vs. FGAs

SGA	N. treatment arms	N. subjects	Mean TD rate ratio vs. FGA (95% CI)	p
Amisulpride	3	558	0.37 (0.15-0.91)	0.032
Aripiprazole	1	1,215	0.045 (0.01-0.19)	0.000
Clozapine	2	405	0.39 (0.22-0.70)	0.001
Olanzapine	12	5,624	0.25 (0.19-0.34)	0.000
Paliperidone	1	145	0.70 (0.35-1.36)	0.294
Quetiapine	2	786	0.94 (0.35-0.91)	0.915
Risperidone	8	2,479	0.38 (0.25-0.58)	0.000
Ziprasidone	4	887	0.57 (0.26-1.27)	0.169

FGAs – first-generation antipsychotics, SGAs – second-generation antipsychotics

The estimated TD risk ratio was significantly lower with olanzapine relative to other non-clozapine SGAs (RR=0.67, 95% CI: 0.50-0.90, p=0.008, k=17; NNT=100, 95% CI: 63-250). Similarly, the estimated TD RaR was significantly lower with olanzapine relative to non-clozapine SGAs (RaR=0.66, 95% CI: 0.49-0.88, p=0.006, k=17).

Moderator analyses of TD RaRs were non-significant, including age, gender, study region, olanzapine equivalent dose, anticholinergic use, percentage of subjects with prior FGA exposure, year of study start, illness duration, illness stage (first-episode yes/no), sponsorship, and comparator SGA study design.

Treatment-related TD RaRs were significantly lower with olanzapine relative to risperidone (RaR=0.57, 95% CI: 0.37-0.89, p=0.015, k=6), which represented the largest comparator subgroup. No other differences were found for the other SGAs when analyzed in exploratory pairwise comparisons.

Table 3 Moderating effects of the comparator SGA on annualized rate ratios of treatment emergent tardive dyskinesia in studies comparing FGAs vs. SGAs

SGA comparator in FGA study	Amisulpride	Clozapine	Olanzapine	Quetiapine	Risperidone	Ziprasidone
Aripiprazole	**Q=5.75, p=0.017**	**Q=7.14, p=0.008**	**Q=5.04, p=0.025**	**Q=4.24, p=0.039**	**Q=11.27, p=0.004**	**Q=8.84, p=0.003**
Amisulpride	-	Q=0.07, p=0.94	Q=0.65, p=0.42	Q=1.84, p=0.17	Q=2.84, p=0.28	Q=0.49, p=0.49
Clozapine		-	Q=1.72, p=0.19	Q=2.72, p=0.13	Q=2.52, p=0.28	Q=0.58, p=0.45
Olanzapine			-	**Q=6.12, p=0.013**	**Q=8.5, p=0.015**	Q=3.6, p=0.06
Quetiapine				-	Q=4.15, p=0.15	Q=0.608, p=0.43
Risperidone					-	Q=2.58, p=0.27

Significant effects are highlighted in bold prints. FGAs – first-generation antipsychotics, SGAs – second-generation antipsychotics

To address the possibility that the head-to-head studies comparing olanzapine vs. risperidone were driving the favorable TD RaRs of olanzapine vs. non-clozapine SGAs, we excluded those studies in a sensitivity analysis. The TD RaR remained significantly lower with olanzapine relative to all other non-clozapine and non-risperidone SGAs (RaR=0.68, 95% CI: 0.46-0.99, p=0.047, k=12).

Clozapine vs. non-clozapine SGAs

There were only three studies comparing clozapine against another SGA, with two of them vs. olanzapine[61,62], and one vs. mixed SGAs[60], where olanzapine was used in 50% of subjects. These studies included only subjects with a schizophrenia spectrum disorder diagnosis. Subjects were 71% male, 77% Caucasian, with a mean age of 38.7±1.1 years, a mean illness duration of 17.3±5.1 years, and mean doses of 318.0±17.1 mg/day for clozapine and 21.0±2.2 mg/day for olanzapine.

There was no difference regarding probable treatment-emergent TD between clozapine and non-clozapine SGAs (RR=1.07, 95% CI: 0.49-2.34, p=0.86, k=3; RaR=1.10, 95% CI: 0.66-1.90, p=0.71, k=3).

Risperidone/paliperidone vs. non-olanzapine SGAs

There were six studies comparing risperidone or paliperidone against a different non-olanzapine/non-clozapine SGA, four of which used long-acting injectables (LAIs) (risperidone-LAI = 3, paliperidone-LAI = 1), and additional treatment arms from CATIE were also included in this subgroup comparison. The comparators were aripiprazole (k=2), lurasidone (k=1), quetiapine (k=1), ziprasidone (k=1) and mixed SGAs (k=1).

Five of the studies included subjects with a schizophrenia spectrum disorder diagnosis, and one was performed in bipolar disorder patients[64]. Subjects were 56% male, 51% Caucasian, with a mean age of 36.8±5.7 years, and a mean illness duration of 10.9±3.3 years.

There was no difference in RRs or RaRs of probable treatment-emergent TD between risperidone/paliperidone treatment arms and SGA comparators (RR=0.88; 95% CI: 0.50-1.56, p=0.66, k=7; RaR=0.94, 95% CI: 0.65-1.35, p=0.72, k=7).

To address the influence of risperidone-olanzapine head-to-head comparisons that were grouped primarily into the first analysis, we added these studies in a sensitivity analysis to the third set analyses. Doing so, there was no difference in RaRs for probable treatment-emergent TD between risperidone/paliperidone treatment arms and all other SGA comparators including olanzapine (RaR=1.24, 95% CI: 0.91-1.60, p=0.20, k=12).

DISCUSSION

Class-wise TD risk

This meta-analysis on treatment-emergent TD during comparative randomized controlled trials shows an overall low, but still clinically relevant incidence of TD (annualized incidence: FGAs=6.5%; SGAs=2.6%). However, these rates are based only on probable and not persistent TD and may be inflated due to rating scale based assessments that do not take into consideration severity or impact. Conversely, the raw TD incidence rates based on clinical observation (FGAs=3.8%, SGAs=0.9%) have clinical face validity, but may be underestimates.

We confirmed our hypothesis of lower incidence rates of TD during SGA vs. FGA treatment, with a relative risk and annualized rate reduction to one third of the FGA rate and a NNT=20. Interestingly, the rate of clinically reported, but not screening-based TD, which may reflect the more severe and clinically relevant cases, matches earlier incidence reports of TD risk during clinical FGA use and during early clinical trials with SGA treatment[3].

The significant TD risk reduction with SGAs found in this meta-analysis contrasts to the findings of the UK-based CUtLASS-1 study[4] and the US-based CATIE study[5], which both conveyed the impression that the TD risk of FGAs and SGAs did not differ. While our dataset consisted predominantly of studies that used haloperidol as the FGA comparator, our analyses

exploring the role of haloperidol as the comparator showed that it did not drive the FGA-SGA class difference. TD RaRs in studies with haloperidol as the FGA comparator did not differ from those that used pooled groups of mixed FGAs (including CATIE and CUtLASS); and both subgroups of FGA comparators were independently associated with significantly greater TD risk than the respective SGA comparators.

Notably, CATIE and CUtLASS had methodological limitations regarding the specific question of EPS, as they were designed to assess antipsychotic efficacy. Contrasting to the remaining body of studies included in this meta-analysis, the FGA arms of the CATIE study[5] did not include subjects with a history of TD, as these subjects are at increased risk of TD upon re-exposure to FGAs. Similarly, contrasting to the remaining body of studies included in this meta-analysis, the FGA arm of the CUtLASS-1 study[4] was dominated by the use of sulpiride (58% of the FGA group). This antipsychotic has been categorized as atypical based on its high dissociation constant at the D2 receptor[69], consistent with its exceptionally low EPS risk[70]. Interestingly, our results are close to the findings of earlier large incidence studies in clinical cohorts that demonstrated a significant risk reduction for SGAs vs. FGAs of 0.51[71] and of 0.55 after adjustment for lifetime antipsychotic exposure[72].

Due to the relatively low absolute annual risk of TD even during FGA treatment, the NNT to achieve a risk reduction for TD of 20 may appear too high to warrant an influence on clinical antipsychotic choice. On the other hand, the annual risk underestimates the individual lifetime risk in chronic mental illness, where antipsychotic exposure times range around six years in 40 year olds[72], being closer to 15 years in people with schizophrenia spectrum disorders[5]. As the individual risk to develop TD is cumulative, at least during the first five years, the risk reduction should also be seen as a cumulative gain throughout the expected treatment period in each individual.

Contrary to earlier suggestions, the advantage of SGAs over FGAs was not driven by high dose FGA studies, as dose was excluded as a systematic confound of the FGA-SGA risk comparisons. The potential dose effect was consistently ruled out via different approaches: a) mean/median study dose; b) <500 mg vs. ≥500 mg chlorpromazine equivalent and <3 mg vs. ≥3 mg haloperidol; and c) <10 mg vs. ≥10 mg haloperidol. Importantly, this finding does contradict the repeated observation that the individual risk for TD increases with higher FGA doses and clinically relevant EPS[12]. Other risk factors, which are established for individual TD risk, such as age and gender, also had no influence on the comparison of TD risks by antipsychotic class, suggesting that antipsychotic dose, age and gender generalize as risk factors across different antipsychotics.

Consistent with our hypothesis, however, moderator analyses showed that TD risk reduction with SGAs differed with specific SGA comparator. The relative risk reduction in comparisons vs. FGAs was most pronounced in studies with aripiprazole or olanzapine as SGA comparator. However, this information has to be interpreted cautiously, as TD RRs and RaRs vs. FGAs were accentuated in industry-sponsored studies. The majority of olanzapine studies was industry sponsored and, although the mixed regression model formally confirmed the independent effect of SGA comparator, while no effect of sponsorship was found, no definite statement on this interaction can be made due to the effect of factor collinearity between olanzapine treatment and industry sponsorship.

Conversely, why would industry-sponsored trials with predominantly blinded ratings provide more favourable TD rates for SGAs? The industry-sponsored studies differed regarding the following factors: a) a higher number of subjects in FGA treatment arms (but the number of subjects per treatment arm is controlled by weighting in meta-analyses); b) higher dose FGA comparator studies; c) higher mean age; d) fewer early psychosis cohorts; and e) more clinical reporting-based TD assessment, which likely disfavours treatments with more severe TD expression, such as FGAs. Nevertheless, none of these factors was a significant moderator of RaRs in this meta-analysis.

Within-SGA class comparisons

Consistent with the TD risk reduction of olanzapine and aripiprazole relative to FGAs, SGA-SGA comparisons seemed to confirm that agents with the lowest acute EPS risk[73], i.e., clozapine and olanzapine, also have the lowest TD risk. However, TD rate differences within the SGA class, based on an NNT of 100 for olanzapine vs. other non-clozapine SGAs, were rather subtle, and quetiapine – that meta-analytically has very similar acute EPS risk to olanzapine and aripiprazole[73] – had the highest individual TD risk among SGA comparator trials.

Contrasting to the FGA-SGA comparisons, no effect of study sponsor was present in the SGA-SGA analyses. However, most of these studies were industry-sponsored, thus reducing the ability to assess the impact of this factor. Nevertheless, the NNT of 100 for the advantage of olanzapine underscores the need to consider other adverse effects, particularly cardiometabolic risk, when making SGA choices.

Interestingly, and consistent with an earlier clinical TD incidence study[72], the TD risk of clozapine and olanzapine was similar in the three relevant studies. However, clozapine was relatively under-dosed in those studies. Furthermore, there were no studies of clozapine vs. non-olanzapine SGAs, a shortcoming that is all the more unfortunate, as olanzapine and clozapine share the major disadvantage of very significant weight gain. Since aripiprazole also seemed to have a particularly lower TD risk vs. FGAs, comparing a D2 partial agonist to clozapine for TD risk would be valuable.

In contrast to our initial hypothesis that SGAs with higher EPS potential, such as risperidone, would also have a higher TD risk, there were no notable differences of TD risks between non-olanzapine, non-clozapine SGAs, including risperidone/paliperidone in particular, but analyzable data were limited.

Limitations

Multiple limitations have to be taken into consideration when interpreting these results.

First, despite attempts to access this information, 75 randomized studies with an appropriate duration to provide information on TD risk could not be included in the analysis, as they did not provide any information on TD (and we were unable to obtain such information from the authors). This number of studies signifies a tremendous loss of information. Importantly, 32 of these studies reported on systematic screening in their methods, but did not provide information on TD cases. Instead, reporting of continuous sum scores of various dyskinesia-rating scales is common, but not informative, as has been discussed before[10,11].

Second, meta-analyses cannot estimate the risk of non-occurrence. Thus, studies that did not observe any cases of TD in either arm did not contribute to risk calculations. They contribute to RARs though, thus their effect is reflected in the NNTs. On the one hand, this is an unfortunate shortcoming of the methodology; on the other hand, the data quality of large studies in adults with chronic mental illness without a single case of probable incident TD may also be questionable. Altogether, five studies including four treatment arm pairs for FGA-SGA comparisons[22,26,28,45] and two treatment arm pairs for SGA-SGA comparisons[22,52] did not contribute to the meta-analytic risk estimates. Subsequently, the estimated raw TD rate is likely slightly lower under controlled study conditions, and similarly the observed differences may also be minimally lower.

Third, as the minimal exposure duration before TD can be diagnosed is >3 months[6,7], this time frame was chosen during title/abstract screening for eligible studies. However, within the time bin of studies lasting <6 months, only one (4-month) study formally reported treatment-emergent dyskinesia. We, thus, refrained from systematically requesting data on TD from studies lasting <6 months (unless a formal screening for dyskinesia was part of the study design). For all studies lasting ≥6 months, data were requested (even if the study design did not include formal TD screening). It is often argued that shorter observation periods are contaminated by high rates of withdrawal dyskinesia. While this argument can only be tested on a case-by-case evaluation of emerging dyskinesia cases, the observation of significantly increasing hyperkinesias during the early treatment phase with haloperidol, but not with risperidone[74], argues against this notion. The median study duration was one year, but longer observation periods would be desirable.

Fourth, most studies reported probable mild TD, likely providing an overestimate of clinically relevant cases. While this is a cautious, safety-oriented and well-recognized strategy, it would be important to learn about the rates of persistent TD along with scale-based severity measures. On the other, conservative end of the sensitivity spectrum, some studies, and increasingly more in SGA-SGA comparator trials, only provide rates of clinically reported TD cases, not relying on rating scale information. Our moderator analysis did not identify a systematic influence of screening techniques on RRs or RaRs, but differences in primary data acquisition need to be considered when interpreting the results.

Fifth, the prevalence of TD in FGA-naïve cohorts treated with SGAs has been reported to be lower than in subjects with a history of FGA exposure[11]. There was, however, not a single head-to-head RCT to test the effect of prior FGA exposure on TD rates in the FGA-SGA comparison, and relatively sparse information on this issues in the SGA-SGA comparisons.

Sixth, as the absolute TD risk gets lower (in SGA studies in general, relative to all FGA studies), the influence of potential co-factors, such as co-treatment with other potentially TD-causing medications (e.g., metoclopramide[75,76], flunarizine[77], or antidepressants[78]) may represent a confound that has not been addressed in the primary sources.

EPS have been characterized as the most important factor predicting non-adherence during FGA treatment[79]. By contrast, akathisia, parkinsonism and dyskinesia failed to predict adherence during the EUFEST and CATIE studies[80], but predicted adherence in another large study including more than 2,000 patients treated with SGAs or FGAs[81]. Considering the high incidence of clinically meaningful weight gain during SGA treatment[82], EPS partly lost their weight in clinical decision-making. Nevertheless, TD remains a clinical problem in movement disorders clinics: SGA-related tardive syndromes lately outnumbered FGA-related TD cases in a movement disorders clinic[83], likely due to increasing on-label, as well as off-label use of SGAs.

CONCLUSIONS

In conclusion, the historic notion of an SGA advantage with regard to TD risk, which had been challenged by CATIE and CUtLASS, has been confirmed.

Importantly, the quality of data acquisition and reporting of TD needs to return to earlier standards in future comparative studies to understand whether or not SGAs differ among one another with regard to TD risk. Risk estimates should span incidence, severity and persistence as well as impact on functioning and quality of life. Ideally, patient-level data repositories should help to accurately estimate rare side effects in antipsychotic studies.

ACKNOWLEDGEMENTS

The authors thank C. Arango, L. Citrome, M.R.K. Dossenbach, R. Rosenheck, S. W. Lewis, K.P. Hayhurst, P. Tyrer, S. Marder, D. Naber, S.G. Potkin, A. Parabiaghi, S.E. Purdon, A.I. Green, R.S. Keefe, R. Kongsakon, G. Gründer, B.J. Kinon, M.J. Cuesta, E. Johnsen, and the Lilly data source management team for providing unpublished data, and for clarifying published data.

REFERENCES

1. Yassa R, Jones BD. Complications of tardive dyskinesia: a review. Psychosomatics 1985;26:305-7.
2. Strassnig M, Rosenfeld A, Harvey PD. Tardive dyskinesia: motor system impairments, cognition and everyday functioning. CNS Spectr 2017;7:1-8.
3. Correll CU, Leucht S, Kane JM. Lower risk for tardive dyskinesia associated with second-generation antipsychotics: a systematic review of 1-year studies. Am J Psychiatry 2004;161:414-25.
4. Peluso MJ, Lewis SW, Barnes TR et al. Extrapyramidal motor side-effects of first- and second-generation antipsychotic drugs. Br J Psychiatry 2012;200: 387-92.
5. Miller DD, Caroff SN, Davis SM et al. Extrapyramidal side-effects of antipsychotics in a randomised trial. Br J Psychiatry 2008;193:279-88.
6. Schooler NR, Kane JM. Research diagnoses for tardive dyskinesia. Arch Gen Psychiatry 1982;39:486-7.
7. American Psychiatric Association. Diagnostic and statistical manual of mental disorders, 5th ed. Arlington: American Psychiatric Publishing, 2013.
8. Komossa K, Rummel-Kluge C, Hunger H et al. Olanzapine versus other atypical antipsychotics for schizophrenia. Cochrane Database Syst Rev 2010; 3:CD006654.
9. Komossa K, Rummel-Kluge C, Schmid F et al. Quetiapine versus other atypical antipsychotics for schizophrenia. Cochrane Database Syst Rev 2010;1:CD006625.
10. Haddad PM, Das A, Keyhani S et al. Antipsychotic drugs and extrapyramidal side effects in first episode psychosis: a systematic review of head-head comparisons. J Psychopharmacol 2012;26:15-26.
11. Carbon M, Hsieh CH, Kane JM et al. Tardive dyskinesia prevalence in the period of second-generation antipsychotic use: a meta-analysis. J Clin Psychiatry 2017;78:e264-78.
12. Jankelowitz SK. Treatment of neurolept-induced tardive dyskinesia. Neuropsychiatr Dis Treat 2013;9:1371-80.
13. Leucht S, Kissling W, Davis JM. How to read and understand and use systematic reviews and meta-analyses. Acta Psychiatr Scand 2009;119:443-50.
14. Leucht S, Samara M, Heres S et al. Dose equivalents for antipsychotic drugs: the DDD method. Schizophr Bull 2016;42(Suppl. 1):S90-4.
15. Beasley CM, Dellva MA, Tamura RN et al. Randomised double-blind comparison of the incidence of tardive dyskinesia in patients with schizophrenia during long-term treatment with olanzapine or haloperidol. Br J Psychiatry 1999;174:23-30.
16. Bouchard RH. Longitudinal comparative study of risperidone vs classical neuroleptics in the treatment of schizophrenia: 24 months of observation. Encéphale 2002;28:S31-2.
17. Carrière P, Bonhomme D, Lempérière T. Amisulpride has a superior benefit/risk profile to haloperidol in schizophrenia: results of a multicentre, double-blind study (the Amisulpride Study Group). Eur Psychiatry 2000; 15:321-9.
18. Colonna L, Saleem P, Dondey-Nouvel L et al. Long-term safety and efficacy of amisulpride in subchronic or chronic schizophrenia. Amisulpride Study Group. Int Clin Psychopharmacol 2000;15:13-22.
19. Csernansky JG, Mahmoud R, Brenner R et al. A comparison of risperidone and haloperidol for the prevention of relapse in patients with schizophrenia. N Engl J Med 2002;346:16-22.
20. de Jesus Mari J, Lima MS, Costa AN et al. The prevalence of tardive dyskinesia after a nine month naturalistic randomized trial comparing olanzapine with conventional treatment for schizophrenia and related disorders. Eur Arch Psychiatry Clin Neurosci 2004;254:356-61.
21. Dossenbach MR, Folnegovic-Smalc V, Hotujac L et al. Double-blind, randomized comparison of olanzapine versus fluphenazine in the long-term treatment of schizophrenia. Prog Neuropsychopharmacol Biol Psychiatry 2004;28:311-8.
22. Findling RL, Johnson JL, McClellan J et al. Double-blind maintenance safety and effectiveness findings from the Treatment of Early-Onset Schizophrenia Spectrum (TEOSS) study. J Am Acad Child Adolesc Psychiatry 2010; 49:583-94.
23. Gaebel W, Riesbeck M, Wolwer W et al. Maintenance treatment with risperidone or low-dose haloperidol in first-episode schizophrenia: 1-year results of a randomized controlled trial within the German Research Network on Schizophrenia. J Clin Psychiatry 2007;68:1763-74.
24. Gharabawi GM, Bossie CA, Zhu Y. New-onset tardive dyskinesia in patients with first-episode psychosis receiving risperidone or haloperidol. Am J Psychiatry 2006;163:938-9.
25. Girgis RR, Phillips MR, Li X et al. Clozapine v. chlorpromazine in treatment-naive, first-episode schizophrenia: 9-year outcomes of a randomised clinical trial. Br J Psychiatry 2011;199:281-8.
26. Glick ID, Marder SR. Long-term maintenance therapy with quetiapine versus haloperidol decanoate in patients with schizophrenia or schizoaffective disorder. J Clin Psychiatry 2005;66:638-41.
27. Green AI, Lieberman JA, Hamer RM et al. Olanzapine and haloperidol in first episode psychosis: two-year data. Schizophr Res 2006;86:234-43.
28. Grunder G, Heinze M, Cordes J et al. Effects of first-generation antipsychotics versus second-generation antipsychotics on quality of life in schizophrenia: a double-blind, randomised study. Lancet Psychiatry 2016;3:717-29.
29. Hirsch SR, Kissling W, Bauml J et al. A 28-week comparison of ziprasidone and haloperidol in outpatients with stable schizophrenia. J Clin Psychiatry 2002;63:516-23.
30. Kahn RS, Fleischhacker WW, Boter H et al. Effectiveness of antipsychotic drugs in first-episode schizophrenia and schizophreniform disorder: an open randomised clinical trial. Lancet 2008;371:1085-97.
31. Keefe RS, Young CA, Rock SL et al. One-year double-blind study of the neurocognitive efficacy of olanzapine, risperidone, and haloperidol in schizophrenia. Schizophr Res 2006;81:1-15.
32. Kinon BJ, Kollack-Walker S, Jeste D et al. Incidence of tardive dyskinesia in older adult patients treated with olanzapine or conventional antipsychotics. J Geriatr Psychiatry Neurol 2015;28:67-79.
33. Kongsakon R, Trinidad-Onate P, Chaudhry HR et al. Asian outpatients with schizophrenia: a double-blind randomized comparison of quality of life and clinical outcomes for patients treated with olanzapine or haloperidol. J Med Assoc Thai 2006;89:1157-70.
34. Lewis SW, Davies L, Jones PB et al. Randomised controlled trials of conventional antipsychotic versus new atypical drugs, and new atypical drugs versus clozapine, in people with schizophrenia responding poorly to, or intolerant of, current drug treatment. Health Technol Assess 2006;10:1-165.
35. Marder SR, Glynn SM, Wirshing WC et al. Maintenance treatment of schizophrenia with risperidone or haloperidol: 2-year outcomes. Am J Psychiatry 2003;160:1405-12.
36. 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.
37. Meltzer HY, Bobo WV, Lee MA et al. A randomized trial comparing clozapine and typical neuroleptic drugs in non-treatment-resistant schizophrenia. Psychiatry Res 2010;177:286-93.
38. Miller DD, Eudicone JM, Pikalov A et al. Comparative assessment of the incidence and severity of tardive dyskinesia in patients receiving aripiprazole or haloperidol for the treatment of schizophrenia: a post hoc analysis. J Clin Psychiatry 2007;68:1901-6.
39. Parabiaghi A, Tettamanti M, D'Avanzo B et al. Metabolic syndrome and drug discontinuation in schizophrenia: a randomized trial comparing aripiprazole olanzapine and haloperidol. Acta Psychiatr Scand 2016;133:63-75.
40. Potkin SG, Weiden PJ, Loebel AD et al. Remission in schizophrenia: 196-week, double-blind treatment with ziprasidone vs. haloperidol. Int J Neuropsychopharmacol 2009;12:1233-48.
41. Purdon SE, Jones BD, Stip E et al. Neuropsychological change in early phase schizophrenia during 12 months of treatment with olanzapine, risperidone, or haloperidol. The Canadian Collaborative Group for research in schizophrenia. Arch Gen Psychiatry 2000;57:249-58.
42. Rosenheck R, Cramer J, Xu W et al. A comparison of clozapine and haloperidol in hospitalized patients with refractory schizophrenia. Department of Veterans Affairs Cooperative Study Group on Clozapine in Refractory Schizophrenia. N Engl J Med 1997;337:809-15.
43. Tohen M, Baker RW, Altshuler LL et al. Olanzapine versus divalproex in the treatment of acute mania. Am J Psychiatry 2002;159:1011-7.
44. Tollefson GD, Beasley CM Jr, Tran PV et al. Olanzapine versus haloperidol in the treatment of schizophrenia and schizoaffective and schizophreniform disorders: results of an international collaborative trial. Am J Psychiatry 1997;154:457-65.
45. Tyrer P, Oliver-Africano PC, Ahmed Z et al. Risperidone, haloperidol, and placebo in the treatment of aggressive challenging behaviour in patients with intellectual disability: a randomised controlled trial. Lancet 2008;371:57-63.
46. Tollefson GD, Beasley CM Jr, Tamura RN et al. Blind, controlled, long-term study of the comparative incidence of treatment-emergent tardive dyskinesia with olanzapine or haloperidol. Am J Psychiatry 1997;154:1248-54.

47. Buchanan RW, Panagides J, Zhao J et al. Asenapine versus olanzapine in people with persistent negative symptoms of schizophrenia. J Clin Psychopharmacol 2012;32:36-45.
48. Arango C, Robles O, Parellada M et al. Olanzapine compared to quetiapine in adolescents with a first psychotic episode. Eur Child Adolesc Psychiatry 2009;18:418-28.
49. Breier A, Berg PH, Thakore JH et al. Olanzapine versus ziprasidone: results of a 28-week double-blind study in patients with schizophrenia. Am J Psychiatry 2005;162:1879-87.
50. Ciudad A, Alvarez E, Bousono M et al. Safety and tolerability of olanzapine versus risperidone: a one-year randomized study in outpatients with schizophrenia with prominent negative symptoms. Actas Esp Psiquiatr 2007;35:105-14.
51. Cuesta MJ, Garcia de Jalon E, Campos MS et al. Cognitive effectiveness of olanzapine and risperidone in first-episode psychosis. Br J Psychiatry 2009;194:439-45.
52. Johnsen E, Jorgensen HA, Kroken RA et al. Neurocognitive effectiveness of quetiapine, olanzapine, risperidone, and ziprasidone: a pragmatic, randomized trial. Eur Psychiatry 2013;28:174-84.
53. Kane JM, Osuntokun O, Kryzhanovskaya LA et al. A 28-week, randomized, double-blind study of olanzapine versus aripiprazole in the treatment of schizophrenia. J Clin Psychiatry 2009;70:572-81.
54. Keks NA, Ingham M, Khan A et al. Long-acting injectable risperidone v. olanzapine tablets for schizophrenia or schizoaffective disorder. Randomised, controlled, open-label study. Br J Psychiatry 2007;191:131-9.
55. Kinon BJ, Lipkovich I, Edwards SB et al. A 24-week randomized study of olanzapine versus ziprasidone in the treatment of schizophrenia or schizoaffective disorder in patients with prominent depressive symptoms. J Clin Psychopharmacol 2006;26:157-62.
56. Kinon BJ, Noordsy DL, Liu-Seifert H et al. Randomized, double-blind 6-month comparison of olanzapine and quetiapine in patients with schizophrenia or schizoaffective disorder with prominent negative symptoms and poor functioning. J Clin Psychopharmacol 2006;26:453-61.
57. McIntyre RS, Cohen M, Zhao J et al. Asenapine for long-term treatment of bipolar disorder: a double-blind 40-week extension study. J Affect Disord 2010;126:358-65.
58. Schoemaker J, Naber D, Vrijland P et al. Long-term assessment of asenapine vs. olanzapine in patients with schizophrenia or schizoaffective disorder. Pharmacopsychiatry 2010;43:138-46.
59. Tran PV, Hamilton SH, Kuntz AJ et al. Double-blind comparison of olanzapine versus risperidone in the treatment of schizophrenia and other psychotic disorders. J Clin Psychopharmacol 1997;17:407-18.
60. Lewis SW, Barnes TR, Davies L et al. Randomized controlled trial of effect of prescription of clozapine versus other second-generation antipsychotic drugs in resistant schizophrenia. Schizophr Bull 2006;32:715-23.
61. McEvoy JP, Lieberman JA, Stroup TS et al. Effectiveness of clozapine versus olanzapine, quetiapine, and risperidone in patients with chronic schizophrenia who did not respond to prior atypical antipsychotic treatment. Am J Psychiatry 2006;163:600-10.
62. Tollefson GD, Birkett MA, Kiesler GM et al. Double-blind comparison of olanzapine versus clozapine in schizophrenic patients clinically eligible for treatment with clozapine. Biol Psychiatry 2001;49:52-63.
63. Addington DE, Labelle A, Kulkarni J et al. A comparison of ziprasidone and risperidone in the long-term treatment of schizophrenia: a 44-week, double-blind, continuation study. Can J Psychiatry 2009;54:46-54.
64. Chengappa KNR, Turkin SR, Schlicht PJ et al. A pilot, 15-month, randomised effectiveness trial of risperidone long-acting injection (RLAI) versus oral atypical antipsychotic agents (AAP) in persons with bipolar disorder. Acta Neuropsychiatrica 2010;22:68-80.
65. Citrome L, Cucchiaro J, Sarma K et al. Long-term safety and tolerability of lurasidone in schizophrenia: a 12-month, double-blind, active-controlled study. Int Clin Psychopharmacol 2012;27:165-76.
66. Gaebel W, Schreiner A, Bergmans P et al. Relapse prevention in schizophrenia and schizoaffective disorder with risperidone long-acting injectable vs quetiapine: results of a long-term, open-label, randomized clinical trial. Neuropsychopharmacology 2010;35:2367-77.
67. Macfadden W, Ma YW, Thomas Haskins J et al. A prospective study comparing the long-term effectiveness of injectable risperidone long-acting therapy and oral aripiprazole in patients with schizophrenia. Psychiatry 2010;7:23-31.
68. Naber D, Hansen K, Forray C et al. Qualify: a randomized head-to-head study of aripiprazole once-monthly and paliperidone palmitate in the treatment of schizophrenia. Schizophr Res 2015;168:498-504.
69. Seeman P. Atypical antipsychotics: mechanism of action. Can J Psychiatry 2002;47:27-38.
70. Leucht S, Pitschel-Walz G, Engel RR et al. Amisulpride, an unusual "atypical" antipsychotic: a meta-analysis of randomized controlled trials. Am J Psychiatry 2002;159:180-90.
71. Tenback DE, van Harten PN, Slooff CJ et al. Effects of antipsychotic treatment on tardive dyskinesia: a 6-month evaluation of patients from the European Schizophrenia Outpatient Health Outcomes (SOHO) Study. J Clin Psychiatry 2005;66:1130-3.
72. Woods SW, Morgenstern H, Saksa JR et al. Incidence of tardive dyskinesia with atypical versus conventional antipsychotic medications: a prospective cohort study. J Clin Psychiatry 2010;71:463-74.
73. Leucht S, Cipriani A, Spineli L et al. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. Lancet 2013;382:951-62.
74. Emsley RA. Risperidone in the treatment of first-episode psychotic patients: a double-blind multicenter study. Risperidone Working Group. Schizophr Bull 1999;25:721-9.
75. Kenney C, Hunter C, Davidson A et al. Metoclopramide, an increasingly recognized cause of tardive dyskinesia. J Clin Pharmacol 2008;48:379-84.
76. Matson JL, Mayville EA, Bielecki J et al. Tardive dyskinesia associated with metoclopramide in persons with developmental disabilities. Res Dev Disabil 2002;23:224-33.
77. Chouza C, Scaramelli A, Caamano JL et al. Parkinsonism, tardive dyskinesia, akathisia, and depression induced by flunarizine. Lancet 1986;1:1303-4.
78. Vandewalle W, Boon E, Sienaert P. Movement disorders due to modern antidepressants and mood stabilizers. Tijdschr Psychiatr 2015;57:132-7.
79. Van Putten T. Why do schizophrenic patients refuse to take their drugs? Arch Gen Psychiatry 1974;31:67-72.
80. Czobor P, Van Dorn RA, Citrome L et al. Treatment adherence in schizophrenia: a patient-level meta-analysis of combined CATIE and EUFEST studies. Eur Neuropsychopharmacol 2015;25:1158-66.
81. Garcia-Cabeza I, Gomez JC, Sacristan JA et al. Subjective response to antipsychotic treatment and compliance in schizophrenia. A naturalistic study comparing olanzapine, risperidone and haloperidol (EFESO Study). BMC Psychiatry 2001;1:7.
82. Manu P, Dima L, Shulman M et al. Weight gain and obesity in schizophrenia: epidemiology, pathobiology, and management. Acta Psychiatr Scand 2015;132:97-108.
83. Zutshi D, Cloud LJ, Factor SA. Tardive syndromes are rarely reversible after discontinuing dopamine receptor blocking agents: experience from a university-based movement disorder clinic. Tremor Other Hyperkinet Mov 2014;4:266.

DOI:10.1002/wps.20579

Management of common adverse effects of antipsychotic medications

T. Scott Stroup[1], Neil Gray[2]

[1]New York State Psychiatric Institute, Columbia University College of Physicians and Surgeons, New York, NY, USA; [2]Department of Psychiatry, John A. Burns School of Medicine, University of Hawaii, Honolulu, HI, USA

The benefits of antipsychotic medications are sometimes obscured by their adverse effects. These effects range from relatively minor tolerability issues (e.g., mild sedation or dry mouth) to very unpleasant (e.g., constipation, akathisia, sexual dysfunction) to painful (e.g., acute dystonias) to disfiguring (e.g., weight gain, tardive dyskinesia) to life-threatening (e.g., myocarditis, agranulocytosis). Importantly, adverse effect profiles are specific to each antipsychotic medication and do not neatly fit into first- and second-generation classifications. This paper reviews management strategies for the most frequent side effects and identifies common principles intended to optimize net antipsychotic benefits. Only use antipsychotics if the indication is clear; only continue antipsychotics if a benefit is discernible. If an antipsychotic is providing substantial benefit, and the adverse effect is not life-threatening, then the first management choice is to lower the dose or adjust the dosing schedule. The next option is to change the antipsychotic; this is often reasonable unless the risk of relapse is high. In some instances, behavioral interventions can be tried. Finally, concomitant medications, though generally not desirable, are necessary in many instances and can provide considerable relief. Among concomitant medication strategies, anticholinergic medications for dystonias and parkinsonism are often effective; beta-blockers and anticholinergic medications are useful for akathisia; and metformin may lead to slight to moderate weight loss. Anticholinergic drops applied sublingually reduce sialorrhea. Usual medications are effective for constipation or dyslipidemias. The clinical utility of recently approved treatments for tardive dyskinesia, valbenazine and deutetrabenazine, is unclear.

Key words: Antipsychotics, adverse effects, schizophrenia, akathisia, tardive dyskinesia, parkinsonism, dystonias, impulse control disorders, sialorrhea, sedation, sexual function, orthostatic hypotension, neuroleptic malignant syndrome, metabolic effects, agranulocytosis

(World Psychiatry 2018;17:341–356)

Antipsychotics are the first-line evidence-based treatment for schizophrenia and other primary psychotic disorders. Some antipsychotics are also approved for treatment of bipolar disorder, treatment-resistant depression, autism, or Tourette's disorder. In addition, these medications are prescribed off-label for individuals with other conditions, such as borderline personality disorder, obsessive-compulsive disorder, anorexia nervosa, insomnia, delirium, and various dementia syndromes including Alzheimer's disease. The utility of these drugs is hampered by their adverse effects, which must be weighed against their variable benefits for these conditions.

In persons with schizophrenia, antipsychotic medications often provide dramatic symptomatic relief for hallucinations and delusions, and improvement for disorganized thoughts and behavior. However, because they are associated with a multitude of adverse effects, some of which are medically serious and many of which affect patient attitudes toward treatment, discussions about these medications are often dominated by their side effects rather than their benefits. This is highlighted by the fact that experts and guidelines commonly recommend choosing antipsychotic medications based on side effect profiles, which vary considerably, rather than efficacy, which is considered to be similar[1,2]. For non-psychotic disorders and for off-label uses, for which the evidence of antipsychotic benefits is often unclear, side effects are vitally important, because the ratio of benefits to risks is lower and significantly influences the decision to use these medications.

Risk-benefit assessments about whether to prescribe an antipsychotic medication for an individual should be made according to specific drugs (as opposed to "generation" or "class" of drug) and the specific situation (i.e., actual benefits and harms expected or experienced by an individual). Because the benefits of antipsychotics are sometimes obscured by the adverse effects and medical risks, understanding how such problems can be avoided and successfully managed is essential to optimize the use of these important but sometimes controversial medications.

RISKS AND SIDE EFFECTS OVERVIEW

The adverse effects of antipsychotic medications range from relatively minor tolerability issues (e.g., mild sedation or dry mouth) to very unpleasant (e.g., constipation, akathisia, sexual dysfunction) to painful (e.g., acute dystonias) to disfiguring (e.g., weight gain, tardive dyskinesia) to life threatening (e.g., myocarditis, agranulocytosis). Some adverse effects have little short-term clinical implications (e.g., increased prolactin or serum lipid levels), but may involve long-term risk of medical complications.

Each antipsychotic medication has a unique side effect profile, which affects individuals differently. Because the incidence of the side effects varies considerably across the large number of antipsychotic medications, we provide Table 1, which estimates the relative liability of commonly used drugs to cause specific adverse effects. The table demonstrates that the drugs' profiles do not adhere closely to first- and second-generation classifications of antipsychotics. With the important exception of tardive dyskinesia, which is more common among patients treated with older (first-generation) medications such as chlorpromazine and haloperidol, no adverse effect is class-specific. Weight gain is not unique to newer drugs,

Table 1 Side effect profiles of selected antipsychotic drugs

Adverse effects	AMI	ARI	CPZ	CLO	HAL	LUR	OLA	PAL	PER	QUE	RIS	SER	ZIP
Anticholinergic effects	0	0	++	+++	0	0	++	0	0/+	+/++	0	0	0
Acute parkinsonism	+	+	+	0	+++	+/++	0/+	++	++	0	++	0/+	+
Akathisia	+	++	+	+	+++	+/++	+	+	++	+	+	+	+/++
Tardive dyskinesia	0/+	0/+	++	0	++	0/+	0/+	0/+	++	0/+	0/+	0/+	0/+
Diabetes	0/+	0/+	+++	+++	0/+	0/+	+++	+	+	++	+	+	0/+
Weight gain	0/+	0/+	+++	+++	+	0/+	+++	++	++	++	++	++	0/+
Increased lipids	+	0/+	+++	++	0/+	0/+	+++	+	+	++	+	+	0/+
Sialorrhea	0	0	0	++	0	0	0	0	0	0	0	0	0
Neutropenia	0/+	0/+	0/+	+++	0/+	0/+	0/+	0/+	0/+	0/+	0/+	0/+	0/+
Orthostatic hypotension	0/+	0/+	++	++	0	0/+	+	+	+	++	+	++	0
Hyperprolactinemia	+++	0	+	+	++	+	+	+++	++	0	+++	+	+
Increased QTc interval	++	0/+	0/+	++	0+	0/+	0/+	+	+	+	+	++/+++	++
Sedation	0/+	0/+	++	+++	+	+/++	+/++	0/+	+	++ b	+	0/+	+
Seizures	0/+	0/+	0/+	++	0/+	0/+	0/+	0/+	0/+	0/+	0/+	0/+	0/+

AMI – amisulpride, ARI – aripiprazole, CPZ – chlorpromazine, CLO – clozapine, HAL – haloperidol, LUR – lurasidone, OLA – olanzapine, PAL – paliperidone, PER – perphenazine, QUE – quetiapine, RIS – risperidone, SER – sertindole, ZIP – ziprasidone, 0: none or equivocal, 0/+: minimal/rare, +: mild/sometimes occurs, ++: moderate/occurs frequently, +++: severe/occurs very often

nor is it present in all of the newer medications. Similarly, akathisia and parkinsonism are common with older drugs and some newer drugs. Several adverse effects – seizures, neutropenia, sialorrhea – are virtually unique to clozapine.

Some population groups respond distinctively to antipsychotics. For example, children, adolescents and the elderly are more likely to experience certain adverse effects or experience them more severely. Youth are more susceptible to weight gain and sedation, while the elderly are more vulnerable to consequences of orthostatic hypotension (falls) and anticholinergic effects (cognitive impairment). In addition, individuals vary considerably in their risk of side effects and how these effects are experienced.

PRINCIPLES FOR ANTIPSYCHOTIC PRESCRIBING

Before discussing the management of specific adverse effects, we propose some general principles for optimal prescribing of antipsychotic medications. First, only prescribe antipsychotics when a clear benefit can be expected and there is no safer or feasible alternative. Second, choose an antipsychotic based on the clinical situation and preferences of the patient (e.g., avoid medications that cause orthostatic hypotension in the elderly; avoid medications associated with substantial weight gain in patients who prioritize weight control; avoid QTc-prolonging drugs in patients with a history of heart disease, arrhythmia or syncope). Third, use the lowest effective dose of antipsychotic medication, which must be determined empirically for each individual. Fourth, discontinue the antipsychotic if there is no benefit. Often there is at least some benefit, signaling the need for an individualized risk-benefit assessment if there are side effects. Finally, monitor for known side effects regularly (see Table 2). The rest of this paper addresses what to do when adverse effects occur.

GENERAL STRATEGIES FOR MANAGING THE ADVERSE EFFECTS OF ANTIPSYCHOTICS

Antipsychotics that are not beneficial or are not required should be discontinued. The main strategies for managing adverse effects are as follows:

Lower the dose. This is relevant when the antipsychotic has provided benefit, and the adverse effect is dose-related and not medically urgent. Using the lowest dose that is effective at achieving treatment goals is widely recommended and reduces dose-related effects such as parkinsonism, sedation, hyperprolactinemia, orthostatic hypotension, and anticholinergic effects. In practice, finding the optimal, lowest effective dose is an individualized, empirical process that must balance the desires for maximal efficacy and minimal adverse effects[3].

Switch to an antipsychotic with a different adverse effect profile. Switching to a medication not likely to cause the problematic effect is a common strategy proven effective for at least some adverse effects, for example to address dyslipidemias or reduce weight[4,5]. Variability among antipsychotic medications in the risk for akathisia, parkinsonism and hyperprolactinemia makes switching an attractive approach to these problems, and evidence from observational and randomized trials sup-

Table 2 Suggested monitoring schedule for individuals taking antipsychotic medications

	Baseline	Each visit	During titration	At 3 months	Quarterly	Every 6 months	Annually
Weight	X		X		X		
Tardive dyskinesia (Abnormal Involuntary Movement Scale)	X		X			X	
Parkinsonism, akathisia	X		X				X
Glucose metabolism (fasting blood sugar, HbA1C)	X			X			X
Lipid metabolism (fasting lipids)	X			X			X
Blood pressure and pulse	X		X	X			X
Sexual/reproductive function	X		X				X
Sedation	X	X					
ECG (based on history and symptoms)	X						
Prolactin	If symptoms of hyperprolactinemia develop						

If taking clozapine, monitor for neutropenia, myocarditis and sialorrhea; if taking aripiprazole, cariprazine or brexpiprazole, monitor for impulse control disorders/behavioral addictions

ports this[4,5]. Switching is ideally done gradually rather than abruptly, to avoid symptom exacerbation and other rebound phenomena. A cross-titration completed within two to four weeks was adequate in one randomized controlled trial[5]. A risk when switching from an antipsychotic that has been effective is that the new medication may not be as efficacious; therefore patients undergoing switches should be monitored carefully for symptom exacerbations. Unless an individual has only responded to clozapine, switching antipsychotics is a preferred approach to deal with adverse effects that cannot be addressed with dosage adjustments.

Use a non-pharmacologic intervention. Non-pharmacologic interventions to reduce adverse effects are appealing but generally unavailable. Diet and exercise programs are modestly effective in addressing weight gain and related lipid abnormalities[6].

Treat with a concomitant medication. Using medications to manage antipsychotic side effects is a common but often suboptimal approach, because the beneficial effects of concomitant medications are often modest, they also may have adverse effects, and drug interactions may occur. For example, anticholinergic medications used to treat parkinsonism are associated with cognitive impairment and constipation. Further, few concomitant medication approaches are supported by evidence from randomized controlled trials.

In the following section, we describe common antipsychotic adverse effects and approaches to their prevention and management (see also Table 3). We focus on the most common and consequential adverse effects rather than the many possible but relatively rare effects. Our emphasis is on evidence-based management strategies, but in many instances the evidence is based on common sense and case reports rather than randomized controlled trials.

SPECIFIC ADVERSE EFFECTS

Neurologic side effects

Neurologic side effects known as extrapyramidal symptoms are prominent with antipsychotic medications, and the risk varies considerably among the individual antipsychotics, with high-potency drugs such as haloperidol carrying the greatest risk (Table 1). Principal manifestations include dystonias, akathisia and parkinsonism; tardive syndromes are discussed separately below. Dystonias are involuntary contractions of antagonistic muscle groups, leading to twisting, sustained and repetitive motions or abnormal postures, most commonly in the head, face and neck. These can be painful and highly distressing. Akathisia refers to a feeling of restlessness and tension that usually (but not always) compels the sufferer to near-constant motion, inducing dysphoria and even suicidality[7]. Parkinsonism includes a number of drug-induced symptoms resembling Parkinson's disease, such as bradykinesia, rigidity and tremor.

Dystonias typically occur within hours to days of antipsychotic administration or dose increase, almost always within the first five days[8]. Prevalence varies widely based on specific medication and risk factors[9]. A history of extrapyramidal side effects is the most significant risk factor, with a relative risk of about six[10]. Young age and male sex are also clear risk factors[10-12]. The two most concerning presentations are laryngospasm, which is rare but life-threatening[13], and oculogyric crisis, a highly painful and distressing tonic deviation of the eyes that can become recurrent or chronic[14].

Because dystonias are painful and highly distressing, prevention is the best management strategy. The mainstay of prophylaxis for dystonias is anticholinergic medication. Benztropine prophylaxis is effective for high-potency antipsycho-

Table 3 Common antipsychotic adverse effects and management strategies

Adverse effects	First choice	Second choice	Third choice	Others/Comments
Dystonias	Anticholinergic medication	Antihistaminic medication	Benzodiazepine	
Parkinsonism (tremor, rigidity, bradykinesia)	Lower dose	Change to antipsychotic with lower risk	Concomitant use of anticholinergic agent	
Akathisia	Lower dose	Change antipsychotic	Concomitant use of beta-blocker	Anticholinergics and benzodiazepines
Tardive dyskinesia	Lower dose	Valbenazine or deutetrabenazine	Gingko biloba or clonazepam	
Sialorrhea	Conservative approaches such as sugarless gum during day, towel over pillow at night	Anticholinergic drops (ipratropium or atropine) topically/sublingually		
Sedation	Dose at night before sleep	Lower dose	Change to less sedating antipsychotic	Stimulants have unclear benefit
Prolactin elevation, sexual side effects	Dose reduction	Change to a prolactin-sparing antipsychotic	Add aripiprazole	Phosphodiesterase inhibitors for sexual dysfunction
Orthostatic hypotension	Adjust dose or dosing schedule	Behavioral changes including adequate hydration	Change antipsychotic	Concomitant medication strategies are limited
QT prolongation	Change antipsychotic			Avoid other QT-prolonging agents
Neuroleptic malignant syndrome	Discontinue antipsychotic	Supportive measures including IV hydration and cooling	Dantrolene and bromocriptine	
Neutropenia/agranulocytosis	Discontinue clozapine or other causative agent	Colony-stimulating factors (e.g., filgastram)		
Impulse control disorders/ behavioral addictions	Change antipsychotic			
Myocarditis	Discontinue clozapine or other causative agent			
Weight gain, dyslipidemia	Behavioral modification (diet, exercise)	Change antipsychotic	Metformin	
Anticholinergic effects (dry mouth, blurry vision, tachycardia, constipation)	Lower dose	Change antipsychotic	Treat symptoms, e.g., constipation with osmotic agents, stimulant laxatives; tachycardia with beta-blocker	Limit other anticholinergic agents

tics[15-18], but may be ineffective for low-potency medications[19]. There is not yet consensus on when prophylaxis is indicated, but clinical judgment of relative risk of dystonias versus risk of anticholinergic side effects and polypharmacy suggests many cases in which prophylaxis is clearly indicated (e.g., a young male starting a high-potency antipsychotic). Prophylaxis should always be used if a patient is getting a first dose of a high-potency antipsychotic, such as haloperidol, by injection. Once a patient is on a stable dose of antipsychotic and anticholinergic, gradual withdrawal of the anticholinergic may be possible[20], though a quarter of patients may require reinstatement[21]. For patients who have dystonias on a high-potency antipsychotic, switching to a lower potency antipsychotic may reduce the risk of dystonias as well as parkinsonism[22].

In acute dystonic reactions requiring urgent treatment, intramuscular anticholinergics (e.g., biperiden 5 mg) or antihistaminics (e.g., diphenhydramine 50 mg) are indicated[9]. Multiple doses may be required for initial response, and are usually needed for 1-2 days to maintain response until the causative antipsychotic is cleared. Benzodiazepines are also thought to be effective in treating dystonias[23,24].

Parkinsonism typically presents insidiously over days to months[8]. In contrast to dystonias, risk of parkinsonism is greater in women and older patients[25]. Additional risk factors include pre-existing rigidity[26] and AIDS[27,28]. Treatment of psychosis in patients with Parkinson's disease is complex, and reviewed elsewhere[29-31].

In the treatment of antipsychotic-induced parkinsonism, reducing dose should be a first consideration[32]; switching from an antipsychotic with high risk to one of low risk can also be an effective strategy[33]. Concomitant medications are a third common approach that is useful if switching antipsychotics is not

desirable. Anticholinergic medications are useful in the treatment of parkinsonism, but this has not been thoroughly studied[34,35]; the risks of anticholinergic agents are greater in the elderly (who are more likely to be affected by parkinsonism). Benztropine, which is in common use, and ethopropazine, which may not be widely available, are anticholinergic medications known to be effective for parkinsonism[36,37]. Amantadine at 100-400 mg daily also has good support in the literature[36,38-40], and may be particularly helpful in elderly patients who need to avoid anticholinergic effects[32].

Akathisia typically develops gradually over days to weeks of treatment, though it can present more acutely[41]. There is not strong evidence for risk factors, other than current antipsychotic dose and rate of dose increment[42,43]. Akathisia occurs with many antipsychotics, with high-potency agents and aripiprazole being particularly prone to this side effect, while clozapine, olanzapine and quetiapine are low-risk[44,45].

Centrally-acting beta-adrenergic antagonists, primarily propranolol, have long been used as first-line therapy for akathisia with moderate efficacy[45,46], supported by multiple small placebo-controlled trials[47-50]. Orthostatic hypotension and bradycardia are significant drawbacks to beta-blockers. Anticholinergics such as benztropine have also been used clinically for akathisia, but their usefulness has not been demonstrated in a systematic way[51]. Anticholinergics may work best for akathisia when it co-occurs with parkinsonism[52].

Serotonergic treatments have gathered increasing attention for treatment of akathisia. The antidepressant mirtazapine at 15 mg/day has shown propranolol equivalency in several trials and seems to be well tolerated in the short term[43,50,51], though its potential to cause weight gain is a particular consideration among those receiving antipsychotics. The specific 5-HT2A/C antagonists mianserin and ritanserin have also shown efficacy in small open-label studies[52-58]. Zolmitriptan (a 5-HT1B/1D agonist) and cyproheptadine (which has 5-HT2 antagonism in addition to anticholinergic and antihistaminergic properties) were both found to be as effective as propranolol in small randomized trials[59,60].

Benzodiazepines are also commonly used to treat akathisia. In severe, acute cases, intravenous diazepam has produced rapid resolution of symptoms[61]. Clonazepam and lorazepam have shown utility in several small trials[62-64], with at least some evidence of a dose-response relationship. Further studies, particularly long-term trials given the tolerance that develops to these medications, are required[65]. A possible association of benzodiazepines with increased mortality rates in schizophrenia dampens enthusiasm for this approach[66].

Several other approaches to akathisia have been explored, but have very limited empirical support. High dose vitamin B6 (600 mg/day) was shown to provide subjective improvement in a small blinded trial[67], but this dose risks peripheral neuropathy in long-term treatment[68]. Clonidine has shown similar efficacy to beta-blockers, but with poorer tolerability[52]. Diphenhydramine has produced mixed results in small trials of akathisia induced by metoclopramide and prochlorperazine[69-71].

Tardive syndromes

Tardive dyskinesia is one of the most dreaded complications of antipsychotic treatment, though it may also occur with other medication classes[72]. It typically develops after months or years of exposure, and is characterized by involuntary athetoid or choreiform movements of the lower face, extremities and/or trunk muscles. Most commonly, these present as grimacing, lip-smacking/puckering, tongue movements, and excessive blinking. Most distressingly, symptoms persist long after the offending medication is discontinued, and may be permanent in some cases (dyskinesia lasting less than a month after withdrawal is considered a separate clinical entity, withdrawal dyskinesia). Other tardive manifestations may include akathisia, stereotypies, dystonias, parkinsonism, tremor, myoclonus, and tourettism[73].

Estimates of prevalence have varied, but a large systematic review of nearly 40,000 patients published in 1992 suggested that about 24% of those treated with antipsychotics had tardive dyskinesia[74]; the prevalence is thought to have declined since then due to the use of newer medications and more moderate dosing. Risk factors for the syndrome include early presence of extrapyramidal symptoms[75], and possibly African ethnicity and older age[72,74,76]. Female sex may also increase the risk[72,74], though there is conflicting evidence[76-78]. The early presence of extrapyramidal symptoms is a particularly useful risk factor, potentially allowing clinicians to reduce dose or switch antipsychotic before tardive dyskinesia is induced. There is an association of anticholinergic medication use with tardive dyskinesia which remains unexplained[77]; perhaps the presence of extrapyramidal symptoms explains this correlation.

Many studies have attempted to characterize genetic risk factors for tardive dyskinesia. In general, there have as yet been no findings conclusive enough, and with sufficient effect size, to warrant screening. Variations in catechol-O-methyltransferase[79,80], brain-derived neurotrophic factor (BDNF)[81], dopamine receptor 2[82], and manganese superoxide dismutase[79] genes have modest evidence for increasing risk. There is also mounting evidence that polymorphisms in genes involved in GABA and serotonergic signaling may confer risk[83-85]. It seems likely that, with continued effort, a clinically useful genetic screening test for tardive dyskinesia risk might be developed in the near future[83].

Newer (second-generation) antipsychotics are less likely to cause tardive dyskinesia[85], with annual incidence estimated at 3.9% (vs. 5.5% for first-generation drugs) in a review of twelve trials[86]. This differential of risk may be more pronounced in the elderly[87,88]. In a patient who has developed tardive dyskinesia on a first-generation antipsychotic, common clinical prac-

tice is to switch to a second-generation drug, but the empirical evidence to support this is weak; this has only been studied in small trials of risperidone and olanzapine[89-91]. Dosage reduction is also commonly recommended to prevent worsening of tardive dyskinesia, but again there is little evidence for this practice[92].

Many pharmaceutical strategies for tardive dyskinesia have been explored. Inhibitors of vesicular monoamine transporter 2 (VMAT2) are most notable: valbenazine was recently approved by the US Food and Drug Administration (FDA)[93]. The closely related medication tetrabenazine, approved for Huntington's disease but used off-label for a variety of hyperkinetic movement disorders, has also shown utility in treating tardive dyskinesia[94,95]. It is unclear to what extent these drugs differ in safety or efficacy[96]. Deutetrabenazine, an isotopic isomer of tetrabenazine, was also recently approved by the FDA as a treatment for tardive dyskinesia[97]. The impact of these new treatments is currently uncertain.

Most GABA agonists tested – including valproate, baclofen, progabide and tetrahydroisoxazolopyridine – have not shown any compelling benefit, and may worsen mental state[98]. However, clonazepam demonstrated moderate efficacy in one of the few double-blind randomized clinical trials for tardive dyskinesia[99]; tolerance developed to its antidyskinetic effect, but this could be restored by brief washout. Also of note, efficacy was more marked in those with primarily dystonic symptoms, as opposed to choreoathetoid dyskinesia.

A single fairly large randomized controlled trial found evidence that ginkgo biloba extract improved tardive dyskinesia symptoms and was well-tolerated[100]. This effect is possibly mediated by increases in BDNF[101]. Other supplement-based strategies include vitamin B6 (pyridoxal 5'-phosphate), with a recent meta-analysis providing weak but supportive evidence[102]. There is also weak evidence that vitamin E may protect against worsening of tardive dyskinesia, but this finding also requires further study[103].

A number of potential tardive dyskinesia treatments have very limited or conflicting evidence bases, including calcium channel blockers, other VMAT inhibitors such as reserpine, cholinergic and anticholinergic drugs, amantadine, and levetiracetam[104-106].

As a final resort, there is growing evidence that brain stimulation and surgical approaches may provide sustained relief of severe tardive dyskinesia, with particularly promising data for stimulation of the globus pallidus[107-109]. There have also been some case reports suggesting potential benefits of lesioning surgeries of the globus pallidus or thalamus[110].

Overall, a variety of treatment options exist for tardive dyskinesia but, with the exception of valbenazine and deutetrabenazine, none has met a level of clinical efficacy and safety sufficient to be approved by regulators. Prior to their development, the evidence-based guidelines of the American Academy of Neurology reported the strongest ("moderate") evidence of efficacy for clonazepam and ginkgo biloba[104].

Sialorrhea

Sialorrhea, the excessive production of saliva, is a side effect most commonly observed in patients treated with clozapine (possibly more than 90% of patients)[111], but can occur with other antipsychotics as well. It is believed to be related to actions on muscarinic and adrenergic receptors in the salivary glands[112,113]. It is often uncomfortable, embarrassing and stigmatizing, and can even result in aspiration pneumonia[114,115]. In some cases, painful swelling of the parotid can co-occur[116,117].

As with many antipsychotic side effects, using the lowest necessary dose and observing a gradual titration schedule are thought to minimize development of sialorrhea[118]. A number of treatments have been explored, principally antimuscarinic and alpha-adrenergic agents. Studies have focused almost exclusively on clozapine-induced sialorrhea[119], so the generalizability of findings to other antipsychotics is an open question.

Topical therapy with anticholinergics, typically by administering an ophthalmic or inhaler preparation sublingually, has been shown to improve symptoms. Atropine appears effective, though the short half-life limits its utility overnight[120-122]. Ipratropium has also shown good effect in several case studies [122-124], though a randomized controlled trial did not detect efficacy[125].

Among systemic antimuscarinic agents, there is evidence for efficacy of benztropine[21,126], trihexylphenidyl[127], glycopyrrolate[128], and pirenzepine[129,130]. Amitriptyline has also been tried in a small case series with promising results[131]. However, systemic antimuscarinic drugs present their own risks (confusion, blurred vision, constipation), which may be additive to clozapine's own anticholinergic effects.

Adrenergic agents also appear useful in antipsychotic-induced sialorrhea, though the mechanism is not clear. Clonidine has shown encouraging results in individual cases[132,133]. Another alpha-2 agonist, guanfacine, was effective in a single case[134]. The alpha-1 antagonist terazosin showed significant promise in a small trial[126], but has not subsequently been studied. Though these studies have not reported major side effects, the potential for worsening antipsychotic-induced orthostatic hypotension must be considered.

Several other pharmacologic strategies have been explored. The antipsychotics sulpiride and amisulpride have shown promising results in several small trials[135-137], as has the monoamine oxidase inhibitor moclobemide[136,138]. Finally, botulinum toxin injection has been shown to substantially improve antipsychotic-induced sialorrhea for 8-16 weeks[139,140].

If conservative, non-pharmacologic approaches are ineffective, we suggest that topical treatment with ipratropium or atropine be the initial approach to antipsychotic-induced sialorrhea, given the relative safety and tolerability. If these agents are ineffective, systemic medication can be used, selecting from the above-mentioned agents based on the patient's clinical picture (e.g., using clonidine in a patient with hypertension, benztropine in one with other extrapyramidal symptoms, amisulpride in one with resistant psychotic symptoms).

Sedation

All antipsychotic medications have been observed to cause sedation, but the severity and frequency vary widely among agents[141]. Sedation may be a causative factor in the increased risk for venous thromboembolism in patients treated with antipsychotics[142].

Although it is a common side effect and a frequently cited reason for medication non-adherence, the management of sedation has not been widely studied. Shifting dosing to nighttime, and reducing total daily dose, are the initially recommended approaches[143], followed by transitioning to a less sedating antipsychotic. Additionally, other sedating medications should be discontinued or changed when possible. The use of caffeine is also common, though it has not been systematically studied.

Stimulants and modafinil may improve cognitive and negative symptoms in schizophrenia[144], but relatively little research has focused on their potential utility in antipsychotic-induced sedation. In two cases, methylphenidate was reportedly useful and safe in treating patients with severe and unremitting sedation due to clozapine[145]. A small double-blind crossover study of methylphenidate did not specifically address antipsychotic-related sedation, but failed to find any benefit on a variety of clinical measures[146]. Moreover, methylphenidate has also been shown to worsen disorganization in patients with schizophrenia[147]. Likewise, despite case reports suggesting that modafinil may treat sedation[148], a systematic review of the literature found little or no evidence to support this[149], and a randomized controlled trial also found no significant effect[150]. A concern is that these medications may lead to worsening of movement disorders[151,152].

Prolactin, sexual function, and bone mineral density

Many antipsychotics can increase the release of prolactin, which can lead to a number of acute side effects: sexual dysfunction, anovulation, inappropriate lactation (galactorrhea), and gynecomastia. Antipsychotics can be imperfectly divided into prolactin-inducing and prolactin-sparing groups. The former include all first-generation antipsychotics, risperidone, paliperidone and amisulpride; the latter include clozapine, quetiapine, ziprasidone and aripiprazole[153]. Long-term hyperprolactinemia is also associated with decreased bone mineral density and osteoporosis[154].

Sexual dysfunction – including reduced libido, anorgasmia and erectile dysfunction – is common in patients taking antipsychotics[155,156] and must be monitored by prescribers. One measure to use is the Antipsychotics and Sexual Function Questionnaire[157]. Assessment of a patient with sexual dysfunction should include obtaining prolactin levels, reviewing other medications that may contribute, and ruling out potential comorbid causes[158]. Treatment strategies are largely dose reduction or switching to a prolactin-sparing antipsychotic (though sexual dysfunction is also common with clozapine and olanzapine)[159]. Evidence for specific symptom treatments (other than phosphodiesterase inhibitors for erectile dysfunction) is lacking[160].

Multiple studies have also identified an increased rate of osteopenia and osteoporosis in patients with schizophrenia[161,162]; however, multiple factors beyond antipsychotic use may contribute, including smoking, alcohol use, sedentary lifestyle, and poor nutrition[153]. Studies have shown that reduced bone mineral density and increased rate of hip fractures are associated with prolactin-inducing antipsychotics[163,164]. There has also been concern that elevated prolactin levels may be partly responsible for the observed increase in breast cancer rate among women with schizophrenia[165], though evidence is far from conclusive, due to multiple associated lifestyle and metabolic factors[166].

There is not yet a consensus on the appropriate monitoring for and management of hyperprolactinemia in people treated with antipsychotics[167-169]. In general, patients should be asked about baseline sexual dysfunction, menstrual irregularity and galactorrhea prior to initiation of antipsychotics. There is not a consensus on obtaining baseline prolactin levels. A conservative approach is to ask patients periodically about symptoms of hyperprolactinemia and to check the prolactin level in any patient developing symptoms. Another rational approach is to obtain a prolactin level at baseline and then approximately three months after starting an antipsychotic, as prolactin levels will have peaked by then[167].

Several specific populations are thought to be at particularly high risk for morbidity due to hyperprolactinemia and, if clinically feasible, should be placed on antipsychotics with minimal risk of raising prolactin levels[169]. First, in patients with established osteopenia or osteoporosis, a prolactin-sparing antipsychotic is obviously preferable. This may also apply to patients under the age of 25 who have not yet achieved peak bone mass, particularly women, who may be at increased risk for later osteoporosis[170]. Second, in female patients intending to become pregnant, a prolactin-sparing antipsychotic will be less likely to interfere with reproductive function. Third, and quite speculatively, in patients with a history of, or otherwise at elevated risk for breast cancer, there may be a greater danger of cancer or recurrence if treated with prolactin-elevating drugs[171].

Development of hyperprolactinemia in a patient on antipsychotics often presents a dilemma to the treating psychiatrist regarding further workup. If a baseline prolactin was obtained, and the elevation in prolactin appears clearly related to the antipsychotic, further workup is likely not necessary. More concerning signs include symptoms of pituitary disease (headaches, visual changes) and prolactin levels more than four times the upper limit of normal (>150 ng/mL), in which case evaluation by an endocrinologist and imaging (preferably magnetic resonance imaging) is warranted[167,169]. In cases of uncertainty (and where the risk of destabilizing the patient is low), a prolactin level assessment may be made after 3-4 days

off antipsychotics; a significant reduction in prolactin is reassuring that there is not an underlying pathology.

In cases of confirmed antipsychotic-induced hyperprolactinemia that are symptomatic, management is dose reduction or switch to a prolactin-sparing antipsychotic. If the clinical risk of dose reduction or switch is felt to be too high, an alternative strategy is to augment with aripiprazole, which has been shown to reduce prolactin levels in patients treated with risperidone[172]. A more experimental strategy is the use of dopamine agonists such as bromocriptine or cabergoline, which have been found to decrease prolactin and improve sexual function, though these may lead to worsening of psychotic symptoms[173,174].

An important but unanswered question is the role of bone density screening in patients on antipsychotics. The US Preventative Services Task Force recommends screening all women at age 65, while the US-based National Osteoporosis Foundation also recommends screening men over 70, as well as menopausal women with risk factors. Because individuals with schizophrenia often have multiple risk factors beyond antipsychotic use (e.g., smoking, obesity, diabetes), more aggressive screening is warranted than for the general population.

Orthostatic hypotension

All antipsychotics carry some risk of orthostatic hypotension, defined as a ≥20 mmHg drop in systolic or a ≥10 mmHg drop in diastolic blood pressure within three minutes of standing. Orthostatic hypotension can lead to dizziness, syncope, falls, and worsening of angina, and it should be evaluated by both history and measurement. Risk factors include systemic diseases causing autonomic instability (e.g., diabetes, alcohol dependence, Parkinson's disease), dehydration, drug-drug interactions, and age[175]. Chlorpromazine, sertindole, clozapine and quetiapine appear to have the greatest risk[176,177], and data suggest iloperidone is also high-risk[178]. Blockade of alpha-1 adrenoceptors and anticholinergic effects are believed to be the mechanism[179].

Switching to an antipsychotic that is rarely associated with orthostatic hypotension is a preferred management approach. Prevention of orthostatic hypotension relies on antipsychotic choice, gradual titration, and dosing distributed throughout the day (in order to minimize peak levels)[175]. Ample consumption of water and increased salt intake (supplementing 1-2 g/day), if not contraindicated, can reduce symptomatic hypotension[180]. Abdominal binders and leg compression stockings can reduce venous pooling and improve symptoms[181].

Pharmacologic treatment may be required in rare cases. Caffeine consumption may have a beneficial, mild pressor effect[180]. Fludrocortisone is widely used for treating orthostatic hypotension, and has been administered successfully in clozapine-associated orthostatic hypotension[182]; deleterious effects on blood sugar and electrolytes are a significant drawback, particularly in patients who already have metabolic side effects[175]. The alpha-1 agonist midodrine may also be considered[175,183], but has been linked to acute dystonias when combined with antipsychotics[184,185].

Sudden cardiac death and QT prolongation

Antipsychotics are associated with a 1.5 to 4-fold increase in risk of sudden cardiac death[186-189]. Risk factors include use of high dose or rapid administration, thioridazine or butyrophenone antipsychotics, and pre-existing hypertension or ischemic heart disease[188,190,191]. There are conflicting data for an association with age[188,192]. There is no evidence that second-generation antipsychotics are safer than first-generation drugs as a class[187].

The leading proposed mechanism is blockade of repolarizing potassium currents and prolongation of the QT interval, which are thought to lead to ventricular arrhythmias. Measurement of QT provides limited guidance in terms of risk; nevertheless, QTc >500 ms or an increase of 60 ms above baseline is regarded as a clear concern[193]. It is critical for the practitioner to consider all medications the patient is taking, as a diverse set of drugs cause QT prolongation[194]. A number of risk factors can make a modest QT prolongation dangerous, including bradycardia, hypokalemia, hypomagnesemia, congestive heart failure, atrial fibrillation, female gender, ion channel polymorphisms[194], and cocaine and chronic alcohol use[193].

Some experts argue that an electrocardiogram (ECG) should be obtained prior to, and shortly after, starting antipsychotic medications as a matter of course[195]. To support this view, they cite the significantly higher absolute risk of sudden cardiac death than clozapine-induced agranulocytosis, for which an extensive monitoring system is in place. Others recommend monitoring only with certain antipsychotics or when other risk factors are present[196]. The American Psychiatric Association's latest guidance recommends thorough physical exam and laboratory screening, with ECG indicated when thioridazine, ziprasidone, pimozide or mesoridazine are prescribed; family history of sudden cardiac death or long-QT syndrome are present; there is a personal history of syncope or known heart disease; or electrolyte abnormalities are present[197]. The UK National Health Service includes haloperidol, sertindole and pimozide as "high-risk" and requiring routine ECG, and recommends ECG if risk factors are present with "moderate-risk" drugs, including chlorpromazine, amisulpride, lurasidone, quetiapine, zotepine, promazine and melperone[198]. Patients who take more than one QT-prolonging drug warrant careful screening and monitoring.

Neuroleptic malignant syndrome

Neuroleptic malignant syndrome is one of the most dangerous adverse effects of antipsychotics. Hallmarks of the syndrome are fever, autonomic instability, rigidity and altered

mental status, associated with leukocytosis and elevated creatine phosphokinase. Mortality has been estimated at about 5%[199]. Neuroleptic malignant syndrome related to second-generation antipsychotics, particularly clozapine, may be less likely to present with signs of parkinsonism[200,201]. Incidence estimates vary widely, with the largest recent studies reporting rates of 0.02 to 0.04%[199,202,203]. The most important risk factor is a prior history of the syndrome. Pharmacologic risk factors include antipsychotic polypharmacy, high-potency antipsychotics, parenteral administration, rapid dose escalation, aripiprazole, lithium and benzodiazepine use[199,202,203]. Multiple medical comorbidities, heat exposure, dehydration, and the use of restraints are also associated with the syndrome[196,202-208].

Neuroleptic malignant syndrome is a medical emergency, often requiring intensive care. Evidence is from case reports rather than randomized clinical trials. For the psychiatrist, first steps are immediate withdrawal of all antipsychotics and related drugs (e.g., metoclopramide), cooling measures, and transfer to higher level of care[203]. Aggressive intravenous hydration and correction of electrolyte abnormalities are essential. Benzodiazepines may be helpful in treating the syndrome, and are preferable to physical restraint in agitated patients[200]. The skeletal muscle relaxant dantrolene and the D2-agonist bromocriptine are among first-line medications for moderate or severe neuroleptic malignant syndrome[199]. Electroconvulsive therapy has been successfully used in treatment-refractory cases[209].

Rechallenging a patient who has recovered from neuroleptic malignant syndrome with an antipsychotic is a clinical dilemma. The likelihood of recurrence is not well known, but likely in the range of 10-40%[210-212]. Longer delay from resolution of the syndrome to rechallenge is associated with reduced risk of recurrence[209]. In some cases, it may be necessary to treat a patient with electroconvulsive therapy to maintain symptom control for an extended period prior to rechallenge[213]. When reintroduction of an antipsychotic is necessary, it is prudent to select a drug with lower D2 potency (e.g., quetiapine or clozapine), pursue very gradual dose titration, and monitor closely.

Neutropenia/agranulocytosis

Neutropenia, the presence of too few infection-fighting neutrophils in the blood, and its extreme form, agranulocytosis, are most commonly associated with clozapine. These conditions and related increased susceptibility to infection are significant enough to warrant monitoring of granulocyte counts throughout a course of clozapine treatment. Clozapine has been associated with agranulocytosis ever since 16 cases, including eight deaths, were reported soon after the drug was introduced in Finland in 1975[214]. While many subsequent cases of clozapine-associated agranulocytosis have been reported[215], rare case reports with phenothiazines, including chlorpromazine, began appearing in the 1950s[216-218]. Case reports also implicate olanzapine[219] and risperidone[220]. About 3% of clozapine-treated patients will develop neutropenia; about 1% will develop agranulocytosis[221]. The risk for other antipsychotics is thought to be far lower.

The most important management strategy for neutropenia or agranulocytosis is early detection, which will prevent opportunistic infections. Because the period of highest risk is during the first months of treatment[215], neutrophil counts are measured more frequently in those months (weekly for 6 months in the US), then fortnightly for the remainder of the first year, and then monthly for the duration of treatment.

If neutropenia occurs, guidelines specify more frequent monitoring and when to interrupt treatment. For patients with stable but marginally adequate neutrophil counts, some clinicians use lithium to raise granulocyte counts above threshold levels to avoid increased monitoring requirements[222,223]. The mechanism by which lithium increases granulocyte counts is unknown[224].

Pharmaceutical versions of granulocyte colony-stimulating factor, a glycoprotein that induces bone marrow to produce and release granulocytes, may be used to treat agranulocytosis acutely[225,226]. For patients who have responded only to clozapine, such drugs may have a longer-term role in preventing agranulocytosis. For example, filgastram can be used over extended periods to maintain adequate neutrophil counts to avoid infections. Challenges for the use of filgastram include the need for parenteral administration and high cost.

Dose reduction is not an effective approach to clozapine-associated neutropenia[224]. Discontinuation of clozapine is the definitive solution to clozapine-induced neutropenia. This approach generally requires switching to another antipsychotic. For those patients who only responded to clozapine, clozapine re-challenge after agranulocytosis has not been successful, but case reports describe successful re-introduction of clozapine after neutropenia using either lithium or filgastram to increase neutrophil counts[227].

Behavioral addictions/impulse control disorders

Aripiprazole has been associated with the onset or exacerbation of impulse control disorders or behavioral addictions, including pathological gambling and compulsive eating, spending, shopping and sexual behaviors[228,229]. Because the dopamine agonists used to treat Parkinson's disease also cause impulse control disorders in a significant portion of patients, aripiprazole's partial dopamine agonist effect is presumed to be the mechanism[230-232]. Therefore, it is likely that other antipsychotics with dopamine agonist activity, such as cariprazine and brexpiprazole, may also have this effect.

The key to management of these compulsive behaviors is recognition that they are medication-induced and not simply part of an underlying mental or behavioral condition. In all reported cases, reducing the dose or discontinuing the causative medication was effective in ending the uncontrollable behav-

ior within weeks[228,233,234]. If an antipsychotic is necessary, one without dopamine agonist effects should be selected.

Myocarditis

Myocarditis, or inflammation of the heart muscle, is a rare but important medical risk of clozapine treatment that almost always occurs within the first two months of treatment[235,236]. Because myocarditis can progress quickly to cardiomyopathy and congestive heart failure, the best management strategy is to monitor for it, so that it can be recognized quickly. Slow titration may help[237,238]. At a minimum, patients initiating clozapine should be monitored weekly for signs and symptoms of myocarditis, including chest pain, dyspnea, orthopnea, peripheral edema, palpitations, fatigue, flu-like symptoms including fever, nausea and vomiting, and diaphoresis[239]. An ECG should be obtained and cardiac enzymes assessed as soon as myocarditis is suspected. Laboratory tests suggesting myocarditis in the context of recently started clozapine include elevated eosinophil count, C-reactive protein, sedimentation rate, and troponins. If myocarditis is suspected, an echocardiogram can assess ventricular and cardiac valve functioning; baseline echocardiograms are not necessary[239,240].

If a diagnosis of myocarditis is highly suspected or confirmed, then clozapine should be discontinued promptly, and general or specialty cardiac follow-up care is needed. In many cases cardiac function returns to normal after clozapine is stopped. Recurrence rates of clozapine-induced myocarditis are high; if the possible benefits of the drug are thought to justify this risk, it should be re-initiated in hospital with close monitoring[241].

Metabolic effects

Many antipsychotic medications are associated, to variable degrees, with weight gain, hypertension, and adverse effects on lipid and glucose metabolism.

Several antipsychotics are associated with significant weight gain, and virtually all antipsychotics are known to cause weight gain among youth[3]. Weight gain is among the most important antipsychotic side effects, because it is distressing to individuals and increases the risk of adverse health outcomes such as degenerative joint disease, type 2 diabetes mellitus and its complications, cardiovascular and cerebrovascular disease, as well as some types of cancer, and liver and kidney disease. Although weight gain commonly accompanies other adverse metabolic effects, adverse changes in lipids and insulin sensitivity may occur independently of weight gain[3].

Anyone taking an antipsychotic medication should be regularly monitored for metabolic side effects. If these effects occur, lifestyle modifications are widely recommended and are a reasonable first step for individuals taking antipsychotic medications. Several structured behavioral programs have been tested and found effective in individuals with severe mental illnesses[242-245]. Switching to an antipsychotic with lower risk for metabolic problems can be effective in helping individuals to lose weight and improve metabolic profiles[4,5].

Metabolic problems that develop in the context of successful antipsychotic treatment can also be treated symptomatically, as they are in the general population. For example, statins are used to treat dyslipidemias, and antihypertensive medications are used to treat hypertension. Metformin has repeatedly been shown in randomized controlled trials to be modestly effective in helping patients taking antipsychotics to lose weight, even if the weight gain was not recent[246-249]. Recently approved weight loss drugs – including lorcaserin, bupropion/naltrexone, and liraglutide – have not been tested specifically for antipsychotic-induced weight gain. Preliminary data on naltrexone alone suggests that it may be helpful[250]. Stimulant weight loss medications are not recommended due to their psychotogenic potential.

Anticholinergic effects

Anticholinergic side effects of antipsychotics include decreased salivation leading to dry mouth, decreased intestinal mobility leading to constipation, inhibition of visual accommodation leading to blurred vision, increased pupil size, and tachycardia[251]. These effects may lead to medical complications such as dental caries, ileus, and angina or myocardial infarction. Because increased pupil size can exacerbate narrow-angle glaucoma, this condition should be treated before initiating antipsychotic treatment; an antipsychotic with minimal anticholinergic effects should be selected. Similarly, prostatic hypertrophy should be treated and an antipsychotic with little anticholinergic effect should be used[251].

Decreasing the antipsychotic dose is the first-choice management strategy for anticholinergic side effects. Changing to a medicine with less anticholinergic effects may also be effective[251]. Finally, symptomatic management is a reasonable approach, but there is little evidence specific to antipsychotic-induced anticholinergic effects.

Constipation due to antipsychotics, particularly clozapine, can be severe and can lead to ileus[252,253]. Prevention and early recognition are critical. Recommended management strategies include adequate hydration; use of osmotic agents such as sorbitol, lactulose, or polyethylene glycol, and stimulant laxatives such as senna or bisacodyl. The benefits of stool softeners such as docusate sodium are unclear[4]. Bulk-forming, fiber-based laxatives are generally not recommended for slow-transit constipation such as that caused by anticholinergic effects[254].

CONCLUSIONS

The considerable benefits of antipsychotic medications are countered, to some extent, by their adverse effects. Appropri-

ate prevention and early management of these effects can enhance the net benefits of antipsychotics. Our review found that, in general, few management approaches are supported by strong empirical evidence; recommendations are often based at least in part on expert opinion.

Nevertheless, a few key principles apply broadly. Only use antipsychotics if the indication is clear; only continue antipsychotics if a benefit is discernible. If an antipsychotic is providing substantial benefit, and the adverse effect is not life-threatening, then the first management choice is to lower the dose or adjust the dosing schedule. Next is to change the antipsychotic; this is often reasonable unless the risk of relapse is high, such as when an individual has only responded to clozapine. In some instances, behavioral interventions can be tried. Finally, concomitant medications, though generally not desirable, are necessary in many instances.

Evidence suggests that adverse effects are not the main reason why individuals discontinue an antipsychotic medication[255]. Nevertheless, optimal management of adverse effects will improve patients' quality of life and their functional outcomes.

ACKNOWLEDGEMENTS

This paper was supported in part by the Global Mental Health Program at Columbia University. The authors thank S. Wetzler for her assistance.

REFERENCES

1. Goff DC. Maintenance treatment with long-acting injectable antipsychotics: comparing old with new. JAMA 2014;311:1973-4.
2. International College of Neuropsychopharmacology. CINP schizophrenia guidelines. http://cinp.org/cinp-schizophrenia-guidelines/.
3. Schimmelmann BG, Schmidt SJ, Carbon M et al. Treatment of adolescents with early-onset schizophrenia spectrum disorders: in search of a rational, evidence-informed approach. Curr Opin Psychiatry 2013;26:219-30.
4. Newcomer JW, Weiden PJ, Buchanan RW. Switching antipsychotic medications to reduce adverse event burden in schizophrenia: establishing evidence-based practice. J Clin Psychiatry 2013;74:1108-20.
5. Stroup TS, McEvoy JP, Ring KD et al. A randomized trial examining the effectiveness of switching from olanzapine, quetiapine, or risperidone to aripiprazole to reduce metabolic risk: comparison of antipsychotics for metabolic problems (CAMP). Am J Psychiatry 2011;168:947-56.
6. Brar JS, Ganguli R, Pandina G et al. Effects of behavioral therapy on weight loss in overweight and obese patients with schizophrenia or schizoaffective disorder. J Clin Psychiatry 2005;66:205-12.
7. Seemuller F, Lewitzka U, Bauer M. The relationship of akathisia with treatment emergent suicidality among patients with first-episode schizophrenia treated with haloperidol or risperidone. Pharmacopsychiatry 2012;45:292-6.
8. Tarsy D. Neuroleptic-induced extrapyramidal reactions: classification, description, and diagnosis. Clin Neuropharmacol 1983;6(Suppl. 1):S9-26.
9. van Harten PN, Hoek HW, Kahn RS. Acute dystonia induced by drug treatment. BMJ 1999;319:623-6.
10. Keepers GA, Casey DE. Use of neuroleptic-induced extrapyramidal symptoms to predict future vulnerability to side effects. Am J Psychiatry 1991; 148:85-9.
11. Aguilar EJ, Keshavan MS, Martinez-Quiles MD et al. Predictors of acute dystonia in first-episode psychotic patients. Am J Psychiatry 1994;151:1819-21.
12. Swett C Jr. Drug-induced dystonia. Am J Psychiatry 1975;132:532-4.
13. Koek RJ, Pi EH. Acute laryngeal dystonic reactions to neuroleptics. Psychosomatics 1989;30:359-64.
14. Gardner DM, Abidi S, Ursuliak Z et al. Incidence of oculogyric crisis and long-term outcomes with second-generation antipsychotics in a first-episode psychosis program. J Clin Psychopharmacol 2015;35:715-8.
15. Arana GW, Goff DC, Baldessarini RJ et al. Efficacy of anticholinergic prophylaxis for neuroleptic-induced acute dystonia. Am J Psychiatry 1988;145: 993-6.
16. Goff DC, Arana GW, Greenblatt DJ et al. The effect of benztropine on haloperidol-induced dystonia, clinical efficacy and pharmacokinetics: a prospective, double-blind trial. J Clin Psychopharmacol 1991;11:106-12.
17. Stern TA, Anderson WH. Benztropine prophylaxis of dystonic reactions. Psychopharmacology 1979;61:261-2.
18. Winslow RS, Stillner V, Coons DJ et al. Prevention of acute dystonic reactions in patients beginning high-potency neuroleptics. Am J Psychiatry 1986;143:706-10.
19. Swett C Jr, Cole JO, Shapiro S et al. Extrapyramidal side effects in chlorpromazine recipients: emergence according to benztropine prophylaxis. Arch Gen Psychiatry 1977;34:942-3.
20. Desmarais JE, Beauclair L, Annable L et al. Effects of discontinuing anticholinergic treatment on movement disorders, cognition and psychopathology in patients with schizophrenia. Ther Adv Psychopharmacol 2014;4:257-67.
21. Caradoc-Davies G, Menkes DB, Clarkson HO et al. A study of the need for anticholinergic medication in patients treated with long-term antipsychotics. Aust N Z J Psychiatry 1986;20:225-32.
22. Chan HY, Chang CJ, Chiang SC et al. A randomised controlled study of risperidone and olanzapine for schizophrenic patients with neuroleptic-induced acute dystonia or parkinsonism. J Psychopharmacol 2010;24:91-8.
23. Altamura AC, Mauri MC, Mantero M et al. Clonazepam/haloperidol combination therapy in schizophrenia: a double blind study. Acta Psychiatr Scand 1987;76:702-6.
24. Gagrat D, Hamilton J, Belmaker RH. Intravenous diazepam in the treatment of neuroleptic-induced acute dystonia and akathisia. Am J Psychiatry 1978;135:1232-3.
25. Thanvi B, Treadwell S. Drug induced parkinsonism: a common cause of parkinsonism in older people. Postgrad Med J 2009;85:322-6.
26. Caligiuri MP, Lohr JB. Instrumental motor predictors of neuroleptic-induced parkinsonism in newly medicated schizophrenia patients. J Neuropsychiatry Clin Neurosci 1997;9:562-7.
27. Hriso E, Kuhn T, Masdeu JC et al. Extrapyramidal symptoms due to dopamine-blocking agents in patients with AIDS encephalopathy. Am J Psychiatry 1991;148:1558-61.
28. Lera G, Zirulnik J. Pilot study with clozapine in patients with HIV-associated psychosis and drug-induced parkinsonism. Mov Disord 1999;14: 128-31.
29. Chang A, Fox SH. Psychosis in Parkinson's disease: epidemiology, pathophysiology, and management. Drugs 2016;76:1093-18.
30. Samudra N, Patel N, Womack KB et al. Psychosis in Parkinson disease: a review of etiology, phenomenology, and management. Drugs Aging 2016; 33:855-63.
31. Taylor J, Anderson WS, Brandt J et al. Neuropsychiatric complications of Parkinson disease treatments: importance of multidisciplinary care. Am J Geriatr Psychiatry 2016;24:1171-80.
32. Mamo DC, Sweet RA, Keshavan MS. Managing antipsychotic-induced parkinsonism. Drug Saf 1999;20:269-75.
33. Cortese L, Caligiuri MP, Williams R et al. Reduction in neuroleptic-induced movement disorders after a switch to quetiapine in patients with schizophrenia. J Clin Psychopharmacol 2008;28:69-73.
34. Magnus RV. A comparison of biperiden hydrochloride (Akineton) and benzhexol (Artane) in the treatment of drug-induced Parkinsonism. J Int Med Res 1980;8:343-6.
35. Dayalu P, Chou KL. Antipsychotic-induced extrapyramidal symptoms and their management. Expert Opin Pharmacother 2008;9:1451-62.
36. Ananth J, Sangani H, Noonan JP. Amantadine in drug-induced extrapyramidal signs: a comparative study. Int J Clin Pharmacol Biopharm 1975; 11:323-6.
37. Chouinard G, Annable L, Ross-Chouinard A et al. Ethopropazine and benztropine in neuroleptic-induced parkinsonism. J Clin Psychiatry 1979; 40:147-52.
38. Kelly JT, Zimmermann RL, Abuzzahab FS et al. A double-blind study of amantadine hydrochloride versus benztropine mesylate in drug-induced parkinsonism. Pharmacology 1974;12:65-73.

39. Mindham RH, Gaind R, Anstee BH et al. Comparison of amantadine, orphenadrine, and placebo in the control of phenothiazine-induced Parkinsonism. Psychol Med 1972;2:406-13.
40. Silver H, Geraisy N, Schwartz M. No difference in the effect of biperiden and amantadine on parkinsonian- and tardive dyskinesia-type involuntary movements: a double-blind crossover, placebo-controlled study in medicated chronic schizophrenic patients. J Clin Psychiatry 1995;56:167-70.
41. Van Putten T. Vulnerability to extrapyramidal side effects. Clin Neuropharmacol 1983;6(Suppl. 1):S27-34.
42. Miller CH, Hummer M, Oberbauer H et al. Risk factors for the development of neuroleptic induced akathisia. Eur Neuropsychopharmacol 1997; 7:51-5.
43. Sachdev P, Kruk J. Clinical characteristics and predisposing factors in acute drug-induced akathisia. Arch Gen Psychiatry 1994;51:963-74.
44. Juncal-Ruiz M, Ramirez-Bonilla M, Gomez-Arnau J et al. Incidence and risk factors of acute akathisia in 493 individuals with first episode non-affective psychosis: a 6-week randomised study of antipsychotic treatment. Psychopharmacology 2017;234:2563-70.
45. Poyurovsky M. Acute antipsychotic-induced akathisia revisited. Br J Psychiatry 2010;196:89-91.
46. Poyurovsky M, Pashinian A, Weizman R et al. Low-dose mirtazapine: a new option in the treatment of antipsychotic-induced akathisia. A randomized, double-blind, placebo- and propranolol-controlled trial. Biol Psychiatry 2006;59:1071-7.
47. Adler L, Angrist B, Peselow E et al. A controlled assessment of propranolol in the treatment of neuroleptic-induced akathisia. Br J Psychiatry 1986;149:42-5.
48. Kramer MS, Gorkin R, DiJohnson C. Treatment of neuroleptic-induced akathisia with propranolol: a controlled replication study. Hillside J Clin Psychiatry 1989;11:107-19.
49. Kramer MS, Gorkin RA, DiJohnson C et al. Propranolol in the treatment of neuroleptic-induced akathisia (NIA) in schizophrenics: a double-blind, placebo-controlled study. Biol Psychiatry 1988;24:823-7.
50. Lima AR, Bacalcthuk J, Barnes TR et al. Central action beta-blockers versus placebo for neuroleptic-induced acute akathisia. Cochrane Database Syst Rev 2004;4:CD001946.
51. Rathbone J, Soares-Weiser K. Anticholinergics for neuroleptic-induced acute akathisia. Cochrane Database Syst Rev 2006;4:CD003727.
52. Fleischhacker WW, Roth SD, Kane JM. The pharmacologic treatment of neuroleptic-induced akathisia. J Clin Psychopharmacol 1990;10:12-21.
53. Hieber R, Dellenbaugh T, Nelson LA. Role of mirtazapine in the treatment of antipsychotic-induced akathisia. Ann Pharmacother 2008;42:841-6.
54. Praharaj SK, Kongasseri S, Behere RV et al. Mirtazapine for antipsychotic-induced acute akathisia: a systematic review and meta-analysis of randomized placebo-controlled trials. Ther Adv Psychopharmacol 2015;5:307-13.
55. Miller CH, Fleischhacker WW, Ehrmann H et al. Treatment of neuroleptic induced akathisia with the 5-HT2 antagonist ritanserin. Psychopharmacol Bull 1990;26:373-6.
56. Miller CH, Hummer M, Pycha R et al. The effect of ritanserin on treatment-resistant neuroleptic induced akathisia: case reports. Prog Neuropsychopharmacol Biol Psychiatry 1992;16:247-51.
57. Poyurovsky M, Fuchs C, Weizman A. Low-dose mianserin in treatment of acute neuroleptic-induced akathisia. J Clin Psychopharmacol 1998;18: 253-4.
58. Poyurovsky M, Shardorodsky M, Fuchs C et al. Treatment of neuroleptic-induced akathisia with the 5-HT2 antagonist mianserin. Double-blind, placebo-controlled study. Br J Psychiatry 1999;174:238-42.
59. Avital A, Gross-Isseroff R, Stryjer R et al. Zolmitriptan compared to propranolol in the treatment of acute neuroleptic-induced akathisia: a comparative double-blind study. Eur Neuropsychopharmacol 2009;19: 476-82.
60. Fischel T, Hermesh H, Aizenberg D et al. Cyproheptadine versus propranolol for the treatment of acute neuroleptic-induced akathisia: a comparative double-blind study. J Clin Psychopharmacol 2001;21:612-5.
61. Hirose S, Ashby CR. Immediate effect of intravenous diazepam in neuroleptic-induced acute akathisia: an open-label study. J Clin Psychiatry 2002; 63:524-7.
62. Bartels M, Heide K, Mann K et al. Treatment of akathisia with lorazepam. An open clinical trial. Pharmacopsychiatry 1987;20:51-3.
63. Kutcher S, Williamson P, MacKenzie S et al. Successful clonazepam treatment of neuroleptic-induced akathisia in older adolescents and young adults: a double-blind, placebo-controlled study. J Clin Psychopharmacol 1989;9:403-6.
64. Pujalte D, Bottai T, Hue B et al. A double-blind comparison of clonazepam and placebo in the treatment of neuroleptic-induced akathisia. Clin Neuropharmacol 1994;17:236-42.
65. Lima AR, Soares-Weiser K, Bacaltchuk J et al. Benzodiazepines for neuroleptic-induced acute akathisia. Cochrane Database Syst Rev 2002; 1:CD001950.
66. Tiihonen J, Mittendorfer-Rutz E, Torniainen M et al. Mortality and cumulative exposure to antipsychotics, antidepressants, and benzodiazepines in patients with schizophrenia: an observational follow-up study. Am J Psychiatry 2016;173:600-6.
67. Lerner V, Bergman J, Statsenko N et al. Vitamin B6 treatment in acute neuroleptic-induced akathisia: a randomized, double-blind, placebo-controlled study. J Clin Psychiatry 2004;65:1550-4.
68. Bernstein AL. Vitamin B6 in clinical neurology. Ann N Y Acad Sci 1990; 585:250-60.
69. Erdur B, Tura P, Aydin B et al. A trial of midazolam vs diphenhydramine in prophylaxis of metoclopramide-induced akathisia. Am J Emerg Med 2012;30:84-91.
70. Friedman BW, Bender B, Davitt M et al. A randomized trial of diphenhydramine as prophylaxis against metoclopramide-induced akathisia in nauseated emergency department patients. Ann Emerg Med 2009;53:379-85.
71. Vinson DR. Diphenhydramine in the treatment of akathisia induced by prochlorperazine. J Emerg Med 2004;26:265-70.
72. Cornett EM, Novitch M, Kaye AD et al. Medication-induced tardive dyskinesia: a review and update. Ochsner J 2017;17:162-74.
73. Bhidayasiri R, Boonyawairoj S. Spectrum of tardive syndromes: clinical recognition and management. Postgrad Med J 2011;87:132-41.
74. Yassa R, Jeste DV. Gender differences in tardive dyskinesia: a critical review of the literature. Schizophr Bull 1992;18:701-15.
75. Tenback DE, van Harten PN, Slooff CJ et al. Evidence that early extrapyramidal symptoms predict later tardive dyskinesia: a prospective analysis of 10,000 patients in the European Schizophrenia Outpatient Health Outcomes (SOHO) study. Am J Psychiatry 2006;163:1438-40.
76. Tenback DE, van Harten PN, van Os J. Non-therapeutic risk factors for onset of tardive dyskinesia in schizophrenia: a meta-analysis. Mov Disord 2009;24:2309-15.
77. Miller DD, McEvoy JP, Davis SM et al. Clinical correlates of tardive dyskinesia in schizophrenia: baseline data from the CATIE schizophrenia trial. Schizophr Res 2005;80:33-43.
78. Zhang XY, Chen DC, Qi LY et al. Gender differences in the prevalence, risk and clinical correlates of tardive dyskinesia in Chinese schizophrenia. Psychopharmacology 2009;205:647-54.
79. Bakker PR, van Harten PN, van Os J. Antipsychotic-induced tardive dyskinesia and polymorphic variations in COMT, DRD2, CYP1A2 and MnSOD genes: a meta-analysis of pharmacogenetic interactions. Mol Psychiatry 2008;13:544-56.
80. Zai CC, Tiwari AK, Muller DJ et al. The catechol-O-methyl-transferase gene in tardive dyskinesia. World J Biol Psychiatry 2010;11:803-12.
81. Miura I, Zhang JP, Nitta M et al. BDNF Val66Met polymorphism and antipsychotic-induced tardive dyskinesia occurrence and severity: a meta-analysis. Schizophr Res 2014;152:365-72.
82. Inada T, Koga M, Ishiguro H et al. Pathway-based association analysis of genome-wide screening data suggest that genes associated with the gamma-aminobutyric acid receptor signaling pathway are involved in neuroleptic-induced, treatment-resistant tardive dyskinesia. Pharmacogenet Gen 2008;18:317-23.
83. Lanning RK, Zai CC, Muller DJ. Pharmacogenetics of tardive dyskinesia: an updated review of the literature. Pharmacogenomics 2016;17:1339-51.
84. Son WY, Lee HJ, Yoon HK et al. GABA transporter SLC6A11 gene polymorphism associated with tardive dyskinesia. Nordic J Psychiatry 2014; 68:123-8.
85. Carbon M, Hsieh CH, Kane JM et al. Tardive dyskinesia prevalence in the period of second-generation antipsychotic use: a meta-analysis. J Clin Psychiatry 2017;78:e264-78.
86. Correll CU, Schenk EM. Tardive dyskinesia and new antipsychotics. Curr Opin Psychiatry 2008;21:151-6.
87. Correll CU, Leucht S, Kane JM. Lower risk for tardive dyskinesia associated with second-generation antipsychotics: a systematic review of 1-year studies. Am J Psychiatry 2004;161:414-25.
88. O'Brien A. Comparing the risk of tardive dyskinesia in older adults with first-generation and second-generation antipsychotics: a systematic review and meta-analysis. Int J Geriatr Psychiatry 2016;31:683-93.

89. Bai YM, Yu SC, Chen JY et al. Risperidone for pre-existing severe tardive dyskinesia: a 48-week prospective follow-up study. Int Clin Psychopharmacol 2005;20:79-85.
90. Chan HY, Chiang SC, Chang CJ et al. A randomized controlled trial of risperidone and olanzapine for schizophrenic patients with neuroleptic-induced tardive dyskinesia. J Clin Psychiatry 2010;71:1226-33.
91. Chouinard G. Effects of risperidone in tardive dyskinesia: an analysis of the Canadian multicenter risperidone study. J Clin Psychopharmacol 1995;15:36S-44S.
92. Bergman H, Walker DM, Nikolakopoulou A et al. Systematic review of interventions for treating or preventing antipsychotic-induced tardive dyskinesia. Health Technol Assess 2017;21:1-218.
93. Hauser RA, Factor SA, Marder SR et al. KINECT 3: a phase 3 randomized, double-blind, placebo-controlled trial of valbenazine for tardive dyskinesia. Am J Psychiatry 2017;174:476-84.
94. Leung JG, Breden EL. Tetrabenazine for the treatment of tardive dyskinesia. Ann Pharmacother 2011;45:525-31.
95. Ondo WG, Hanna PA, Jankovic J. Tetrabenazine treatment for tardive dyskinesia: assessment by randomized videotape protocol. Am J Psychiatry 1999;156:1279-81.
96. Citrome L. Valbenazine for tardive dyskinesia: a systematic review of the efficacy and safety profile for this newly approved novel medication – what is the number needed to treat, number needed to harm and likelihood to be helped or harmed? Int J Clin Pract (in press).
97. Fernandez HH, Factor SA, Hauser RA et al. Randomized controlled trial of deutetrabenazine for tardive dyskinesia: the ARM-TD study. Neurology 2017;88:2003-10.
98. Alabed S, Latifeh Y, Mohammad HA et al. Gamma-aminobutyric acid agonists for neuroleptic-induced tardive dyskinesia. Cochrane Database Syst Rev 2011;4:CD000203.
99. Thaker GK, Nguyen JA, Strauss ME et al. Clonazepam treatment of tardive dyskinesia: a practical GABAmimetic strategy. Am J Psychiatry 1990;147: 445-51.
100. Zhang WF, Tan YL, Zhang XY et al. Extract of Ginkgo biloba treatment for tardive dyskinesia in schizophrenia: a randomized, double-blind, placebo-controlled trial. J Clin Psychiatry 2011;72:615-21.
101. Zhang XY, Zhang WF, Zhou DF et al. Brain-derived neurotrophic factor levels and its Val66Met gene polymorphism predict tardive dyskinesia treatment response to Ginkgo biloba. Biol Psychiatry 2012;72:700-6.
102. Adelufosi AO, Abayomi O, Ojo TM. Pyridoxal 5 phosphate for neuroleptic-induced tardive dyskinesia. Cochrane Database Syst Rev 2015;4: CD010501.
103. Soares-Weiser K, Maayan N, McGrath J. Vitamin E for neuroleptic-induced tardive dyskinesia. Cochrane Database Syst Rev 2011;2:CD000209.
104. Bhidayasiri R, Fahn S, Weiner WJ et al. Evidence-based guideline: treatment of tardive syndromes: report of the Guideline Development Subcommittee of the American Academy of Neurology. Neurology 2013;81: 463-9.
105. El-Sayeh HG, Lyra da Silva JP, Rathbone J et al. Non-neuroleptic catecholaminergic drugs for neuroleptic-induced tardive dyskinesia. Cochrane Database Syst Rev 2006;1:CD000458.
106. Essali A, Deirawan H, Soares-Weiser K et al. Calcium channel blockers for neuroleptic-induced tardive dyskinesia. Cochrane Database Syst Rev 2011; 11:CD000206.
107. Pouclet-Courtemanche H, Rouaud T, Thobois S et al. Long-term efficacy and tolerability of bilateral pallidal stimulation to treat tardive dyskinesia. Neurology 2016;86:651-9.
108. Saleh C, Gonzalez V, Coubes P. Role of deep brain stimulation in the treatment of secondary dystonia-dyskinesia syndromes. Handb Clin Neurol 2013;116:189-208.
109. Sobstyl M, Zabek M. Deep brain stimulation for intractable tardive dystonia: literature overview. Neurol Neurochir Pol 2016;50:114-22.
110. Thobois S, Poisson A, Damier P. Surgery for tardive dyskinesia. Int Rev Neurobiol 2011;98:289-96.
111. Maher S, Cunningham A, O'Callaghan N et al. Clozapine-induced hypersalivation: an estimate of prevalence, severity and impact on quality of life. Ther Adv Psychopharmacol 2016;6:178-84.
112. Corrigan FM, MacDonald S, Reynolds GP. Clozapine-induced hypersalivation and the alpha 2 adrenoceptor. Br J Psychiatry 1995;167:412.
113. Ukai Y, Taniguchi N, Ishima T et al. Muscarinic supersensitivity and subsensitivity induced by chronic treatment with atropine and diisopropylfluorophosphate in rat submaxillary glands. Arch Int Pharmacodyn Ther 1989;297:148-57.
114. Saenger RC, Finch TH, Francois D. Aspiration pneumonia due to clozapine-induced sialorrhea. Clin Schizophr Relat Psychoses 2016;9: 170-2.
115. Trigoboff E, Grace J, Szymanski H et al. Sialorrhea and aspiration pneumonia: a case study. Innov Clin Neurosci 2013;10:20-7.
116. Acosta-Armas AJ. Two cases of parotid gland swelling in patients taking clozapine. Hosp Med 2001;62:704-5.
117. Patkar AA, Alexander RC. Parotid gland swelling with clozapine. J Clin Psychiatry 1996;57:488.
118. Bird AM, Smith TL, Walton AE. Current treatment strategies for clozapine-induced sialorrhea. Ann Pharmacother 2011;45:667-75.
119. Essali A, Rihawi A, Altujjar M et al. Anticholinergic medication for non-clozapine neuroleptic-induced hypersalivation in people with schizophrenia. Cochrane Database Syst Rev 2013;12:CD009546.
120. Antonello C, Tessier P. Clozapine and sialorrhea: a new intervention for this bothersome and potentially dangerous side effect. J Psychiatry Neurosci 1999;24:250.
121. Sharma A, Ramaswamy S, Dahl E et al. Intraoral application of atropine sulfate ophthalmic solution for clozapine-induced sialorrhea. Ann Pharmacother 2004;38:1538.
122. Tessier P, Antonello C. Clozapine and sialorrhea: update. J Psychiatry Neurosci 2001;26:253.
123. Calderon J, Rubin E, Sobota WL. Potential use of ipratropium bromide for the treatment of clozapine-induced hypersalivation: a preliminary report. Int Clin Psychopharmacol 2000;15:49-52.
124. Freudenreich O, Beebe M, Goff DC. Clozapine-induced sialorrhea treated with sublingual ipratropium spray: a case series. J Clin Psychopharmacol 2004;24:98-100.
125. Sockalingam S, Shammi C, Remington G. Treatment of clozapine-induced hypersalivation with ipratropium bromide: a randomized, double-blind, placebo-controlled crossover study. J Clin Psychiatry 2009;70:1114-9.
126. Reinstein M, Sirotovskaya L, Chasanov M et al. Comparative efficacy and tolerability of benztropine and terazosin in the treatment of hypersalivation secondary to clozapine. Clin Drug Invest 1999;17:97-102.
127. Spivak B, Adlersberg S, Rosen L et al. Trihexyphenidyl treatment of clozapine-induced hypersalivation. Int Clin Psychopharmacol 1997;12: 213-5.
128. Liang CS, Ho PS, Shen LJ et al. Comparison of the efficacy and impact on cognition of glycopyrrolate and biperiden for clozapine-induced sialorrhea in schizophrenic patients: a randomized, double-blind, crossover study. Schizophr Res 2010;119:138-44.
129. Bai YM, Lin CC, Chen JY et al. Therapeutic effect of pirenzepine for clozapine-induced hypersalivation: a randomized, double-blind, placebo-controlled, cross-over study. J Clin Psychopharmacol 2001;21: 608-11.
130. Schneider B, Weigmann H, Hiemke C et al. Reduction of clozapine-induced hypersalivation by pirenzepine is safe. Pharmacopsychiatry 2004;37:43-5.
131. Copp PJ, Lament R, Tennent TG. Amitriptyline in clozapine-induced sialorrhoea. Br J Psychiatry 1991;159:166.
132. Grabowski J. Clonidine treatment of clozapine-induced hypersalivation. J Clin Psychopharmacol 1992;12:69-70.
133. Praharaj SK, Verma P, Roy D et al. Is clonidine useful for treatment of clozapine-induced sialorrhea? J Psychopharmacol 2005;19:426-8.
134. Webber MA, Szwast SJ, Steadman TM et al. Guanfacine treatment of clozapine-induced sialorrhea. J Clin Psychopharmacol 2004;24:675-6.
135. Kreinin A, Epshtein S, Sheinkman A et al. Sulpiride addition for the treatment of clozapine-induced hypersalivation: preliminary study. Isr J Psychiatry Relat Sci 2005;42:61-3.
136. Kreinin A, Miodownik C, Sokolik S et al. Amisulpride versus moclobemide in treatment of clozapine-induced hypersalivation. World J Biol Psychiatry 2011;12:620-6.
137. Kreinin A, Novitski D, Weizman A. Amisulpride treatment of clozapine-induced hypersalivation in schizophrenia patients: a randomized, double-blind, placebo-controlled cross-over study. Int Clin Psychopharmacol 2006;21:99-103.
138. Kreinin A, Miodownik C, Libov I et al. Moclobemide treatment of clozapine-induced hypersalivation: pilot open study. Clin Neuropharmacol 2009;32:151-3.
139. Kahl KG, Hagenah J, Zapf S et al. Botulinum toxin as an effective treatment of clozapine-induced hypersalivation. Psychopharmacology 2004; 173:229-30.

140. Steinlechner S, Klein C, Moser A et al. Botulinum toxin B as an effective and safe treatment for neuroleptic-induced sialorrhea. Psychopharmacology 2010;207:593-7.
141. Leucht S, Cipriani A, Spineli L et al. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. Lancet 2013;382:951-62.
142. Masopust J, Maly R, Valis M. Risk of venous thromboembolism during treatment with antipsychotic agents. Psychiatry Clin Neurosci 2012;66: 541-52.
143. Miller DD. Atypical antipsychotics: sleep, sedation, and efficacy. Prim Care Companion J Clin Psychiatry 2004;6(Suppl. 2):3-7.
144. Andrade C, Kisely S, Monteiro I et al. Antipsychotic augmentation with modafinil or armodafinil for negative symptoms of schizophrenia: systematic review and meta-analysis of randomized controlled trials. J Psychiatr Res 2015;60:14-21.
145. Burke M, Sebastian CS. Treatment of clozapine sedation. Am J Psychiatry 1993;150:1900-1.
146. Carpenter MD, Winsberg BG, Camus LA. Methylphenidate augmentation therapy in schizophrenia. J Clin Psychopharmacol 1992;12:273-5.
147. Szeszko PR, Bilder RM, Dunlop JA et al. Longitudinal assessment of methylphenidate effects on oral word production and symptoms in first-episode schizophrenia at acute and stabilized phases. Biol Psychiatry 1999;45:680-6.
148. Makela EH, Miller K, Cutlip WD 2nd. Three case reports of modafinil use in treating sedation induced by antipsychotic medications. J Clin Psychiatry 2003;64:485-6.
149. Saavedra-Velez C, Yusim A, Anbarasan D et al. Modafinil as an adjunctive treatment of sedation, negative symptoms, and cognition in schizophrenia: a critical review. J Clin Psychiatry 2009;70:104-12.
150. Freudenreich O, Henderson DC, Macklin EA et al. Modafinil for clozapine-treated schizophrenia patients: a double-blind, placebo-controlled pilot trial. J Clin Psychiatry 2009;70:1674-80.
151. Miller SC. Methylphenidate for clozapine sedation. Am J Psychiatry 1996; 153:1231-2.
152. Saraf G, Viswanath B, Narayanaswamy JC et al. Modafinil for the treatment of antipsychotic-induced excessive daytime sedation: does it exacerbate tics? J Neuropsychiatry Clin Neurosci 2013;25:E35-6.
153. Chen CY, Lane HY, Lin CH. Effects of antipsychotics on bone mineral density in patients with schizophrenia: gender differences. Clin Psychopharmacol Neurosci 2016;14:238-49.
154. Kollin E, Gergely I, Szucs J et al. Decreased bone density in hyperprolactinemic women. N Engl J Med 1981;304:1362-3.
155. Gonzalez-Blanco L, Greenhalgh AM, Garcia-Rizo C et al. Prolactin concentrations in antipsychotic-naive patients with schizophrenia and related disorders: a meta-analysis. Schizophr Res 2016;174:156-60.
156. La Torre A, Conca A, Duffy D et al. Sexual dysfunction related to psychotropic drugs: a critical review part II: antipsychotics. Pharmacopsychiatry 2013;46:201-8.
157. de Boer MK, Castelein S, Bous J et al. The Antipsychotics and Sexual Functioning Questionnaire (ASFQ): preliminary evidence for reliability and validity. Schizophr Res 2013;150:410-5.
158. Compton MT, Miller AH. Antipsychotic-induced hyperprolactinemia and sexual dysfunction. Psychopharmacol Bull 2002;36:143-64.
159. Serretti A, Chiesa A. A meta-analysis of sexual dysfunction in psychiatric patients taking antipsychotics. Int Clin Psychopharmacol 2011;26: 130-40.
160. Berner MM, Hagen M, Kriston L. Management of sexual dysfunction due to antipsychotic drug therapy. Cochrane Database Syst Rev 2007;1: CD003546.
161. Renn JH, Yang NP, Chueh CM et al. Bone mass in schizophrenia and normal populations across different decades of life. BMC Musculoskelet Disord 2009;10:1.
162. Stubbs B, De Hert M, Sepehry AA et al. A meta-analysis of prevalence estimates and moderators of low bone mass in people with schizophrenia. Acta Psychiatr Scand 2014;130:470-86.
163. Howard L, Kirkwood G, Leese M. Risk of hip fracture in patients with a history of schizophrenia. Br J Psychiatry 2007;190:129-34.
164. Lin CH, Lin CY, Huang TL et al. Sex-specific factors for bone density in patients with schizophrenia. Int Clin Psychopharmacol 2015;30:96-102.
165. Catala-Lopez F, Suarez-Pinilla M, Suarez-Pinilla P et al. Inverse and direct cancer comorbidity in people with central nervous system disorders: a meta-analysis of cancer incidence in 577,013 participants of 50 observational studies. Psychother Psychosom 2014;83:89-105.
166. De Hert M, Peuskens J, Sabbe T et al. Relationship between prolactin, breast cancer risk, and antipsychotics in patients with schizophrenia: a critical review. Acta Psychiatr Scand 2016;133:5-22.
167. Holt RI, Peveler RC. Antipsychotics and hyperprolactinaemia: mechanisms, consequences and management. Clin Endocrinol 2011;74:141-7.
168. Lehman AF, Lieberman JA, Dixon LB et al. Practice guideline for the treatment of patients with schizophrenia, second edition. Am J Psychiatry 2004;161:1-56.
169. Peveler RC, Branford D, Citrome L et al. Antipsychotics and hyperprolactinaemia: clinical recommendations. J Psychopharmacol 2008;22:98-103.
170. Hernandez CJ, Beaupre GS, Carter DR. A theoretical analysis of the relative influences of peak BMD, age-related bone loss and menopause on the development of osteoporosis. Osteoporos Int 2003;14:843-7.
171. Tworoger SS, Eliassen AH, Sluss P et al. A prospective study of plasma prolactin concentrations and risk of premenopausal and postmenopausal breast cancer. J Clin Oncol 2007;25:1482-8.
172. Kane JM, Correll CU, Goff DC et al. A multicenter, randomized, double-blind, placebo-controlled, 16-week study of adjunctive aripiprazole for schizophrenia or schizoaffective disorder inadequately treated with quetiapine or risperidone monotherapy. J Clin Psychiatry 2009;70:1348-57.
173. Chang SC, Chen CH, Lu ML. Cabergoline-induced psychotic exacerbation in schizophrenic patients. Gen Hosp Psychiatry 2008;30:378-80.
174. Costa AM, Lima MS, Mari Jde J. A systematic review on clinical management of antipsychotic-induced sexual dysfunction in schizophrenia. Sao Paulo Med J 2006;124:291-7.
175. Gugger JJ. Antipsychotic pharmacotherapy and orthostatic hypotension: identification and management. CNS Drugs 2011;25:659-71.
176. Drici MD, Priori S. Cardiovascular risks of atypical antipsychotic drug treatment. Pharmacoepidemiol Drug Safety 2007;16:882-90.
177. Stroup TS, Lieberman JA, McEvoy JP et al. Results of phase 3 of the CATIE schizophrenia trial. Schizophr Res 2009;107:1-12.
178. Citrome L. Iloperidone for schizophrenia: a review of the efficacy and safety profile for this newly commercialised second-generation antipsychotic. Int J Clin Pract 2009;63:1237-48.
179. Leung JY, Barr AM, Procyshyn RM et al. Cardiovascular side-effects of antipsychotic drugs: the role of the autonomic nervous system. Pharmacol Ther 2012;135:113-22.
180. Freeman R. Current pharmacologic treatment for orthostatic hypotension. Clin Autonom Res 2008;18(Suppl. 1):14-8.
181. Figueroa JJ, Basford JR, Low PA. Preventing and treating orthostatic hypotension: as easy as A, B, C. Cleve Clin J Med 2010;77:298-306.
182. Testani M Jr. Clozapine-induced orthostatic hypotension treated with fludrocortisone. J Clin Psychiatry 1994;55:497-8.
183. Kurihara J, Takata Y, Suzuki S et al. Effect of midodrine on chlorpromazine-induced orthostatic hypotension in rabbits: comparison with amezinium, etilefrine and droxidopa. Biol Pharmaceut Bull 2000;23:1445-9.
184. Castrioto A, Tambasco N, Rossi A et al. Acute dystonia induced by the combination of midodrine and perphenazine. J Neurol 2008;255:767-8.
185. Takahashi H. Acute dystonia induced by adding midodrine, a selective alpha 1 agonist, to risperidone in a patient with catatonic schizophrenia. J Neuropsychiatry Clin Neurosci 2000;12:285-6.
186. Hou PY, Hung GC, Jhong JR et al. Risk factors for sudden cardiac death among patients with schizophrenia. Schizophr Res 2015;168:395-401.
187. Ray WA, Chung CP, Murray KT et al. Atypical antipsychotic drugs and the risk of sudden cardiac death. N Engl J Med 2009;360:225-35.
188. Straus SM, Bleumink GS, Dieleman JP et al. Antipsychotics and the risk of sudden cardiac death. Arch Int Med 2004;164:1293-7.
189. Wu CS, Tsai YT, Tsai HJ. Antipsychotic drugs and the risk of ventricular arrhythmia and/or sudden cardiac death: a nation-wide case-crossover study. J Am Heart Assoc 2015;4:e001568.
190. Mehtonen OP, Aranko K, Malkonen L et al. A survey of sudden death associated with the use of antipsychotic or antidepressant drugs: 49 cases in Finland. Acta Psychiatr Scand 1991;84:58-64.
191. Reilly JG, Ayis SA, Ferrier IN et al. Thioridazine and sudden unexplained death in psychiatric in-patients. Br J Psychiatry 2002;180:515-22.
192. Ray WA, Meredith S, Thapa PB et al. Antipsychotics and the risk of sudden cardiac death. Arch Gen Psychiatry 2001;58:1161-7.
193. Haddad PM, Anderson IM. Antipsychotic-related QTc prolongation, torsade de pointes and sudden death. Drugs 2002;62:1649-71.
194. Heist EK, Ruskin JN. Drug-induced proarrhythmia and use of QTc-prolonging agents: clues for clinicians. Heart Rhythm 2005;2:S1-8.
195. Schneeweiss S, Avorn J. Antipsychotic agents and sudden cardiac death – how should we manage the risk? N Engl J Med 2009;360:294-6.

196. Shah AA, Aftab A, Coverdale J. QTc prolongation with antipsychotics: is routine ECG monitoring recommended? J Psychiatr Pract 2014;20:196-206.
197. Lieberman JA, Merrill D, Parameswaran S. APA guidance on the use of antipsychotic drugs and cardiac sudden death. https://www.omh.ny.gov.
198. Kavanaugh G, Brown G, Khan I et al. Guidance on the use of antipsychotics, Version 3.2. London: National Health Service, 2015.
199. Pileggi DJ, Cook AM. Neuroleptic malignant syndrome. Ann Pharmacother 2016;50:973-81.
200. Trollor JN, Chen X, Chitty K et al. Comparison of neuroleptic malignant syndrome induced by first- and second-generation antipsychotics. Br J Psychiatry 2012;201:52-6.
201. Trollor JN, Chen X, Sachdev PS. Neuroleptic malignant syndrome associated with atypical antipsychotic drugs. CNS Drugs 2009;23:477-92.
202. Nielsen RE, Wallenstein Jensen SO, Nielsen J. Neuroleptic malignant syndrome – an 11-year longitudinal case-control study. Can J Psychiatry 2012;57:512-8.
203. Strawn JR, Keck PE Jr, Caroff SN. Neuroleptic malignant syndrome. Am J Psychiatry 2007;164:870-6.
204. Su YP, Chang CK, Hayes RD et al. Retrospective chart review on exposure to psychotropic medications associated with neuroleptic malignant syndrome. Acta Psychiatr Scand 2014;130:52-60.
205. Tse L, Barr AM, Scarapicchia V et al. Neuroleptic malignant syndrome: a review from a clinically oriented perspective. Curr Neuropharmacol 2015; 13:395-406.
206. Lee JW. Serum iron in catatonia and neuroleptic malignant syndrome. Biol Psychiatry 1998;44:499-507.
207. Patil BS, Subramanyam AA, Singh SL et al. Low serum iron as a possible risk factor for neuroleptic malignant syndrome. Int J Appl Basic Med Res 2014;4:117-8.
208. Rosebush PI, Mazurek MF. Serum iron and neuroleptic malignant syndrome. Lancet 1991;338:149-51.
209. Patel AL, Shaikh WA, Khobragade AK et al. Electroconvulsive therapy in drug resistant neuroleptic malignant syndrome. J Assoc Physicians India 2008;56:49-50.
210. Caroff SN, Mann SC. Neuroleptic malignant syndrome. Med Clin North Am 1993;77:185-202.
211. Wells AJ, Sommi RW, Crismon ML. Neuroleptic rechallenge after neuroleptic malignant syndrome: case report and literature review. Drug Intell Clin Pharm 1988;22:475-80.
212. Rosebush PI, Stewart TD, Gelenberg AJ. Twenty neuroleptic rechallenges after neuroleptic malignant syndrome in 15 patients. J Clin Psychiatry 1989;50:295-8.
213. Trollor JN, Sachdev PS. Electroconvulsive treatment of neuroleptic malignant syndrome: a review and report of cases. Aust N Z J Psychiatry 1999; 33:650-9.
214. Idänpään-Heikkilä J, Alhava E, Olkinuora M et al. Clozapine and agranulocytosis. Lancet 1975;306:611.
215. Alvir JM, Lieberman JA, Safferman AZ et al. Clozapine-induced agranulocytosis. Incidence and risk factors in the United States. N Engl J Med 1993; 329:162-7.
216. Canoso RT, Lewis ME, Yunis EJ. Association of HLA-Bw44 with chlorpromazine-induced autoantibodies. Clin Immunol Immunopathol 1982;25: 278-82.
217. Fiore JM, Francis MN. Agranulocytosis due to mepazine (phenothiazine): report of three cases. N Engl J Med 1959;260:375-8.
218. Goldman D. Agranulocytosis associated with administration of chlorpromazine: report of three cases with description of treatment and recovery. AMA Arch Int Med 1955;96:496.
219. Stergiou V, Bozikas VP, Garyfallos G et al. Olanzapine-induced leucopenia and neutropenia. Prog Neuropsychopharmacol Biol Psychiatry 2005;29:992-4.
220. Sluys M, Güzelcan Y, Casteelen G et al. Risperidone-induced leucopenia and neutropenia: a case report. Eur Psychiatry 2004;19:117.
221. Munro J, O'Sullivan D, Andrews C et al. Active monitoring of 12,760 clozapine recipients in the UK and Ireland. Beyond pharmacovigilance. Br J Psychiatry 1999;175:576-80.
222. Sporn A, Gogtay N, Ortiz-Aguayo R et al. Clozapine-induced neutropenia in children: management with lithium carbonate. J Child Adolesc Psychopharmacol 2003;13:401-4.
223. Papetti F, Darcourt G, Giordana J-Y et al. Treatment of clozapine-induced granulocytopenia with lithium (two observations). Encéphale 2004;30: 578-82.
224. Yadav D, Burton S, Sehgal C. Clozapine-induced neutropenia reversed by lithium. Prog Neurol Psychiatry 2016;20:13-5.
225. Nielsen H. Recombinant human granulocyte colony-stimulating factor (rhG-CSF; filgrastim) treatment of clozapine-induced agranulocytosis. J Int Med 1993;234:529-31.
226. Lamberti JS, Bellnier TJ, Schwarzkopf SB et al. Filgrastim treatment of three patients with clozapine-induced agranulocytosis. J Clin Psychiatry 1995;56:256-9.
227. Simon L, Cazard F. Clozapine rechallenge after neutropenia in resistant schizophrenia: a review. Encéphale 2016;42:346-53.
228. Gaboriau L, Victorri-Vigneau C, Gérardin M et al. Aripiprazole: a new risk factor for pathological gambling? A report of 8 case reports. Addict Behav 2014;39:562-5.
229. Smith N, Kitchenham N, Bowden-Jones H. Pathological gambling and the treatment of psychosis with aripiprazole: case reports. Br J Psychiatry 2011;199:158-9.
230. Nirenberg MJ, Waters C. Compulsive eating and weight gain related to dopamine agonist use. Mov Disord 2006;21:524-9.
231. Claassen DO, van den Wildenberg WP, Ridderinkhof KR et al. The risky business of dopamine agonists in Parkinson disease and impulse control disorders. Behav Neurosci 2011;125:492-500.
232. Etminan M, Sodhi M, Samii A et al. Risk of gambling disorder and impulse control disorder with aripiprazole, pramipexole, and ropinirole: a pharmacoepidemiologic study. J Clin Psychopharmacol 2017;37:102-4.
233. Cheon E, Koo BH, Seo SS et al. Two cases of hypersexuality probably associated with aripiprazole. Psychiatry Invest 2013;10:200-2.
234. Cohen J, Magalon D, Boyer L et al. Aripiprazole-induced pathological gambling: a report of 3 cases. Curr Drug Saf 2011;6:51-3.
235. Haas SJ, Hill R, Krum H et al. Clozapine-associated myocarditis: a review of 116 cases of suspected myocarditis associated with the use of clozapine in Australia during 1993-2003. Drug Saf 2007;30:47-57.
236. Ronaldson K, Fitzgerald P, Taylor A et al. Clinical course and analysis of ten fatal cases of clozapine-induced myocarditis and comparison with 66 surviving cases. Schizophr Res 2011;128:161-5.
237. Chopra N, de Leon J. Clozapine-induced myocarditis may be associated with rapid titration: a case report verified with autopsy. Int J Psychiatry Med 2016;51:104-15.
238. Youssef DL, Narayanan P, Gill N. Incidence and risk factors for clozapine-induced myocarditis and cardiomyopathy at a regional mental health service in Australia. Australas Psychiatry 2016;24:176-80.
239. Freudenreich O. Clozapine-induced myocarditis: prescribe safely but do prescribe. Acta Psychiatr Scand 2015;132:240-1.
240. Ellis PM, Cameron CR. Clozapine-induced myocarditis and baseline echocardiograms. Aust N Z J Psychiatry 2012;46:1005-6.
241. Ittasakul P, Archer A, Kezman J et al. Rapid re-challenge with clozapine following pronounced myocarditis in a treatment-resistance schizophrenia patient. Clin Schizophr Relat Psychoses 2013;10:120-2.
242. Ganguli R. Behavioral therapy for weight loss in patients with schizophrenia. J Clin Psychiatry 2007;68(Suppl. 4):19-25.
243. Daumit GL, Dickerson FB, Wang NY et al. A behavioral weight-loss intervention in persons with serious mental illness. N Engl J Med 2013;368: 1594-602.
244. Bartels SJ, Pratt SI, Aschbrenner KA et al. Pragmatic replication trial of health promotion coaching for obesity in serious mental illness and maintenance of outcomes. Am J Psychiatry 2015;172:344-52.
245. Green CA, Yarborough BJ, Leo MC et al. The STRIDE weight loss and lifestyle intervention for individuals taking antipsychotic medications: a randomized trial. Am J Psychiatry 2015;172:71-81.
246. Wu RR, Zhao JP, Guo XF et al. Metformin addition attenuates olanzapine-induced weight gain in drug-naive first-episode schizophrenia patients: a double-blind, placebo-controlled study. Am J Psychiatry 2008; 165:352-8.
247. Wu RR, Zhao JP, Jin H et al. Lifestyle intervention and metformin for treatment of antipsychotic induced weight gain: a randomized controlled trial. JAMA 2008;299:185-93.
248. Jarskog LF, Hamer RM, Catellier DJ et al. Metformin for weight loss and metabolic control in overweight outpatients with schizophrenia and schizoaffective disorder. Am J Psychiatry 2013;170:1032-40.
249. Baptista T, Rangel N, Fernández V et al. Metformin as an adjunctive treatment to control body weight and metabolic dysfunction during olanzapine administration: a multicentric, double-blind, placebo-controlled trial. Schizophr Res 2007;93:99-108.

250. Tek C, Ratliff J, Reutenauer E et al. A randomized, double-blind, placebo-controlled pilot study of naltrexone to counteract antipsychotic-associated weight gain: proof of concept. J Clin Psychopharmacol 2014;34:608-12.
251. Lieberman JA 3rd. Managing anticholinergic side effects. Prim Care Companion J Clin Psychiatry 2004;6(Suppl. 2):20-3.
252. Nielsen J, Meyer JM. Risk factors for ileus in patients with schizophrenia. Schizophr Bull 2012;38:592-8.
253. Every-Palmer S, Ellis PM. Clozapine-induced gastrointestinal hypomotility: a 22-year bi-national pharmacovigilance study of serious or fatal 'slow gut' reactions, and comparison with international drug safety advice. CNS Drugs 2017;31:699-709.
254. Hayat U, Dugum M, Garg S. Chronic constipation: update on management. Cleve Clin J Med 2017;84:397-408.
255. Lieberman JA, Stroup TS, McEvoy JP et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. N Engl J Med 2005;353:1209-23.

DOI:10.1002/wps.20567

Healthy pregnancy and prevention of psychosis

Do healthier pregnancies reduce the risk of offspring psychosis? Variants of this question have appeared in recent papers, but with little discussion of how to answer it.

As a starting point, we note that current research on prenatal factors and psychotic disorders is relevant to this question but only addresses it partially and indirectly; that we are not aware of studies that do address it holistically; and that we do not yet know how such studies could be done. We begin by offering a perspective on research on prenatal factors and psychotic disorders. Then we discuss three points that would require consideration before explicitly directing research toward the question at hand.

About 25 years ago, numerous established pregnancy/birth cohorts had already reached adulthood, creating new opportunities for life course investigations[1]. In the US, for example, investigators launched studies of prenatal factors and psychotic disorders that made use of archived prenatal maternal sera in two large pregnancy cohorts. The development and linkage of national electronic registries in Scandinavia and elsewhere further transformed capacity to investigate prenatal factors and psychotic disorders[2]. Studies of prenatal exposures based in "natural experiments" also contributed by mitigating sources of confounding that preclude causal inference in traditional observational designs[3].

We are presently at the cusp of another leap forward, as national registries are being linked to archived biological data[4]; pregnancy cohorts of more than 100,000 births are entering the peak age of risk for psychotic disorders, with prenatal genetic and biological data, and ongoing follow-up[5]; and natural experiments are being conducted which include neuroimaging as well as diagnoses[6].

Relatively strong evidence suggests a role for prenatal infection and nutrition, but prenatal toxic exposures, prenatal stress, and interpregnancy intervals are also viable candidates, to name just a few. New methodologies from epidemiology are increasingly incorporated to strengthen causal inference in these data, meeting challenges such as disentangling the contributions of factors that tend to cluster together due to lifestyle or social conditions. Genomics and population neuroscience are contributing to the converging evidence that prenatal factors matter for psychotic disorders, and yielding insights into mechanisms. We still do not have definitive evidence that a specific modifiable prenatal exposure is a cause of psychotic disorders. There is much room for optimism, however, as new approaches and data bases come to fruition.

As we move closer to definitive results, it becomes important to consider how these results could be incorporated into public health initiatives to promote healthy pregnancy. Some results might yield further evidence for preventive actions already incorporated into healthy pregnancy initiatives, such as recommended vaccinations and nutritional supplements. It seems likely, however, that emerging results will require us to consider public health actions that go beyond these simplest scenarios. Therefore, it would be appropriate in the long-term to adopt a more holistic public health framework for research. For this purpose, three central points would require consideration: What do we mean by a healthier pregnancy? Should we broaden the offspring outcomes beyond psychotic disorders? What could we gain by focusing on the population distribution of relevant prenatal factors that lie on a continuum?

A universally applicable definition of "healthier pregnancy" is elusive, and any particular measure needs justification for purpose and context. From a life course perspective, characteristics of a pregnancy may be beneficial for some offspring health outcomes and harmful for others. Among many examples, pregnancy characteristics that increase birthweight may reduce offspring risk of psychotic disorders but increase offspring risk of pre-menopausal breast cancer[2,7], and advanced paternal age at conception may increase risk of psychotic disorders but lower offspring risk of cardiovascular disease[8]. Moreover, across different contexts, the characteristics and outcomes that need most emphasis will be different.

Neurodevelopmental delays, low cognitive performance, and persistent subclinical psychotic experiences in children are associated with increased risk of subsequent psychotic disorders. These outcomes are manifest earlier and are more common than psychotic disorders; therefore, they are often more amenable to investigation. They have been related to prenatal experiences; however, like for psychotic disorders, the evidence is not definitive. Furthermore, recent work on the structure of psychopathology supports a dimensional transdiagnostic perspective[9]. From this perspective, preventing these earlier outcomes could substantially reduce risk of psychotic disorders, and probably other psychiatric disorders too, and could have more public health significance. By contrast, we may also find that certain prenatal factors are related to a subgroup of frank psychotic disorders and not to these earlier antecedents; hence the need to investigate the breadth of related outcomes.

Characteristics of a pregnancy may be related to psychotic disorders on a continuum. A large study found that lower birthweight was associated with increased risk of psychotic disorders, but across a broad continuum, implying that a shift in the entire distribution of birthweight (or the causal factor it represents) in the population might do more for prevention of psychotic disorders than reducing the number of low birthweight babies[2]. Furthermore, across the continuum of birthweight, lower birthweight was associated with all treated psychiatric disorders, not only with psychotic disorders. Caution is needed, however, because the relationship of prenatal factor and outcome may not be linear, but rather J-shaped or U-shaped.

We suggest that, alongside the currently dominant approach to research on prenatal factors and psychotic disorders, it could

be useful to set the stage for a more holistic program of research on healthy pregnancy and prevention of psychosis. We have highlighted three central questions that could be amenable to research and might be significant for public health interventions. We should bear in mind, however, that the results of such research may offer guidance, but may not provide unequivocal answers.

Finally, we note that, with few exceptions[3], studies of prenatal factors and psychotic disorders have been done in high-income countries. This makes it difficult to generalize any holistic framework to lower-resource settings, where maternal exposures and conditions affecting pregnancy are different, and access to prenatal care is more limited. Interventions may need to be integrated into broadly conceived programs, such as the Maternal Health Thematic Fund[10]; and we may need to consider, for example, whether reducing maternal mortality and obstetric fistulas could result in less childhood trauma and thereby benefit offspring neurodevelopment. A global approach to healthy pregnancy and psychosis will depend upon the expansion of research to diverse low- and middle-income country settings.

Ezra Susser[1,2], Katherine Keyes[1], Franco Mascayano[1]
[1]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA; [2]Division of Social Psychiatry, New York State Psychiatric Institute, New York, NY, USA

1. Susser E, Terry MB, Matte T. Paediatr Perinat Epidemiol 2000;14:98-100.
2. Abel KM, Wicks S, Susser ES et al. Arch Gen Psychiatry 2010;67:923-30.
3. Susser E, St Clair D. Soc Sci Med 2013;97:325.
4. Susser E, Keyes KM. JAMA Psychiatry 2017;74:349-50.
5. Magnus P, Birke C, Vejrup K et al. Int J Epidemiol 2016;45:382-8.
6. Roffman JL, Eryilmaz H, Huntington FC et al. Effects of prenatal exposure to population-wide folic acid fortification on cortical thickness in youth. Presented at the Annual Meeting of the Society for Biological Psychiatry. Atlanta, 2016.
7. Michels KB, Xue F. Int J Cancer 2006;119:2007-25.
8. Aviv A, Susser E. Int J Epidemiol 2013;42:457-62.
9. Clark LA, Cuthbert B, Lewis-Fernández R et al. Psychol Sci Public Interest 2017;18:72-145.
10. The Maternal Health Thematic Fund. Towards equality in access, quality of care and accountability. New York: United Nations Population Fund, 2017.

DOI:10.1002/wps.20554

Serotonin, psychedelics and psychiatry

Serotonin is a key neuromodulator known to be involved in brain development, perception, cognition, and mood. However, unlike as with dopamine for example, a compelling unified theory of brain serotonin function has not yet been established. This is likely due to the exceptional complexity of the serotonin system, with its 14+ receptors, over twice the number identified for any of the other major neuromodulator systems[1].

Serotonin has been implicated in several major psychiatric disorders, and most obviously in depression. Chronic medication with selective serotonin reuptake inhibitors (SSRIs) remains the dominant treatment for unipolar depression, and SSRI prescription rates have been increasing year-on-year at record levels. Such widespread SSRI use has not noticeably impacted on depression prevalence, however, and questions continue to be asked about the safety, efficacy and general philosophy of chronic pharmacotherapy.

Historically, psychiatry has been a divided house, with the psychodynamic model dominating the first half of the 20th century, and the biomedical model ever since. It is natural for early perspectives within nascent disciplines to overshoot in confidence before maturing and moderating over time. Such has been the case with psychodynamic psychology for example, and there are reasons to suspect that something similar may be happening in relation to the pharmacological model.

This subtle shift in perspective is especially evident in contemporary serotonin and depression research. Until recently, it was not unusual to hear patients, doctors and even psychiatrists speak with presumed authority about how deficient serotonin functioning is causal of depression, offering solace in the view that "serotonin is to blame". As with genetic determinism, one should be mindful of the emotional function of such explanations – especially in psychiatry, the most personal of medical disciplines.

So what is the relationship between serotonin and depression? A fair (but unsatisfactory) answer to this question is that "it is complex". Not wishing to sit on the fence, however, a more constructive statement is that there is increasing evidence that serotonergic processes play a critical role in mediating an individual's sensitivity to context[2]. For example, within the last decade, seminal work has been done to demonstrate how genetic variation within[3] and pharmacological manipulations of[4] the serotonergic system interact significantly with environmental factors to determine outcomes in mental health. The natural implication is that the pure pharmacological model can explain only part of the mental health picture.

What, then, is the alternative? By implication, we should be looking for a hybrid model, a middle-way, that combines the precision, potency and cost-effectiveness of biomedicine with the depth of insight and roundedness of psychology. There is already evidence that SSRIs, in combination with evidence-based psychotherapies, offer (marginally) superior efficacy over either treatment alone[5] – but should our search stop here?

In 1975, the Czech psychiatrist S. Grof compared the potential impact of psychedelic drugs on psychiatry to that of the

microscope on biology and, while this analogy may strike some as laughable, let us reflect for a moment that human research with psychedelics has been effectively moribund since the restrictive drug policy reforms of the 1960s-70s, and has only recently been revived[6].

Classic serotonergic psychedelics – such as LSD, psilocybin and dimethyltryptamine – all possess agonist properties at the 5-HT2A receptor subtype, and 5-HT2A receptor agonism is known to be the pharmacological trigger of the "psychedelic experience"[1]. Crucially, there is also a wealth of evidence to implicate 5-HT2A receptor signaling in processes of plasticity, such as neurogenesis, neurodevelopment, learning, extinction learning, cognitive flexibility and enhanced environmental sensitivity[1].

Added to this, the subjective quality of a psychedelic experience is highly susceptible to contextual influence, for example from the environment in which it occurs as well as from the expectations of the "tripper" and those around him or her[2]. Moreover, the quality of an acute psychedelic experience appears to be a highly reliable predictor of subsequent long-term mental health outcomes[7]. Another predictor of long-term psychological outcomes is the degree of increase in the complexity or "entropy" of brain activity recorded during the psychedelic experience, and this brain effect is hypothesized to be relatively unique to psychedelics, and key to an understanding of their exceptional phenomenology and therapeutic potential[8].

Within the last 12 years, a growing body of evidence, albeit from mostly small scale pilot studies, has suggested that psychedelics, combined with contextual manipulation (such as music listening and psychological support), can offer a safe and effective treatment for a range of different psychiatric disorders[6]. Where successful, the treatment effect appears to be rapid and enduring. Moreover, promising outcomes have not just been seen in depression, but in addiction and other disorders as well[6]. That just one or two treatment sessions can yield therapeutic effects lasting for several months is unprecedented in modern psychiatry. Of course, incredible claims require credible evidence but, with large randomized controlled trials beginning with psilocybin for depression[9], the required roads are being laid.

A simple and plausible model of therapeutic mechanisms of psychedelic treatments would greatly complement this ongoing clinical work. The thesis is put forward here that serotonin differentially encodes behavioral and physiological responses to uncertainty. More specifically, it is proposed that the limbic-rich inhibitory postsynaptic 5-HT1A receptor subtype provides basal control during normal conditions, via moderating emotion and anxiety, and promoting a generalized patience. On the other hand, the cortically-rich 5-HT2A receptor subtype is hypothesized to engage more during conditions of crisis, when the above-mentioned default mechanism becomes suboptimal, e.g. when an individual's internal and/or external milieu becomes so changeable and/or inconsistent with his/her prior beliefs and behaviors that significant revisions become mandated[1].

Viewed through a Bayesian lens, it is proposed that the principal functional effect of 5-HT2A receptor stimulation is to relax prior assumptions or beliefs, held at multiple levels of the brain's functional hierarchy: perceptually, emotionally, cognitively and philosophically (e.g., in terms of biases). In so doing, it opens a door to heightened sensitivity to context[2], an ideal pre-condition for effective change.

Robin L. Carhart-Harris
Psychedelic Research Group, Neuropsychopharmacology Unit, Imperial College London, London, UK

1. Carhart-Harris RL, Nutt DJ. J Psychopharmacol 2017;31:1091-120.
2. Carhart-Harris RL, Roseman L, Haijen E et al. J Psychopharmacol 2018;32:725-31.
3. Caspi A, Hariri AR, Holmes A et al. Am J Psychiatry 2010;167:509-27.
4. Alboni S, van Dijk RM, Poggini S et al. Mol Psychiatry 2017;22:552-61.
5. Khan A, Faucett J, Lichtenberg P et al. PLoS One 2012;7:e41778.
6. Carhart-Harris RL, Goodwin GM. Neuropsychopharmacology 2017;42:2105-13.
7. Roseman L, Nutt DJ, Carhart-Harris RL et al. Front Pharmacol 2018;8:974.
8. Carhart-Harris RL. Neuropharmacology (in press).
9. COMPASS Pathways and Worldwide Clinical Trials. https://www.businesswire.com/news/home/20171206005237/en/COMPASS-Pathways-Partners-Worldwide-Clinical-Trials-Conduct.
10. Matias S, Lottem E, Dugué GP et al. Elife 2017;6:e20552.

DOI:10.1002/wps.20555

Insomnia and inflammation: a two hit model of depression risk and prevention

Depression is projected to increase by 2030 to a position of the greatest contributor to illness burden, due to its nearly 20% prevalence and its over 75% rate of recurrence. Moreover, even when pharmacological treatments are delivered, only about 30% of depressed adults achieve remission. The National Academy of Medicine has called for efforts to develop, evaluate and implement prevention strategies focused on depression[1]. However, to define those to be targeted for depression prevention, it is first necessary to identify biobehavioral factors of greatest risk salience.

Sleep disturbance (i.e., insomnia) is estimated to occur in 15% of the population, with rates as high as 70% in primary care patients[2]. Among depressed patients, sleep disturbance is one of the most frequent complaints, which often persists to serve as a potent predictor of depression recurrence[2]. Because use of antidepressant medications does not mitigate this risk[3], interventions that target sleep disturbance to prevent depression relapse are needed.

Yet, not all persons with sleep disturbance develop an incident depression, which raises the possibility that other factors act in concert with insomnia to instigate the onset of clinically significant depressive symptoms. In addition, it is not known how insomnia gets converted into biological and affective risk for depression, which is critical for identification of molecular targets for pharmacological interventions, and for refinement of insomnia treatments that target affective responding.

Substantial observational, prospective and experimental data show that sleep disturbance is associated with increases in systemic markers of inflammation, such as C-reactive protein and interleukin-6[4], which have been found to predict depression. Similarly, extremes of sleep duration, such as sleeping less than 6 hours or more than 8 hours per night, lead to elevated levels of systemic inflammation[4]. Finally, experimental sleep loss is found to induce an activation of inflammatory biology dynamics at multiple levels of analysis, including increases in systemic inflammation; monocytic production of pro-inflammatory cytokines; activation of the nuclear factor (NF)-κB transcription control pathway and the signal transducer and activator of transcription (STAT) family proteins; transcription of interleukin-6 and tumor necrosis factor mRNA; and expression of the pro-inflammatory transcriptome[6]. Interestingly, such immune activation in response to sleep loss is more robust in younger aged adults, and in women as compared to men, consistent with epidemiologic evidence showing that younger aged women are at greatest risk of depression.

Given that inflammation can elicit profound behavioral changes, which include the initiation of depressive symptoms such as sad mood, anhedonia, fatigue, psychomotor retardation and social-behavioral withdrawal[5,7,8], the inflammatory biotype induced by sleep disturbance may be a key phenomenon driving depression pathogenesis and recurrence. Indeed, sleep disturbance and depression overlap with several somatic conditions known to have an inflammatory basis, such as asthma, rheumatoid arthritis and cardiovascular disease.

Inflammation is not static, but rather shows dynamic variability, due in part to many contributing factors, including specific diseases (i.e., infections) and psychosocial factors (i.e., interpersonal stress)[5]. Such acute increases in inflammation may account for the prospective association between these multi-level processes and subsequent onset of depression.

For example, in controlled experimental models that mimic exposure to an infectious challenge, inflammatory activation induces increases in depressed mood which correlate with activation of brain regions recognized for their role in the pathophysiology of major depressive disorder, and with decreases in reward processing or anhedonia, which correlate with a down-regulation of ventral striatum activity[5].

Moreover, such experimental strategies have yielded support for a "two hit" model of depression, in which sleep disturbance serves as a vulnerability factor to increase severity of depressive symptoms following exposure to an inflammatory challenge[6], consistent with clinical observations that risk of depression is heightened when sleep disturbance occurs in concert with inflammatory states such as an infectious challenge or psychological stress. Alternatively, there is evidence that inflammation itself can serve as a vulnerability factor and increase the risk of depression when persons with the inflammatory biotype experience sleep disturbance.

If insomnia is associated with both inflammation and depression, and in turn, inflammation signals depressive symptom onset, then a credible hypothesis posits that treatment of sleep disturbance might reverse inflammation, and reduce the risk of depression. Emerging evidence supports this possibility.

Among the various treatment options for insomnia, cognitive behavioral therapy for insomnia (CBT-I) is recognized to be the "gold standard", with effects as robust and more durable than pharmacological therapies. Using randomized controlled trial designs, CBT-I induces robust improvements in insomnia outcomes, which map onto long-term (i.e., one year) and large (>50%) decreases in levels of C-reactive protein, as well as decreases in the proportion of insomnia patients whose C-reactive protein levels are considered high risk (>3.0 mg/dl)[6]. Importantly, these improvements in insomnia and inflammation temporally coincide with decreases in depressive symptoms.

Additionally, mind-body interventions such as tai chi (i.e., a movement meditation) and mindfulness meditation, known to target stress response mechanisms, have been found to be non-inferior to CBT-I in the treatment of insomnia[9], and also to reverse the insomnia related inflammatory leukocyte transcriptional profile (i.e., genes regulated by the pro-inflammatory NF-κB/Rel family) and activation of cellular inflammation, with effects greater than those observed following treatment with CBT-I.

Whereas antagonism of endogenous inflammation by potent cytokine antagonists appears to reduce depressive symptoms, at least in those depressed patients who evidence an inflammatory subtype of depression[5], such treatments are expensive and carry the risk of adverse effects not reported to occur with behavioral or mind-body interventions, making these latter non-pharmacologic therapies scalable for delivery in the community to improve insomnia outcomes and reduce inflammation, with possible effects in preventing depression.

Insomnia and inflammation act in concert as "two hits" to identify a population who is especially vulnerable for the occurrence and/or recurrence of depression. Treatments that target the inflammatory biotype and/or the insomniac behavioral phenotype are emerging as promising strategies to prevent depression.

Michael R. Irwin[1], Dominique Piber[1,2]
[1]Cousins Center for Psychoneuroimmunology, Semel Institute for Neuroscience and Human Behavior, Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, University of California, Los Angeles, CA, USA; [2]Department of Psychiatry, Charité Universitatsmedizin, Berlin, Germany

The authors acknowledge the support provided by US National Institutes of Health grants R01AG051944, R01AG026364, R01CA160245 and R01CA207130, Max Kade Foundation, and the Cousins Center for Psychoneuroimmunology.

1. Institute for Medicine. Preventing mental, emotional, and behavioral disorders among young people: progress and possibilities. Washington: National Academy Press, 2009.
2. Irwin MR. Annu Rev Psychol 2015;66:143-72.
3. Cho HJ, Lavretsky H, Olmstead R et al. Am J Psychiatry 2008;165:1543-50.
4. Irwin MR, Olmstead R, Carroll JE. Biol Psychiatry 2016;80:40-52.
5. Slavich GM, Irwin MR. Psychol Bull 2014;140:774-815.
6. Irwin MR, Opp MR. Neuropsychopharmacology 2017;42:129-55.
7. Dantzer R. Biol Psychiatry 2012;71:4-5.
8. Irwin MR, Cole SW. Nature Rev Immunol 2011;11:625-32.
9. Irwin MR, Olmstead R, Carrillo C et al. J Clin Oncol 2017;35:2656-65.

DOI:10.1002/wps.20556

Conditioned hallucinations: historic insights and future directions

Perception is a process of *unconscious inference*, based on a model of one's surroundings which is combined with incoming sensory evidence. Pavlov suggested that this process involved classical conditioning. The associations formed by learning from experience comprise the world model. We have known since 1895 that such learned expectations may in fact penetrate perceptual inference so profoundly that they induce hallucinations[1].

Dating back to J. Konorski in the 1960s, associatively retrieved internal representations have been implicated in the genesis of hallucinations in rats[2]. For example, a hungry rat is presented with a tone and subsequently a sweet sugar solution. The rat learns after only a few trials that the tone predicts sugar. The tone evokes a highly realistic, sensory representation of the sugar, which the rat has trouble distinguishing from reality. With extended training, rats stop having these cue-induced hallucinations, but not in animal models that recapitulate the biology of psychosis[3].

In humans, consistent pairings between the faint illumination of a bulb and a near-threshold tone presentation caused subjects to report hearing tones, even when none were presented[4]. Voice-hearers with psychosis may be more susceptible to this effect[5]. Auditory stimuli can also cue expectations: a salient 1-kHz tone can, through repeated association with a faint visual stimulus, induce visual hallucinations[6]. These experiences even transfer out of the laboratory: subjects later reported seeing the conditioned visual stimulus on their television screen when none was presented, conditional on hearing a 1-kHz tone[6].

In an adaptation of these classic experiments, we recently recruited four groups of subjects for participation in a functional imaging experiment[7]. The four groups differed in having or not a diagnosis of a psychotic illness and having or not daily hallucinations, resulting in groups of those with psychosis and hallucinations, with psychosis and without hallucinations, without psychosis and with hallucinations, and without either psychosis or hallucinations. After learning the association between the visual and auditory stimuli, all groups confidently reported hearing tones that had not been presented (*conditioned hallucinations*). During these, they activated a network of regions previously identified during symptom-capture of auditory hallucinations (e.g., bilateral anterior insula, association auditory cortex, inferior frontal gyrus, superior temporal gyrus, cerebellar vermis, parahippocampal gyrus, and anterior cingulate). However, those with hallucinations, whether or not they had a diagnosable psychotic illness, reported conditioned hallucinations at a much higher rate than those without.

We next employed a formal computational model of perception: a three-tiered hierarchical Gaussian filter (HGF)[8]. The HGF uses participant responses and the task structure to estimate perceptual belief across three levels of abstraction. The first level of the model (X_1) represents whether the subject believes that a tone was present or not on each trial. The second level (X_2) models belief that visual cues predict tones. The third level (X_3) is the change in belief about the contingency between visual and auditory stimuli (i.e., volatility of X_2). Those with hallucinations demonstrated higher degrees of perceptual belief on the first two layers (X_1 and X_2) and an over-reliance on prior beliefs, which correlated with activity in insula, superior temporal gyrus and other nodes in the network active during conditioned hallucinations. Those with psychosis, regardless of whether they had hallucinations or not, were less likely to detect changes in the statistical structure of the task (X_3) compared to non-psychotic participants, activating cerebellum and parahippocampal gyrus less while encoding the volatility of the light-sound contingency.

The model differentiated those with hallucinations from those without, as well as those with psychosis from those without. These computational metrics may hasten the detection of those at risk for hallucinations and psychosis. The computational dissection of the circuit underlying conditioned hallucinations allows for identification of nodes within that circuit that sub-serve specific computational functions. Results indicate that insula and superior temporal gyrus are particularly involved in encoding low-level stimulus beliefs, while cerebellar vermis and parahippocampal gyrus are critical for encoding the volatility of learned contingencies.

This dissection has important implications for the use of repetitive transcranial magnetic stimulation (rTMS) and other forms of neuromodulation as potential treatments. Different

directions of modulation may be beneficial within each region: hyperactivity within superior temporal gyrus and insula may be ameliorated by slow rTMS, inducing inhibitory plasticity. Decreased cerebellar and parahippocampal activity and belief updating may be remediated by potentiating theta-burst stimulation. Targeting superior temporal gyrus in this manner has shown efficacy in the treatment of auditory hallucinations[9]. Likewise, cerebellar vermis, in addition to being driven by multiple sensory modalities, has been implicated in the etiology of schizophrenia and identified as a potential target for deep brain stimulation in treatment of psychosis.

Mathematically, prior weighting is the ratio of the precision of prior knowledge to the precision of incoming sensory information. Therefore, it may potentially be normalized by either decreasing the precision of prior knowledge or increasing the precision of incoming sensory evidence. The precision of sensory evidence appears to depend critically upon cholinergic signaling: acetylcholine increases auditory discrimination abilities and biases perceptual inference toward sensory evidence. Cholinergic receptor blockade diminishes sensory sensitivity, decreases reliance on incoming sensory evidence during perceptual inference, and can both cause spontaneous hallucinations and enhance conditioned hallucinations[10]. By contrast, increased cholinergic signaling ameliorates psychotic symptoms in humans and rodent models of schizophrenia.

Combining long observed phenomena in humans and animals with state-of-the-art computational neuroimaging, this work has yielded new insights into the biology and psychology of hallucinations, that portend new, more precise, therapeutic approaches.

Philip R. Corlett, Albert R. Powers
Department of Psychiatry, Yale University, Connecticut Mental Health Center, New Haven, CT, USA

1. Seashore CE. Stud Yale Psych Lab 1895;3:1-67.
2. Konorski J. Integrative activity of the brain: an interdisciplinary approach. Chicago: University of Chicago Press, 1967.
3. McDannald M, Schoenbaum G. Schizophr Bull 2009;35:664-7.
4. Ellison DG. J Exp Psychol 1941;28:1-20.
5. Kot T, Serper M. J Nerv Ment Dis 2002;190:282-8.
6. Davies P, Davies GL, Bennett S. Perception 1982;11:663-9.
7. Powers AR, Mathys C, Corlett PR. Science 2017;357:596-600.
8. Mathys CD, Lomakina EI, Daunizeau J et al. Front Hum Neurosci 2014; 8:825.
9. Orlov ND, Giampietro V, O'Daly O et al. Transl Psychiatry 2018;8:46.
10. Warburton DM, Wesnes K, Edwards J et al. Neuropsychobiology 1985;14: 198-202.

DOI:10.1002/wps.20557

Balancing validity, utility and public health considerations in disorders due to addictive behaviours

The concept of "behavioural (non-chemical) addictions" was introduced close to three decades ago, and a growing body of literature has emerged more recently on this and related constructs[1,2]. Simultaneously, some authors have noted that the classification of behavioural addictions requires further effort[3,4]. Here we provide an update on this area, emphasizing recent work undertaken during the development of the ICD-11, and addressing the question of whether it is useful to have a separate section on disorders due to addictive behaviours in this classification.

Both the DSM and ICD systems have long avoided the term "addiction" in favor of the construct of "substance dependence". However, the DSM-5 includes gambling disorder in its chapter on substance-related and addictive disorders, and provides criteria for Internet gaming disorder, considering it an entity requiring further study, and highlighting its similarities to substance use disorders[5-7]. In the draft ICD-11, the World Health Organization has introduced the concept of "disorders due to addictive behaviours" to include gambling and gaming disorders[2,8]. These disorders are characterized by impaired control over engagement in the addictive behaviour, the behaviour occupying a central role in the person's life, and continued engagement in the behaviour despite adverse consequences, with associated distress or significant impairment in personal, family, social, and other important areas of functioning[2,8].

An important focus during the development of DSM-5 was on diagnostic validators. Certainly, there is some evidence for overlap between substance use disorders and disorders due to addictive behaviours, such as gambling disorder, on key validators including comorbidity, biological mechanisms, and treatment response[5-7]. For gaming disorder, there is increasing information on clinical and neurobiological features. For a wide range of other putative behavioural addictions, less evidence exists. Further, several of these conditions may also demonstrate overlap with impulse control disorders (in DSM-IV and ICD-10), including comorbidity, biological mechanisms, and treatment response[9].

The groups working on ICD-11 recognize the importance of validators of mental and behavioural disorders, given that a classification system with greater diagnostic validity may well lead to improved treatment outcomes. At the same time, ICD-11 workgroups have focused in particular on clinical utility and public health considerations in their deliberations, with an explicit focus on improving primary care in non-specialist settings, consistent with the ICD-11's emphasis on global mental health. Fine-grained differentiations of disorders and disorder subtypes, even if supported by empirical work on diagnostic validity, are arguably not as useful in contexts where non-specialists provide care. However, associated disability and impairment are key issues in this perspective, supporting the inclusion of gambling and gaming disorders in ICD-11[2,8].

There are multiple reasons why the recognition of disorders due to addictive behaviours and their inclusion in the nosology together with substance use disorders may contribute to improving public health. Importantly, a public health framework for prevention and management of substance use disorders may well be applicable to gambling disorder, gaming disorder, and perhaps some other disorders due to addictive behaviours (although the draft ICD-11 suggests that it may be premature to include in the classification any other disorder due to addictive behaviours outside of gambling and gaming disorders).

A public health framework to considering disorders due to addictive behaviours arguably has a number of specific advantages. In particular, it places appropriate attention on: a) the spectrum from leisure-related behaviour without any harms to health through to behaviour associated with significant impairment; b) the need for high-quality surveys of prevalence and costs of these behaviours and disorders, and c) the utility of evidence-based policy-making to reduce harm.

Although some may be concerned about the medicalization of ordinary living and lifestyle choices, such a framework overtly recognizes that some behaviours with addictive potential are not necessarily and may never become a clinical disorder, and it emphasizes that prevention and reduction of health and social burden associated with disorders due to addictive behaviours may be achieved in meaningful ways by interventions outside the health sector.

Several other criticisms of the constructs of behavioural disorders or disorders due to addictive behaviours may be raised for discussion. We have previously pointed out in this journal that additional work is needed to make strong claims about diagnostic validity[9], and the draft ICD-11 currently also lists gambling and gaming disorders in the section on "impulse control disorders". Relatedly, there is a reasonable concern that the boundaries of this category may be inappropriately extended beyond gambling and gaming disorder to include many other types of human activity. Some of these arguments overlap with those which emphasize the dangers of a reductionist medical model of substance use disorders.

While cognizant of the importance of these issues, our view is that the potentially large burden of disease due to behavioural addictions requires a proportionate response, and that the optimal framework is a public health one.

Here we have outlined reasons why a public health framework that is useful for substance use disorders may also be usefully applied to gambling disorder, gaming disorder and, potentially, other health conditions due to addictive behav-

iours. This argument provides support for including substance use disorders, gambling disorder and gaming disorder in a single section of the chapter on mental, behavioural or neurodevelopmental disorders in ICD-11.

Dan J. Stein[1], Joël Billieux[2], Henrietta Bowden-Jones[3], Jon E. Grant[4], Naomi Fineberg[5], Susumu Higuchi[6], Wei Hao[7], Karl Mann[8], Hisato Matsunaga[9], Marc N. Potenza[10], Hans-Jürgen M. Rumpf[11], David Veale[12], Rajat Ray[13], John B. Saunders[14], Geoffrey M. Reed[15], Vladimir Poznyak[15]

[1]Department of Psychiatry and Mental Health, SA Medical Research Council Unit on Risk and Resilience in Mental Disorders, University of Cape Town, Cape Town, South Africa; [2]Institute for Health and Behaviour, University of Luxembourg, Luxembourg; [3]Division of Brain Science, Imperial College London, London, UK; [4]Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL, USA; [5]University of Hertfordshire, Hatfield, and School of Clinical Medicine, Cambridge University, Cambridge, UK; [6]National Center for Addiction Services Administration, Yokosuka, Kanagawa, Japan; [7]Mental Health Institute, Central South University, Changsha, China; [8]Central Institute of Mental Health, Mannheim, Germany; [9]Department of Neuropsychiatry, Hyogo College of Medicine, Nishinomiya, Japan; [10]Yale University School of Medicine, and Connecticut Council on Problem Gaming, Clinton, CT, USA; [11]Clinic for Psychiatry and Psychotherapy, University of Lübeck, Lübeck, Germany; [12]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; [13]Department of Psychiatry, Swami Rama Himalayan University, Dehradun, India; [14]Disciplines of Psychiatry and Addiction Medicine, University of Sydney, Sydney, NSW, Australia; [15]Department of Mental Health and Substance Abuse, World Health Organization, Geneva, Switzerland

The authors alone are responsible for the views expressed in this letter and they do not necessarily represent the decisions, policy or views of the World Health Organization. The letter is based in part upon work from Action CA16207 "European Network for Problematic Usage of Internet", supported by the European Cooperation in Science and Technology (COST).

1. Chamberlain SR, Lochner C, Stein DJ et al. Eur Neuropsychopharmacol 2016;26:841-55.
2. Saunders JB, Hao W, Long J et al. J Behav Addict 2017;6:271-9.
3. Starcevic V. Aust N Z J Psychiatry 2016;50:721-5.
4. Aarseth E, Bean AM, Boonen H et al. J Behav Addict 2017;6:267-70.
5. Hasin DS, O'Brien CP, Auriacombe M et al. Am J Psychiatry 2013;170:834-51.
6. Petry NM. Addiction 2006;101(Suppl. 1):152-60.
7. Potenza MN. Addiction 2006;101(Suppl. 1):142-51.
8. Saunders JB. Curr Opin Psychiatry 2017;30:227-37.
9. Grant JE, Atmaca M, Fineberg NA et al. World Psychiatry 2014;13:125-7.

DOI:10.1002/wps.20570

Evidence of the clinical utility of a prolonged grief disorder diagnosis

A substantial body of research has shown that prolonged grief disorder (PGD), characterized by persistent and severe separation distress, constitutes a disorder distinct from bereavement-related major depressive disorder (MDD) and post-traumatic stress disorder (PTSD)[1]. Reviewing the available evidence, the work group covering the Disorders Specifically Associated With Stress section in the ICD-11 decided to slate PGD for inclusion as a new stress response syndrome[2]. Still, mental health professionals and laypersons have expressed concerns that diagnosing PGD represents a "medicalization" of normal grief reactions[3]. Fears of the overdiagnosis of normal responses remain[4-6].

As a new disorder, it is of paramount importance to determine whether PGD is a clinically useful diagnosis. According to First[7], a mental disorder or diagnostic system has clinical utility if it: a) helps communication, b) facilitates effective interventions, c) predicts management needs and outcomes, and d) differentiates disorder from non-disorder and comorbid disorders. Whereas a large body of evidence has demonstrated the construct, predictive and incremental validity of PGD, clinicians' perceptions of its clinical utility have yet to be tested experimentally.

To address this gap, our group recently completed a two-phase National Institute of Mental Health (NIMH)-funded randomized controlled trial in the US that evaluated the clinical utility of PGD by examining the impact of providing information about the diagnosis on clinicians' ability to differentially diagnose PGD in "virtual standardized patients" (VSPs). The use of VSPs allowed us to standardize clinical presentations, control influential confounding variables and patient characteristics, and avoid burdening bereaved participants. Using VSPs (rather than written vignettes or clinicians selecting their own patients[8], as has been done in prior studies) increased the external validity of this investigation.

In Phase 1 of the study, video-recorded case vignettes for the VSPs were developed with the input of seven bereavement experts. They reflected cases of PGD, normative grief not meeting criteria for PGD, MDD, and PTSD. Four blinded, expert diagnosticians were asked to review the VSPs and evaluate the cases to establish "gold" or "criterion" standard diagnoses. There was full agreement on 12 of the cases, which were included in Phase 2 of the study.

In Phase 2, clinicians (N=120 completers) were randomized to receive written information about PGD (informed) or not (not informed). Participants were asked about their background and experience working with the bereaved, and were invited to provide a diagnosis and treatment recommendations for four VSPs depicting normative grief, PGD, MDD and/or PTSD. Participants were also surveyed about PGD's clinical utility. Participants included psychiatrists (17%), psychologists (27%), social workers (43%), and other licensed clinicians (13%). They were 76% female and 66% White.

We found that clinicians provided with information about PGD, compared to those not receiving such information, were 4.5 times more likely to diagnose PGD accurately. There were no significant group differences in the likelihood of clinicians accurately diagnosing normative grief, MDD or PTSD, but there were significant between-group differences in treatment recommendations for PGD cases. Clinical utility ratings of the PGD diagnostic criteria were high, with the majority of clini-

cians rating those criteria as easy to use (97%) and overall clinically useful (95%).

There has been significant concern that introducing a diagnosis of prolonged grief would increase the likelihood that clinicians will medicalize or pathologize grief[4-6]. We found, however, that mental health providers who received information about PGD were no more likely to pathologize normative grief than those who did not receive this information in advance of evaluating standardized patients. Furthermore, clinicians who correctly diagnosed PGD were shown to be *less* likely to recommend antidepressants for individuals they accurately diagnosed with PGD and *more* likely to recommend psychotherapies that have direct relevance to PGD symptoms, such as disbelief (emotion-focused therapy), loss of meaning (existential therapy), and persistent suffering (acceptance and commitment therapy). This may reflect clinicians' perception that PGD is less biologically based than, for example, MDD. Although, like the DSM, the PGD tutorial did not offer treatment recommendations, it did describe risk factors that were psychological in nature, which may have affected the recommendations made.

This study also suggests the clinical value of using straightforward diagnostic criteria to distinguish pathological grief from other clinical presentations. The proposed PGD criteria are highly specific, which should reduce the risk of pathologizing normative grief reactions[1]. At the same time, they are sufficiently sensitive to capture those in need[1]. Under-recognition of PGD and misclassifying it as another diagnosis is likely to lead to suboptimal treatment. PGD improves when specific interventions, such as those recommended by the study participants, target unique pathological grief symptoms[9]. The misdiagnosis of PGD as MDD or PTSD may promote the use of inappropriate interventions.

Although this study was limited by a relatively small sample size and by the biases inherent in those who chose to participate, it demonstrates that PGD is perceived and shown to be clinically useful. We therefore believe that educating clinicians about PGD is likely to improve their ability to distinguish normal from pathological grief; to enhance communication between clinicians, patients, and their families; and to assist in the delivery of effective treatments for PGD[7].

Wendy G. Lichtenthal[1,2], Paul K. Maciejewski[2], Caraline Craig Demirjian[1], Kailey E. Roberts[1], Michael B. First[5], David W. Kissane[1-3], Robert A. Neimeyer[4], William Breitbart[1,2], Elizabeth Slivjak[1], Greta Jankauskaite[1], Stephanie Napolitano[1], Andreas Maercker[6], Holly G. Prigerson[2]

[1]Memorial Sloan Kettering Cancer Center, New York, NY, USA; [2]Weill Cornell Medicine, New York, NY, USA; [3]Monash University, Clayton, VIC, Australia; [4]University of Memphis, Memphis, TN, USA; [5]Columbia University Medical Center, New York, NY, USA; [6]University of Zurich, Zurich, Switzerland

The authors would like to thank the experts who reviewed the VSP scripts, including N. Alston, P. Boelen, J.A. García-García, A. Ghesquiere, A. Papa and S. Schachter. They are also indebted to misterBIT for developing the online study platform and to the mental health providers who gave their time to this study. This research was supported by the NIMH grant R21MH095378 and National Cancer Institute grants R35CA197730, K07CA172216 and P30CA008748.

1. Prigerson HG, Horowitz MJ, Jacobs SC et al. PLoS Med 2009;6:e1000121.
2. Maerker A, Brewin CR, Bryant RA et al. World Psychiatry 2013;12:198-206.
3. Breen LJ, Penman EL, Prigerson HG et al. J Nerv Ment Dis 2015;203:569-73.
4. Wakefield JC. World Psychiatry 2013;12:171-3.
5. Prigerson HG, Maciejewski PK. JAMA Psychiatry 2017;74:435-6.
6. Dodd A, Guerin S, Delaney S et al. Patient Educ Couns 2017;100:1447-58.
7. First MB. Prof Psychol Res Pract 2010;41:465-73.
8. Spitzer RL, First MB, Shedler J et al. J Nerv Ment Dis 2008;196:356-74.
9. Mancini AD, Griffin P, Bonanno GA. Curr Opin Psychiatry 2012;25:46-51.

DOI:10.1002/wps.20544

Diet as a hot topic in psychiatry: a population-scale study of nutritional intake and inflammatory potential in severe mental illness

People with severe mental illnesses (SMIs) – including schizophrenia, major depressive disorder (MDD) and bipolar disorder – have excessive caloric intake, a low-quality diet, and poor nutritional status compared to the general population[1,2]. Poor diet increases the risk of diabetes and cardiovascular mortality in this population[3]. Furthermore, excessive consumption of high-fat and high-sugar foods can increase systemic inflammation[4]. Indeed, all classes of SMI show heightened levels of peripheral inflammatory markers, which is linked to worse prognosis in these conditions. However, there currently is an absence of large-scale studies comparing the nutritional intake and inflammatory profile of the diets of individuals with SMIs.

To address this, we used detailed dietary intake data from the baseline phase (2007-2010) of the UK Biobank study[5] to examine differences in nutritional intake and diet-associated inflammation between people with SMIs and the general population. Full details on the UK Biobank, including approval from the National Health Service (NHS) Research Ethics Committee, are available elsewhere[5]. We used patient hospital records to identify individuals with a ICD-10 diagnostic history of recurrent depressive disorder, bipolar disorder (type I or II) or schizophrenia. Additionally, participants' answers to questions from the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) and Patient Health Questionnaire (PHQ), administered at the UK Biobank baseline, were used to identify additional individuals with MDD or bipolar disorder[6]. Participants who fell into multiple psychiatric categories were assigned hierarchically to only one, in this order: schizophrenia, bipolar disorder, MDD. Healthy controls were derived from all UK Biobank participants who had no indication of any previous or pre-

sent psychotic or mood disorder. Individuals with neurological conditions known to affect memory recall were excluded from all groups.

Diet was assessed using a validated computerized questionnaire, the Oxford WebQ[7]. This brief, self-administered 24-hour recall measure queries previous day intake for >200 food items. To capture intra-individual (day-to-day) variation in food intake, the WebQ was administered on up to five separate occasions over a 16-month period. Missing individual data points were imputed as the mean average across all available time points from that individual. The first assessment was administered on-site at the UK Biobank assessment centre; all subsequent administrations were completed online.

A total of 69,843 eligible UK Biobank subjects (mean age 56.5 years, 46.4% male) provided sufficient data for analyses. Of these, 14,619 individuals had MDD, 952 had bipolar disorder, 262 had schizophrenia, and 54,010 were SMI-free. Multivariable linear regression was used to examine differences in total daily energy intake and each macronutrient between the SMI and control samples. Analyses were adjusted for gender and age.

The greatest differences in diet were observed for the schizophrenia sample. Age- and gender-adjusted comparisons to control subjects showed highly elevated intakes in that sample (all $p<0.001$) for total energy (+553.4 kilojoules (kj)/day, SE= 143.8), carbohydrates (+25.4 g/day, SE=4.86), sugar (+16.0 g/day, SE=2.98), total fat (+6.04 g/day, SE=1.77), saturated fat (+3.76 g/day, SE=0.76) and protein (+5.24 g/day, SE=1.51), with no difference in dietary fibre ($p=0.78$).

Individuals with bipolar disorder similarly showed significantly (all $p\leq0.01$) higher intake of total energy (+298.5 kj/day, SE=75.9), carbohydrates (+11.4 g/day, SE=2.57), protein (+1.97 g/day, SE=0.80), sugar (+9.63 g/day, SE=1.57), total fat (+2.40 g/day SE=0.93) and saturated fat (+1.29 g/day, SE=0.40) compared to controls, with no difference in fibre ($p=0.32$).

The MDD sample also showed significantly greater (all $p< 0.001$) age- and gender-adjusted intake in comparison to controls, for total energy (+189.4 kj/day, SE=21.88), carbohydrates (+5.15 g/day, SE=0.74), sugar (+3.11 g/day, SE=0.45), total fat (+2.19 g/day, SE=0.27), saturated fat (+0.96 g/day, SE=0.12) and protein (+1.12 g/day, SE=0.23), along with a small difference in dietary fibre intake (+0.15 g/day, SE=0.06, $p=0.01$).

Sensitivity analyses were conducted to adjust for ethnicity, body mass index (BMI), education and social deprivation. This did not substantially alter the overall findings.

Beyond examining raw macronutrient intakes, we also explored the inflammatory potential of the diet using the dietary inflammatory index (DII®)[4], which has produced consistent positive associations in over ten studies using inflammatory markers including C-reactive protein, interleukin-6 and tumor necrosis factor-α as construct validators[8].

A total of 68,879 participants had provided sufficient dietary intake data across the 18 macro/micronutrient intake parameters relevant for DII® calculation. Multiple linear regression was used to examine if SMI was positively associated with DII scores, adjusting for age, gender, ethnicity, BMI, social deprivation, education and total energy intake. DII scores were significantly elevated in subjects with schizophrenia (B=0.220, SE=0.084, $p=0.009$) and MDD (B=0.031, SE=0.013, $p=0.014$), but not with bipolar disorder ($p=0.27$), compared to controls.

Overall, this population-scale analysis of nutritional intake confirms that people with SMIs have higher intakes of obesogenic nutrients and more inflammatory diets than the general population. Whereas dietary interventions for SMIs often focus exclusively on over-consumption of obesogenic, pro-inflammatory foods, this study shows that further consideration should be given to increasing consumption of nutrient-dense foods that are known to reduce systemic inflammation[4,8].

In terms of both total caloric intake and excess obesogenic nutrients, the worst dietary patterns were observed among people with schizophrenia. This is a notable finding, as these individuals also have significantly higher rates of metabolic disorders and greater premature mortality than individuals with other classes of SMI[3,9], indicating that diet could be a key factor influencing these outcomes.

Indeed, dynamic weight change algorithms predict that each 100 kj of excess energy intake per day will eventually lead to at least 1 kg increase in body weight[10]. Thus, the 553 kj (132 calories) per day excess observed in the schizophrenia sample suggests that dietary differences alone can account for 5-6 kg of the increased body weight observed in this population. Not only does excess caloric, carbohydrate and fat intake increase inflammation, but the concomitant increase in adipose tissue also enhances chronic, systemic inflammation.

The degree to which the heightened systemic inflammation observed in SMIs is attributable to dietary factors needs to be clarified. Sufficiently sized cohort studies, using detailed dietary and psychiatric data alongside biomarkers of inflammation, can provide new insights into of the role of diet in SMIs. Future work should also aim to establish the extent to which heightened dietary inflammation in SMIs independently contributes to the poor physical, psychological and neurocognitive outcomes observed in these populations, which represent a significant public health challenge.

Joseph Firth[1,2], Brendon Stubbs[3,4], Scott B. Teasdale[5,6], Philip B. Ward[6,7], Nicola Veronese[8,9], Nitin Shivappa[10,11], James R. Hebert[10,11], Michael Berk[12-14], Alison R. Yung[2,15], Jerome Sarris[1,13]

[1]NICM Health Research Institute, Western Sydney University, Penrith, NSW, Australia; [2]University of Manchester, Manchester, UK; [3]South London and Maudsley NHS Foundation Trust, London, UK; [4]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; [5]South Eastern Sydney Local Health District, Sydney, NSW, Australia; [6]University of New South Wales, Sydney, NSW, Australia; [7]Ingham Institute of Applied Medical Research, Liverpool, NSW, Australia; [8]IRCCS S. de Bellis, Castellana Grotte, Bari, Italy; [9]National Research Council, Padua, Italy; [10]University of South Carolina, Columbia, SC, USA; [11]Connecting Health Innovations LLC, Columbia, SC, USA; [12]Deakin University, Barwon Health, VIC, Australia; [13]University of Melbourne, Melbourne, VIC, Australia; [14]Orygen, The National Centre of Excellence in Youth Mental Health and Orygen Youth Health Research Centre, Melbourne, VIC, Australia; [15]Greater Manchester Mental Health NHS Foundation Trust, Manchester, UK

This research was conducted using the UK Biobank Resource (application no. 22125). J. Firth is supported by a Blackmores Institute fellowship and Medical Research Council (MRC) doctoral training grant (P117413F07). B. Stubbs is in part funded by the UK National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. J. Sarris is supported by a research fellowship from the National Health and Medical Research Council (NHMRC), Australia (APP1125000). M. Berk is supported by an NHMRC senior principal research fellowship (1059660). The views expressed are those of the authors and not necessarily those of the above entities.

1. Dipasquale S, Pariante CM, Dazzan P et al. J Psychiatr Res 2013;47: 197-207.
2. Firth J, Carney R, Stubbs B et al. Schizophr Bull (in press).
3. Vancampfort D, Correll CU, Galling B et al. World Psychiatry 2016;15: 166-74.
4. Shivappa N, Steck SE, Hurley TG et al. Public Health Nutr 2014;17:1689-96.
5. UK Biobank. The UK Biobank: rationale, design and development of a large-scale prospective resource. Stockport: UK Biobank Coordinating Centre, 2007.
6. Smith DJ, Nicholl BI, Cullen B et al. PLoS One 2013;8:e75362.
7. Liu B, Young H, Crowe FL et al. Publ Health Nutr 2011;14:1998-2005.
8. Shivappa N, Hebert JR, Marcos A et al. Mol Nutr Food Res 2017;61(6).
9. Liu NH, Daumit GL, Dua T et al. World Psychiatry 2017;16:30-40.
10. Hall KD, Heymsfield SB, Kemnitz JW et al. Am J Clin Nutr 2012;95:989-94.

DOI:10.1002/wps.20571

Global mental health: how are we doing?

The World Health Organization (WHO) has just launched the 2017 edition of the Mental Health Atlas, consisting of updated information from nearly 180 countries[1].

Data from the Atlas are used to monitor mental health policies, laws, programmes and services across WHO Member States[2,3], and to track progress in the implementation of the WHO's Mental Health Action Plan 2013-2020[4]. Additionally, Atlas 2017 is particularly relevant as WHO is embarking on a major transformation to increase its impact at country level and to be fit-for-purpose in the era of the Sustainable Development Goals[5].

With the aim of stimulating the global mental health community to make further progress in relation to mental health policies, laws, programmes and services, we present here the main findings of Atlas 2017, and describe progress towards the achievement of the four objectives of the Mental Health Action Plan.

The production of Atlas 2017 followed a strict methodological process, involving development of a questionnaire and an associated completion guide, management of an online data collection system, validation of information and responses, liaison with Member States and WHO Regional and Country Offices, as well as analysis and interpretation of data[1].

A total of 177 out of 194 WHO Member States (91%) completed, at least partially, the Atlas questionnaire, with a submission rate above 85% in all WHO Regions.

In terms of mental health governance, 72% of Member States reported to have a standalone policy or plan for mental health, and 57% to have a standalone mental health law. Importantly, 94 countries, i.e. 68% of those that responded or 48% of all WHO Member States, have developed or updated their policies or plans for mental health in line with international and regional human rights instruments. Similarly, 76 countries, i.e. 75% of those that responded or 39% of all WHO Member States, have developed or updated their law for mental health in line with international and regional human rights instruments.

In terms of financial and human resources for mental health, Atlas 2017 shows that the levels of public expenditure on mental health are very meagre in low- and middle-income countries, and more than 80% of these funds go to mental hospitals. Globally, the median number of mental health workers is 9 per 100,000 population, with extreme variation, from below 1 in low-income countries to over 70 in high-income countries.

Wide variation was also observed in terms of number of mental health beds, which range from below 7 in low- and lower middle-income countries to over 50 in high-income countries per 100,000 population. Similar variation was documented for child and adolescent beds, which range from below 0.2 in low- and lower middle-income countries to over 1.5 in high-income countries.

A total of 123 countries, equivalent to 69% of those that responded or 63% of all WHO Member States, reported at least two functioning national, multisectoral mental health promotion and prevention programmes. Out of almost 350 functioning programmes, 40% were aimed at improving mental health literacy or combating stigma and 12% were aimed at suicide prevention.

As far as progress towards the achievement of the four objectives of the Mental Health Action Plan is concerned, Atlas 2017 highlighted the following:

Target 1.1: 80% of countries will have developed or updated their policies or plans for mental health in line with international and regional human rights instruments (by the year 2020). The proportion of countries fulfilling this target slightly increased from 45% (Atlas 2014) to 48% (Atlas 2017) of all WHO Member States.

Target 1.2: 50% of countries will have developed or updated their law for mental health in line with international and regional human rights instruments (by the year 2020). The proportion of countries fulfilling this target slightly increased from 34% (Atlas 2014) to 39% (Atlas 2017) of all WHO Member States.

Target 2: Service coverage for severe mental disorders will have increased by 20% (by the year 2020). Although Atlas 2017 made a substantial effort to increase the reliability of data, service coverage for severe mental disorders was not computable. The treated prevalence for psychosis, bipolar disorder and depression was 171.3, 41.0 and 95.6 per 100,000 population, respectively.

Target 3.1: 80% of countries will have at least two functioning national, multisectoral mental health promotion and prevention programmes (by the year 2020). The proportion of countries fulfilling this target increased from 41% (Atlas 2014) to 63% (Atlas 2017) of all WHO Member States.

Target 3.2: The rate of suicide in countries will be reduced by 10% (by the year 2020). According to WHO data on suicide, suicide rate slightly decreased from 11.4 to 10.5 per 100,000 population from 2014 to 2017.

Target 4: 80% of countries will be routinely collecting and reporting at least a core set of mental health indicators every two years through their national health and social information systems (by the year 2020). The proportion of countries fulfilling this target slightly increased from 64 countries, 33% of all WHO Member States (Atlas 2014), to 71 countries, 37% of all WHO Member States (Atlas 2017).

A number of limitations should be recognized when examining Atlas 2017 data. A first limitation is that some countries were not able to provide data for a number of indicators. For example, data on service coverage and utilization were not available for many countries, possibly due to the still limited implementation of national information systems.

Second, although a large number of countries submitted questionnaires for both Atlas 2014 and Atlas 2017, the list of countries completing both data points within each of the questions was sometimes different. This adds some constraints to comparisons of data over time between the two Atlas versions.

Third, it is important to acknowledge the limitations associated with self-reported data, particularly in relation to qualitative assessments or judgements, often being made by a single focal point. It is nevertheless important to note that the process of country-level mental health data collection, started by WHO in 2000 in partnership with WHO Member States, has progressively improved in terms of quality and quantity of information, and is expected to make further progress over the next years.

This continuous effort has also contributed to consolidate an epidemiological and evaluative culture among WHO Member States, which is a major achievement considering the important public health choices that countries are continuously required to make[6]. The next important step is that countries start to use the data they collected to improve their mental health system and to monitor progress.

Data included in the Mental Health Atlas 2017 demonstrate the commitment of countries to track progress towards the implementation of the WHO's Mental Health Action Plan 2013-2020. Progressive development is being made in relation to mental health policies, laws, programmes and services across WHO Member States.

Findings from Atlas 2017 suggest that the global targets established by the Mental Health Action Plan will be reached only if there is a collective global commitment that leads to substantial investment and expanded efforts at country level in relation to mental health policies, laws, programmes and services across WHO Member States.

Fahmy Hanna[1], Corrado Barbui[2,3], Tarun Dua[1], Antonio Lora[4], Marieke van Regteren Altena[5], Shekhar Saxena[1]

[1]Department of Mental Health and Substance Abuse, World Health Organization, Geneva, Switzerland; [2]WHO Collaborating Centre for Research and Training in Mental Health and Service Evaluation, Department of Neuroscience, Biomedicine and Movement Sciences, Section of Psychiatry, University of Verona, Verona, Italy; [3]Cochrane Global Mental Health, University of Verona, Verona, Italy; [4]Department of Mental Health, Lecco Hospital, Lecco, Italy; [5]National Social Inclusion Office, Health Service Executive, Dublin, Ireland

1. World Health Organization. Mental Health Atlas. Geneva: World Health Organization, 2017.
2. Saxena S, Sharan P, Garrido M et al. World Psychiatry 2006;5:179-84.
3. Lora A, Hanna F, Chisholm D. Epidemiol Psychiatr Sci (in press).
4. World Health Organization. Mental health action plan 2013-2020. Geneva: World Health Organization.
5. Saxena S, Maj M. World Psychiatry 2017;16:1-2.
6. Janse van Rensburg B, Bhugra D, Saxena S. World Psychiatry 2018;17:116-7.

DOI:10.1002/wps.20572

Global trends in high impact psychiatry research

The utility of psychiatry research, when indexed as the number of times it is cited by other researchers, is most centrally predicted by the impact factor of the publishing journal[1]. In keeping, scientists' track record in high impact journals is directly related to academic promotions and merit increases in academic psychiatry institutions[2], and the acquisition of research funding[3]. Thus, much pressure exists for psychiatric researchers to publish their research in high impact psychiatry journals.

Trends in high impact psychiatry journals are therefore important to examine. This is especially true when considering that psychiatry is thought to be particularly vulnerable to publication bias. While studies have noted possible biases in the preponderance of underpowered studies stemming from phar-

macological company sponsorship[4], and the underrepresentation of research stemming from low- and middle-income countries[5], less evidence exists in detailing possible bias relating to the content of psychiatric research. We therefore assessed global publication trends in high impact general psychiatry journals according to specific illness typology and methodology of research. In total, we extracted 12,218 manuscripts from six high impact psychiatry journals, over a 16-year period (2000-2016).

In terms of psychopathology, stable trends emerged relating to comorbid psychiatric presentations, mood disorders and psychotic disorders, which remained the most studied psychopathologies throughout this period. Equally robust trends emerged around the least commonly studied psychopathologies – personality, eating and obsessive-compulsive disorders. The coverage of two of them (eating and personality disorders) consistently decreased over time.

In terms of manuscript typology, less stable trends emerged and, although epidemiological and behavioral studies were the most prevalent type across the entire period, they appeared to be steadily decreasing in prevalence over time. By contrast, editorial manuscripts almost tripled in prevalence throughout this 16-year period and, by 2016, these manuscripts made up the most common manuscript type in high impact psychiatry journals.

In interpreting these findings, several possibilities are worth considering. First, while data relating to the ratio of submitted versus accepted manuscripts are unavailable, one possibility is that more manuscripts relating to comorbid presentations, mood disorders and psychotic disorders are submitted to high impact psychiatry journals, relative to other psychiatric disorders, which may account for their greater overall prevalence.

Relatedly, evidence from the US National Institute of Mental Health suggests that research pertaining to mood and psychotic disorders consistently receives more funding than other psychopathologies[6]. Greater financial support may allow for a greater number of large-scale high-quality studies in the domain of mood and psychotic disorders, relative to other disorders, which may account for the greater overall volume of these studies in high impact psychiatry journals.

Similarly, the decreasing prevalence of epidemiological and behaviorally-focused studies in high impact psychiatry journals may reflect a declining submission rate. This would be consistent with the ongoing paradigm shift which posits that innovative research ought to uncover the precise pathophysiology of psychiatric disorders, rather than documenting behavioral sequalae or epidemiological trends[7]. However, it is important to note that the presence of genetic, cellular or neuroimaging studies in high impact psychiatry journals is not increasing.

A second possibility in accounting for these trends, in the event of comparable submission rates, is a potential greater acceptance rate of manuscripts relating to specific psychopathologies and methodologies. In this instance, it would be important to consider what drives a higher acceptance rate, and therefore a greater emphasis on scientific discovery, among some specific psychopathologies relative to others. For example, in considering nationally representative data relating to the prevalence of psychiatric disorders in the US, evidence suggests that anxiety disorders are the most common psychiatric disorders, yet they represent only 4-7% of manuscripts in high impact psychiatry journals, and their coverage is steadily declining over time. In contrast, psychotic disorders feature prominently in high impact psychiatry journals, representing 20-27% of all manuscripts, yet demonstrate a lifetime prevalence of less than 1%.

Burden of illness data suggest that mood disorders are the leading cause of disability-adjusted life years (DALYs) stemming from psychiatric illness[8], which is consistent with the prevalence of manuscripts relating to these disorders in high impact psychiatry journals. However, other major contributors to DALYs stemming from psychiatric illness include anxiety disorders and substance use disorders, which represent a notably more marginal focus in high impact psychiatry journals.

The rapid increase in editorial manuscripts over this period is particularly noteworthy, and it is unclear whether this reflects a sharp increase in the overall volume of editorial submissions, relative to other types of methodologies, or a greater acceptance rate among editorial manuscripts. In fact, editorial manuscripts are unique in that they are typically solicited from editors, although it is unclear as to whether this alters the course of editorial and peer review. This sharp increase is important, however, in that editorials do not offer novel data, and more frequently offer novel perspectives or syntheses of emerging findings.

This may involve the advantage of assisting in the interpretation and contextualization of complex empirical findings, while simultaneously increasing the citation count of articles included in the journal (although not the impact factor, which is based on citations received by articles in the following two years), since editorials typically relate to manuscripts in the same journal edition.

In summary, these findings suggest a demonstrable skew in high impact psychiatry journals towards comorbid, mood and psychotic disorders, alongside editorial type manuscripts. These trends are important as they relate to the dissemination and uptake of psychiatric research, since potential bias may stymie advances in research relating to less commonly featured psychopathologies.

Our results suggest that trends in high impact psychiatry journals may be comparable to trends in psychiatric research funding, which is consistent with the notion that extramural research funding appears to be linked to researchers' track records in high impact journals[3]. The potentially recursive influence between lower extramural research funding and the lower likelihood of high impact publications may present an additional barrier for research innovation among psychopathologies which attract less extramural funding, which ultimately may result in slower advances for patients.

Stuart B. Murray[1], Eva Pila[2], Jonathan M. Mond[3], Deborah Mitchison[4], Emily Nauman[5], Scott Griffiths[6]
[1]Department of Psychiatry, University of California, San Francisco, CA, USA; [2]Centre for Addiction and Mental Health, Toronto, ON, Canada; [3]Center for Rural Health, University of Tasmania, Launceston, TAS, Australia; [4]Department of Psychology, Macquarie University, Sydney, NSW, Australia; [5]PGSP-Stanford Psy.D. Consortium, Palo Alto University, Palo Alto, CA, USA; [6]School of Psychology, University of Melbourne, Melbourne, VIC, Australia

1. Callaham M, Wears RL, Weber E. JAMA 2002;287:2847-50.
2. MacMaster FP, Swansburg R, Rittenbach K. Acad Psychiatry 2017;41:452-4.
3. Saraykar S, Saleh S, Selek S. Acad Psychiatry 2017;41:455-9.
4. Thornley B, Adams C. BMJ 1998;317:1181.
5. Patel V, Sumathipala A. Br J Psychiatry 2001;178:406-9.
6. Insel TR. The anatomy of NIMH funding. https://www.nimh.nih.gov/funding/funding-strategy-for-research-grants/the-anatomy-of-nimh-funding.shtml.
7. Insel TR, Cuthbert BN. Science 2015;348:499-500.
8. Whiteford HA, Degenhardt L, Rehm J et al. Lancet 2013;382:1575-86.

DOI:10.1002/wps.20573

Mental health initiatives in the workplace: models, methods and results from the Mental Health Commission of Canada

Issues related to mental health in the workplace have become of paramount interest, in part because of the recognition that mental health of employees affects productivity, but also because some workplaces have structurally embedded stressors that may increase the risk of mental health problems. For example, police, fire and emergency services have inherent mental health challenges, which necessitate workplace health promotion.

In addition to the humane argument to support the optimal mental health of workers, there is also a financial argument to address mental health in the workplace. Using an economic model simulation, it was found that a comprehensive screening program for depression had a return on investment of approximately 4:1, based on estimates of presenteeism and absenteeism alone[1].

When issues related to lost productivity, increased disability rates, and the indirect costs associated with hiring and training replacement employees are also considered, employers are well advised to promote optimal mental health – which means building structures and a cultural environment that are supportive of mental health in the workplace.

We have addressed issues related to mental health in the workplace through the Opening Minds initiative of the Mental Health Commission of Canada. This initiative recognizes that the structure of the workplace can increase the probability of mental health problems, and that the willingness of workplace leaders to identify and speak about these issues influences mental health outcomes. For example, negative attitudes towards identifying and treating mental health problems, or direct stigma and discrimination from managers or co-workers, can create significant barriers to self-care and reduce the likelihood that an employee will seek care for mental health issues[2].

In concert with the Opening Minds initiative, we have produced reviews of anti-stigma activities in the workplace[3,4] and of the organizational factors that facilitate and hinder mental health and access to services[5]. We have modified and systematically evaluated more than twenty-four implementations of two workplace programs that aim to directly address stigma, encourage open dialogue and promote personal resilience[6]. We have also conducted qualitative studies of how mental ill health is experienced and managed in the workplace, and of worker perspectives on related training programs. It is our belief that these projects will enable a more sophisticated, compassionate and evidence-based approach to mental illness in the workplace, and allow workplaces to both promote mental health and recognize and address mental health challenges when they occur[7].

The two programs we have developed are the Road to Mental Readiness for First Responders (R2MR) and The Working Mind (TWM).

The R2MR is an adaptation of a program that was created by the Department of National Defense in Canada for military personnel. A notable feature of that program is the use of the mental health continuum model, which encourages participants to conceptualize their mental health, in a non-pathologizing way, on a scale that ranges from good functioning (represented as the color green), through varying degrees of increasing distress and behavioural indicators, color-coded as yellow, orange and red, respectively. The program also encourages four coping skills, adapted from cognitive-behavioral therapy, to maintain and restore mental health, as needed.

Our adaptation of the R2MR program for first responders included enhancement to the discussion about stigma and discrimination related to mental illness in the workplace. We adapted the mental health continuum model and coping skills that build personal resilience for the appropriate context. We also leveraged the research literature suggesting that contact-based education is a successful strategy to impart health-related information. Contact-based education includes the use of video materials of first responders who have experienced and overcome mental health problems, and the involvement of trained peers to deliver the program to their colleagues.

This adapted R2MR program has been delivered to approximately 75,000 participants in Canada. We have consistently evaluated the program, using an open trial methodology, with pre-test, post-test, and 3-month follow-up on primary measures related to stigma and mental health resilience. Across 16

sites and multiple types of first responders (N=4,649), we observed an average effect size of 0.26 (range=0.12-0.45) for decreases in stigma, and 0.32 (range=0.20-0.49) for increases in self-reported resilience.

The TWM is a further adaptation of the R2MR program, but for general workplace settings. It incorporates videos and other training materials that are consistent with those settings. The program has been delivered to approximately 25,000 Canadians. Our outcome evaluations in eight diverse settings (N= 1,155) revealed an average effect size of 0.38 (range=0.15-0.51) for reduced stigma, and 0.50 (range=0.41-0.65) for increased resilience.

Qualitative outcomes for both programs suggest that participants seek help earlier and support others to seek help. Versions of both TWM and R2MR exist for frontline employees and managers. The manager version includes an additional module addressing issues related to employer responsibilities (e.g., workplace accommodations) and how to develop and maintain a mentally healthy workplace.

Based on the success of these programs, other variants are being developed and evaluated. For example, The Inquiring Mind is an adaptation for post-secondary students, and is currently under evaluation. Web-based booster sessions are being examined as a means to promote ongoing use of the program's knowledge and skills. A family package was created to assist family members of first responders who took the R2MR program to understand the program's insights. We have both a randomized trial of TWM underway, and an intended return on investment study of the R2MR program.

Despite the work to date, there remains much to learn about these types of programs and their effects. Interested readers can learn more by contacting the Mental Health Commission of Canada at mpietrus@mentalhealthcommission.ca.

Keith S. Dobson[1], Andrew Szeto[1], Stephanie Knaak[1,2], Terry Krupa[3], Bonnie Kirsh[4], Dorothy Luong[5], Robyn McLean[6], Micheal Pietrus[2]
[1]University of Calgary, Calgary, AB, Canada; [2]Mental Health Commission of Canada, Ottawa, ON, Canada; [3]Queen's University, Kingston, ON, Canada; [4]University of Toronto, Toronto, ON, Canada; [5]Toronto Rehabilitation Research Institute, Toronto, ON, Canada; [6]Tapestry Evaluation, Canada

1. McDaid D, King D, Parsonage M. In: Knapp M, McDaid D, Parsonage M (eds). Mental health promotion and prevention: the economic case. London: Department of Health, 2011:20-1.
2. Thorpe K, Chénier L. Building mentally healthy workplaces: perspectives of Canadian workers and front-line managers. Ottawa: Conference Board of Canada, 2011.
3. Szeto ACH, Dobson KS. Appl Preventive Psychol 2010;14:41-56.
4. Hanisch S, Twomey C, Szeto ACH et al. BMC Psychiatry 2016;16:1.
5. Krupa T, Kirsh B, Cockburn L et al. Work 2009;33:413-25.
6. Leppin AL, Bora PR, Tilburt JC et al. PLoS One 2014;9:e111420.
7. Kirsh B, Krupa T. Luong D. Work (in press).

DOI:10.1002/wps.20574

Personality traits and risk of suicide mortality: findings from a multi-cohort study in the general population

Suicide is a global public health concern. While many fewer deaths per year are attributed to suicide (800,000) than to chronic disease, estimates suggest that, for every completed suicide, an additional 30-40 attempts are made. This equates to more than 20 million attempted suicides worldwide each year[1].

While poor mental health[2], low cognition[3], social isolation[4] and socio-economic disadvantage[5] are related to suicide risk, the predictive role of other psychosocial characteristics such as personality type is uncertain. There is a circumstantial case for selected personality types being implicated in the occurrence of suicide. Observational studies, for example, suggest that low extraversion, high neuroticism, and low conscientiousness are associated with an increased prevalence of depressive symptoms[6], a determinant of suicide[2]. Lower conscientiousness has also been linked with an increased risk of heavy alcohol consumption[7], a further risk factor for suicide[8].

For the first time to our knowledge, we simultaneously related the five major personality components – extraversion, neuroticism, agreeableness, conscientiousness, and openness to experience – to suicide death by collating data from seven large cohort studies.

We pooled individual-participant (raw) data from five cohort studies with information on personality, key covariates, and suicide mortality: the UK Health and Lifestyle Survey (HALS), the original US National Health and Nutrition Examination Survey (NHANES 1), the US Health and Retirement Study (HRS), the Wisconsin Longitudinal Study Graduate Sample (WLSGS), and the Wisconsin Longitudinal Study Sibling Sample (WLSSS). We also incorporated results from remote, bespoke analyses of two further studies: the UK Biobank (UKBB) and the Miyagi Cohort Study (MCS) (study summaries available upon request).

Personality was assessed by using a range of questionnaires. In HALS (extraversion, neuroticism), NHANES 1 (extraversion, neuroticism, openness), UKBB (neuroticism), and MCS (extraversion, neuroticism), a selection of personality variables were captured, while in HRS, WLSG and WLSS all of the "Big Five" traits were measured (extraversion, neuroticism, agreeableness, conscientiousness, and openness to experience).

The covariates of education (primary, secondary, tertiary level), smoking (current, former/never), alcohol intake (light, heavy), and marital status (married/cohabiting, other) were self-reported and based on standard enquiries.

In all studies, death was ascertained from mortality records, with suicide denoted by any mention of the following events: suicide and self-inflicted poisoning by solid or liquid substances (E950-E959) and injury undetermined whether accidentally or purposely inflicted (E980-E989) according to ICD-9; and terrorism (U03.1 and U03.9), intentional self-harm (X60-X84), event of undetermined intent (Y10-Y34), sequelae of intentional self-harm, assault and events of undetermined intent (Y87), and sequelae of unspecified external cause (Y89.9) according to ICD-10.

A mean duration of mortality surveillance of 8.1 years in a total of 464,251 participants (3,782,553 person-years) gave rise to 270 suicide deaths. In the five studies for which we had individual-participant data, each of our covariates was related to completed suicide in the expected direction, although statistical significance at conventional levels was not always apparent: age (per decade increase: hazard ratio, HR=1.51, 95% CI: 1.19-1.92), gender (female vs. male: HR=0.37, 95% CI: 0.21-0.66), education (primary vs. secondary/tertiary: HR=2.40, 95% CI: 1.26-4.54), smoking (current vs. former/never: HR=1.89, 95% CI: 0.97-3.67), alcohol intake (heavy vs. light: HR=1.64, 95% CI: 0.36-7.44), and marital status (married/cohabiting vs. not: HR=0.59, 95% CI: 0.32-1.06).

In the main analyses in which the exposures of interest were the five personality types, adjusting for these covariates yielded the same results as those apparent after controlling for age and gender alone; we therefore present multiply-adjusted HRs only. Each one SD increment in neuroticism score was related to a 1.3-fold increase in suicide risk (HR=1.33, 95% CI: 1.18-1.50), while a one SD higher agreeableness score was associated with protection (HR=0.71, 95% CI: 0.53-0.97). After dropping data from UKBB (129 suicides) to examine if the largest study had skewed the results, we found that the risk associated with higher neuroticism was materially unchanged (HR=1.31, 95% CI: 1.11-1.55). We found no evidence that extraversion (HR=0.99, 95% CI: 0.84-1.17), conscientiousness (HR=0.98, 95% CI: 0.69-1.39), or openness to experience (HR=0.94, 95% CI: 0.69-1.29) were related to suicide rates in any of our analyses.

Each personality type was only weakly related to socio-economic status and health behaviours and, as a consequence, controlling for these factors did not have an impact on the personality-suicide relation. This implicates other explanations for the link between neuroticism and agreeableness on the one hand and suicide mortality on the other.

It seems likely that people regarded as being agreeable and less neurotic have a more extended or better established social network relative to individuals with less favourable scores on these traits. Social support, most frequently captured using marital status, is related to a lower risk of suicide[9]. Though the relationship of agreeableness and neuroticism with suicide was robust to the adjustment of marital status herein, we did not measure other potentially important characteristics of social integration – social network size, religious service attendance – that are known to predict suicide[9].

In conclusion, the characteristics of empathy and cooperation that are synonymous with agreeableness appear to be related to lower suicide rates, while people with a tendency towards impulsivity and hostility, typical of a neurotically-prone personality, experience higher risk. Our observation that standard demographic risk factors (gender, education, marital status) were related to suicide risk in the expected direction gives us a degree of confidence in these novel results for personality. Our findings suggest that attention should be paid to selected personality characteristics in suicide prevention.

G. David Batty[1], Catharine R. Gale[2], Fumiya Tanji[3], David Gunnell[4], Mika Kivimäki[1], Ichiro Tsuji[3], Markus Jokela[5]
[1]Department of Epidemiology and Public Health, University College, London, UK; [2]MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, UK; [3]Division of Epidemiology, Tohoku University School of Public Health, Sendai, Japan; [4]School of Social and Community Medicine, University of Bristol, Bristol, UK; [5]Department of Psychology and Logopedics, University of Helsinki, Helsinki, Finland

G.D. Batty is partially supported by the UK Medical Research Council (MR/P023444/1) and the US National Institute on Aging (1R56AG052519-01; 1R01 AG052519-01A1), M. Kivimäki by the UK Medical Research Council, NordForsk (the Nordic Research Programme on Health and Welfare) and the Academy of Finland (311472); and F. Tanji and I. Tsuji by the Ministry of Health, Labour and Welfare, Japan (H29-Junkankitou-Ippan-003). Part of this research was conducted using the UK Biobank Resource (application no. 10279).

1. World Health Organization. Preventing suicide: a global imperative. Geneva: World Health Organization, 2014.
2. Bell S, Russ TC, Kivimaki M et al. JAMA Psychiatry 2015;72:1254-6.
3. Batty GD, Whitley E, Deary IJ et al. BMJ 2010;340:c2506.
4. Tsai AC, Lucas M, Kawachi I. JAMA Psychiatry 2015;72:987-93.
5. Li Z, Page A, Martin G et al. Soc Sci Med 2011;72:608-16.
6. Hakulinen C, Elovainio M, Pulkki-Raback L et al. Depress Anxiety 2015;32:461-70.
7. Hakulinen C, Elovainio M, Batty GD et al. Drug Alcohol Depend 2015;151:110-4.
8. Jee SH, Kivimaki M, Kang HC et al. Eur Heart J 2011;32:2773-80.
9. Batty GD, Kivimaki M, Bell S et al. Transl Psychiatry 2018;8:22.

DOI:10.1002/wps.20575

Strengthening the scientific backbone of the WPA

As per WPA by-laws, the overarching aims of the Scientific Sections comprise the "collection, analysis, presentation, and dissemination of information concerning services, research, and training in the various fields of psychiatry and mental health and the advancement of scientific knowledge in these fields".

This is a lot, and may be considered quite ambitions by some. But, on the other hand, the Sections have become known as the "scientific backbone" of the WPA and, as such, they are meant to carry a lot of weight and provide a framework for our field, that is so diverse and deals with disorders that present with a combined lifetime prevalence of up to 40 percent. This diversity certainly is one of the reasons why the number of Sections has now grown to 72.

Given this importance for WPA, the post of Secretary for Scientific Sections comes with a great deal of responsibility and calls for a continuous thought process on how to further develop the Sections and keep them engaged.

In close collaboration with the President and the other members of WPA's Executive Committee, I will dedicate the next six years to the implementation of the following goals:

- *Improve and streamline communication between the Sections and facilitate research and publication projects.* With the Sections being the backbone for scientific advances within the WPA, the Association will develop an infrastructure that allows for brainstorming on research ideas, swift exchange of research proposals, comments, joint publications, etc.. Coordinated by the Secretariat[1,2], this will be achieved by acquiring novel and flexible information technology (IT) tools, such as web-based conferencing or sharepoints, and social media outlets. Such infrastructure will also be used to build a database on funding opportunities, with the idea to list country by country names and contact details of funding bodies/agencies and current requests for applications (RFAs) relevant to WPA's work. The emphasis should lie on RFAs aimed at international collaborations. The Sections should furthermore be encouraged to take up research projects that are in line with WPA's Action Plan[3] and/or the cities-RISE[4] initiative that WPA has partnered with. Ideally, this could be incentivized through a WPA program to offer seed funding based on a matched funding scheme.
- *Continue and expand the intersectional activities of WPA.* Over the past decade, the WPA has stepped up its efforts to foster intersectional activities at various levels[5,6]. This has ranged from arranging intersectional symposia or workshops at WPA meetings to major conferences like the joint meeting of the Section on Epidemiology and Public Health and the Section on Genetics in Psychiatry in Munich in 2016. Ideally, intersectional activities apply a cross-regional approach.
- *Leverage the Sections' experiences and resources to further WPA's activities for early career psychiatrists.* Investing into early career individuals is an investment into the future of WPA and psychiatry in general. This is why the WPA established an Early Career Psychiatrists (ECP) Section, following the successful example of the WPA Early Career Psychiatrists Council[7]. I will closely work with this Section to make sure that early career clinicians and researchers are sufficiently represented in the leadership and activities of all Sections. Leveraging the IT framework currently being built, the creation of a mentoring data base will be explored. Within the ECP program at WPA meetings, I consider implementing "science slams" where ECP individuals would present in a very brief form (5 min or less) current research projects/ideas/grant proposals to the audience. Senior representatives from the WPA Executive Committee, Board, Council, and Sections should be present and comment and give advice.
- *Promote gender equity at all levels of Sections and their activities.* Together with the Executive Committee and the Secretariat, I will implement measures to promote gender equity across all Sections, in particular as regards committee members and office bearers.
- *Establish cross-country peer networks of researchers to facilitate and share access to knowledge, resources, and strategies to publish successfully.* Building on the aforementioned resources and strategies, the Sections should be encouraged to establish efficient and long-lasting research networks spanning the globe. These could give rise to powerful consortia tackling important research questions and serve as catalysts for early career clinicians and scientists.
- *Establish truly authentic and compassionate relationships with organizations representing patients and caregivers.* In order to truly bring about change and to improve the lives of millions, WPA needs to bring together researchers/clinicians, patients, and family members/caregivers[8]. Sections are ideally suited to initiate such "trialogue"[9], focusing on specific topics of practical relevance to patients while at the same time establishing links to the research world.

These proposals will be tightly coordinated with the plans and work programs of the new Secretaries for Education and Scientific Publications, as well as the newly established Science Committee. Their implementation and integration in WPA's Action plan is currently discussed at the level of the WPA Executive Committee and Board.

The Secretary for Scientific Sections welcomes any further suggestions as to the future work and visibility of the Sections, WPA's scientific backbone.

Thomas G. Schulze
WPA Secretary for Scientific Sections

1. Kallivayalil RA. World Psychiatry 2017;16:114.
2. Kallivayalil RA. World Psychiatry 2017;16:330-1.
3. Herrman H. World Psychiatry 2017;16:329-30.
4. Herrman H. World Psychiatry 2018;17:236-7.
5. Javed A. World Psychiatry 2016;15:191-2.
6. Javed A. World Psychiatry 2017;16:222.
7. Fiorillo A, Brambhatt P, Elkholy H et al. World Psychiatry 2011;10:159.
8. Wallcraft J, Amering M, Freidin J et al. World Psychiatry 2011;10:229-36.
9. Amering M, Mikus M, Steffen S. Int Rev Psychiatry 2012;24:11-8.

DOI:10.1002/wps.20582

WPA Secretary for Education work plan

WPA vision is to advocate for "a world in which people live in conditions that promote mental health and have access to mental health treatment and care that meet appropriate professional and ethical standards, integrate public health principles and respect human rights".

To achieve the aim of ensuring equitable access to mental health care for people in different parts of the world, high-quality mental health education for professionals taking care of people with mental health problems must be provided. Driving sustainable changes in mental health services also requires collaborative partnerships with service users, carers, and other community stakeholders[1,2].

Therefore, my WPA education vision is "Education for all: no matter who you are, where you are, and how you are, with a strategic intent to improve mental health of all people around the Globe".

Service users and carers, as well as the general public, should have access to evidence-based mental health information. With the assistance of Member Societies and early career psychiatrists, the WPA website will provide hyperlinks to websites with high-quality mental health information[3]. As of today, we have identified websites with this information in seven languages, including English, Spanish, Portuguese, Chinese, Arabic, Hindi and Urdu.

Another initiative is to form a work group to establish a simple guideline for Member Societies and interested individuals to standardize the translation of important WPA documents into major languages.

A related project is to compile a list of recommended, evidence-based mental health apps for mobile devices through the support of the Section of e-Mental Health and other relevant WPA Scientific Sections. Besides, future WPA meetings will be expected to include free public forums for service users, carers and community stakeholders on topics of major concern to the host countries.

A global survey of training provisions involving WPA Member Societies was conducted in 2017[4]. The survey revealed that 30% of respondent countries provided less than 36 months of psychiatric training. A framework for psychiatric training has been developed and released on the WPA website[5]. The next step is to assist those countries to enhance their psychiatric training and education through different educational initiatives. Given that many psychiatrists with limited training are working in low- and middle-income countries (LMIC), an online psychiatric education programme will be an important educational tool[6].

The WPA has recently collaborated with the University of Melbourne to develop a completely online diploma programme on International Psychiatry[7]. The syllabus aims at enhancing knowledge and skills in psychiatric and risk assessment, essential psychopharmacology, basic psychotherapy, and social and cultural psychiatry. Potential sponsors are currently being identified to provide scholarships for these target participants, particularly those from Member Societies with less than 36 months of psychiatric training. Collaboration with other regional psychiatric associations is also underway to promote this online programme.

Apart from online education, face-to-face experiential training remains an important mode of education. A network of volunteer psychiatrists with different psychiatric expertise is now being developed in collaboration with major national psychiatric associations. The WPA will serve as a platform to coordinate continuous training and supervision by experts addressing the identified needs in the recipient countries, and to find potential sponsors to support travel and accommodation costs for the visiting scholars. Such travel fellowships have been found to be valuable in other medical specialties[8]. This programme is now under development in Asia in collaboration with the Asian Federation of Psychiatric Associations (AFPA).

Apart from supporting psychiatrists who have completed psychiatric training, the WPA is determined to support and develop talents for mental health at the start of medical career. In collaboration with the International Federation of Medical Students, a global survey on the psychiatric curriculum in basic medical education has just been completed. The survey results will inform the WPA on how undergraduate psychiatric curriculum can be enriched in medical school education[9].

The WPA also recognizes early career psychiatrists as our next generation in promoting the betterment of global mental health. Therefore, every work group on education will have at least one early career psychiatrist being involved as a regular member. Apart from having their voices and ideas being heard in these work groups, early career psychiatrists may also benefit from learning the leadership styles of senior members. Last but not least, future WPA congresses will endeavor to identify possible means to support medical students and early career psychiatrists to participate in educational symposia and academic exchanges.

The educational needs of other professionals working in mental health will also be addressed. The aforementioned WPA online diploma also targets primary care doctors working with patients with mental health problems. Separate scholarships will also be identified for this target group of potential applicants. Fur-

thermore, joint educational activities will be developed in collaboration with the World Organization of Family Doctors (WONCA).

Given the high prevalence of common mental disorders in the community across the world and the effectiveness of cognitive-behavioral therapy (CBT) for these disorders, health professionals taking care of patients with these disorders in the primary care sector should be equipped with the relevant basic CBT skills[10].

Existing evidence-based CBT manuals targeted for health professionals at various levels of clinical experience in different languages will be identified. CBT therapists and supervisors from the relevant WPA Scientific Sections will be recruited to disseminate these CBT skills during their travel fellowships to LMIC and WPA courses organized in WPA meetings.

Roger Man Kin Ng
WPA Secretary for Education

1. Herrman H. World Psychiatry 2017;16:329-30.
2. Bhugra D, Tasman A, Pathare S et al. Lancet Psychiatry 2017;4:775-818.
3. Kallivayalil RA. World Psychiatry 2017;16:114.
4. Ng RMK, Hermans M, Belfort E et al. Int Rev Psychiatry (in press).
5. Belfort E, Lopez-Ibor MI, Hermans M et al. WPA recommendations: principles and priorities for a framework for training psychiatrists. www.wpanet.org.
6. Saraceno B, van Ommeren M, Batniji R et al. Lancet 2007;370:1164-74.
7. Bhugra D. World Psychiatry 2017;16:221-2.
8. Nchinda T. Soc Sci Med 2002;54:1699-711.
9. Shields G, Ng RMK, Ventriglio A et al. World Psychiatry 2017;16:113-4.
10. Bower P, Gilbody S. BMJ 2005;330:839-42.

DOI:10.1002/wps.20576

Action Plan of the WPA Secretary for Publications

What makes the exceptional value of the WPA is the unique opportunity it offers to accommodate the diversity of psychiatry, not only in its geographical and cultural aspects, but also in its theoretical and practical dimensions. This diversity makes, indeed, the richness of psychiatry as long as it is integrated under a common umbrella like the one provided by the WPA. This integrative mission is explicitly highlighted in the WPA Action Plan 2017-2020[1].

Within this global framework, publications are situated between science (or knowledge) and education (or training), in reference to evidence and value based good practices[2]. It is, therefore, mainly a mediating stage between the input of knowledge and the output of education. The main aim of publications is then to be integrated in this general process to serve WPA goals for the benefit of psychiatry, mental health and their users or patients.

In this general perspective, publications have, however, several specific tasks. Among them, the dissemination of knowledge to improve psychiatric science and practice and the promotion of the visibility, the funding and the academic recognition of Psychiatry, of WPA and of WPA components, in particular Member Societies and the psychiatrists they associate.

For that purpose, we consider crucial to take into account the lack of resources of low- and middle-income countries (LAMIC) and the glass-ceiling effect professionals often face in less scientifically favored contexts, when they want to get published. Linguistic, cultural and educational reasons account for many of these difficulties, but they are far from being the only or the most important obstacles met by these colleagues. Even more crucial is often the lack of theoretical and practical incentives in the less scientifically informed and methodologically trained contexts.

To address this issue, it is crucial to improve the quality of psychiatry worldwide. To reach this objective, we will have to keep in mind that the WPA is not a scientific organization as any other: beside knowledge dissemination and evidence-based training, the WPA should aim at increasing its prestige and its scientific recognition, so that it can use them to reduce the above contextual limitations and promote the production emerging from less favored zones of the world.

In this action, the stake of our work in the WPA will be to encompass, on one hand, the state of the art in the various domains involved in psychiatry, taking into account the huge differences of scientific levels according to each of these domains, and, on the other hand, the reality of psychiatric practices in the various contexts in which psychiatry and psychiatrists are struggling to help as efficiently as possible real patients and carers they meet in their "natural" settings.

Multiple actions can be proposed in the publication domain to reach these, potentially contradictory, overall goals. Building on the extraordinary success of the WPA official scientific journal, *World Psychiatry* (it recently reached the impressive impact factor of 30, under the direction of M. Maj, its Editor), our publication project will try to renew WPA efforts to increase the number, the scientific quality and the dissemination of the products of psychiatric knowledge and experience.

For that purpose, one of our proposals will be to commission and contribute to produce WPA or WPA sponsored books on relevant topics, with recognized publishers and editors, increasing the visibility of these productions through the WPA website and WPA meetings.

As an international organization, we will also do our best to support the translations of *World Psychiatry* in various languages (Russian, French, Spanish, Portuguese, Arabic, Chinese). The same efforts will be made to produce or trans-

late anthologies of important classical papers from various psychiatric traditions, resuming a successful book series that has been interrupted in the past few years.

In line with the focus of the previous committee for scientific publications, under the leadership of M. Riba[3–5], we will also develop, as much as possible, adequate assistance to Member Societies to get their national or regional journals properly indexed. In that perspective, we will look for adequate ways to provide psychiatrists worldwide with assistance and support to increase their ability to publish scientific papers in English, such as bilingual online journals (French-English, Spanish-English, Portuguese-English), methodological and writing assistance, online courses on "how to publish in indexed journals".

Beside their scientific qualities, these productions will aim to cover, from science to practice, topics likely to benefit from multifocal and comprehensive approaches, in close interaction with allied disciplines. Here again, the past productions of WPA components and the feedback from the Member Societies will guide and inspire us for our future works.

In all these projects, WPA diversity will be seen as a strength. In line with the Action Plan, the tasks related to publications will look for the benefit they can get from this diversity, articulating our publication projects with those assigned to the other members of the Executive Committee, particularly the Secretaries for Education and for Scientific Sections, in connection with the new Standing Committee for Science.

Michel Botbol
WPA Secretary for Scientific Publications

1. Herrman H. World Psychiatry 2017;16:329-30.
2. Fulford KWM. J Eval Clin Pract 2001;17:976-87.
3. Riba M. World Psychiatry 2016;15:88.
4. Riba M. World Psychiatry 2017;16:114-5.
5. Szabo C, Castaldelli-Maia JM, Chandra P et al. Global Psychiatry 2018;1:1-2.

DOI:10.1002/wps.20578

Acknowledgement
This publication has been partially supported by an unrestricted
educational grant from Janssen-Cilag SpA,
which is hereby gratefully acknowledged.

© 2018 by WPA

Notice No responsibility is assumed by the Publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made.