# World Psychiatry

OFFICIAL JOURNAL OF THE WORLD PSYCHIATRIC ASSOCIATION (WPA)

*Volume 14, Number 3* *October 2015*





WILEY Blackwell **IMPACT FACTOR: 14.225** ISSN 1723-8617

http://www.arabpsynet.com/Journals/WJ/WP-October2015.Eng.pdf

## 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 135, spanning 118 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 69 scientific sections, aimed to disseminate information and promote collaborative work in specific domains of psychiatry. It has produced several educational programmes and series of books. It has developed ethical guidelines for psychiatric practice, including the Madrid Declaration (1996).

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



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Research Reports containing unpublished data are welcome for submission to the journal.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



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).**

# Antidepressant-placebo differences: is the glass half full or half empty?

**Dan J. Stein**

Department of Psychiatry and Mental Health, MRC Unit on Anxiety and Stress Disorders, Groote Schuur Hospital, University of Cape Town, Cape Town, South Africa

Khan and Brown's work (1) on public domain data from the archives of the U.S. Food and Drug Administration (FDA) emphasizes that drug-placebo differences in recent antidepressant trials are smaller than in early investigations, and draws a number of conclusions about the nature of clinical depression and optimal trial design. Several relevant issues in regards to this and related work on drug-placebo differences in psychiatry research may be worth emphasizing.

First, it is important for psychiatry to steer an even course between the polar positions of scientism and scepticism. Mental disorders such as depression are not analogous to squares; they are not natural kinds that can simply be defined using necessary and sufficient criteria (2,3; see also 4 in this issue of the journal). At the same time, mental disorders are not social constructions that are solely determined by socio-political considerations, and that therefore differ wholly from time to time and place to place. While efforts such as the Research Domain Criteria (RDoC), which aim to ground psychiatric constructs in translational neuroscience, may help lead to advances in psychiatric nosology and clinical trials over the long term (5), iterative improvements of diagnostic criteria and guidelines will anchor clinical practice and interventional research for now and for the foreseeable future (6).

Second, it is relevant to note that psychiatry's approach to mental disorders has a great deal in common with the rest of medicine's approach to physical disorders. The rest of medicine accepts that many conditions are best conceptualized as syndromes (7). Arguably, psychiatry has led the way in terms of providing valid, reliable and useful approaches to the diagnosis of medical conditions where simple biomarkers are not available or helpful (8). While the introduction of paradigm-shifting innovations and personalized medicine initiatives in clinical trials methodology for interventional research in psychiatry may occur over time, in the shorter term we can more certainly expect iterative improvements (perhaps including ideas emphasized by Khan and Brown, such as limiting the number of sites and treatment arms) to FDA and European Medicines Agency (EMA) guidelines for undertaking pivotal clinical trials.

Third, it is crucial to emphasize that a broad range of causal mechanisms are likely involved in the pathogenesis and treatment of mental disorders such as depression (9). Not surprisingly, any specific pharmacological agent, acting on only a limited subset of such mechanisms, may have a relatively low effect size, particularly when inclusion criteria lead to investigation of a heterogeneous phenotype. Antidepressant effect sizes may, however, be higher for some narrower phenotypes (e.g., melancholic depression) that are often excluded from clinical trials (e.g., due to characteristic suicidal ideation). Such effect sizes may also differ in the U.S. and Europe, for a range of reasons (10). Furthermore, effect sizes for psychiatric treatments are as least as high as those in the rest of medicine (11,12).

Steering a course between scientism and scepticism also means finding a balance between over-optimism and over-pessimism with regards to psychiatry in general and antidepressants in particular. We need to acknowledge the enormous advances made in psychopharmacological research over the past several decades, while also emphasizing that there remain significant needs and opportunities for better understanding the relevant proximal and distal psychobiology of mental disorders, for better implementing and scaling-up available treatments, and for more efficacious and effective drugs (13). The level of liquid in our glass is arguably at 50%, and we need to deal with this reality accordingly.

Steering a course between scientism and scepticism may also impact our perspective on the placebo response. The reliance of modern clinical research on randomized placebo-controlled trials has led to an ever growing database demonstrating that placebo is a remarkably powerful intervention for a range of psychiatric and medical conditions, including milder cases of depression (14,15). This should not be cause for embarrassment or despair for psychiatry, but rather an impetus for research on the underlying psychobiology of the placebo response, and on how to better harness such effects in clinical practice (16). Further advances in this direction would arguably help ensure that our glass is more than half-full.

## References

1. Khan A, Brown WA. Antidepressants versus placebo in major depression: an overview. World Psychiatry 2015;14:294-300.
2. Stein DJ. Psychopharmacology and natural kinds: a conceptual framework. In: Kincaid H, Sullivan JA (eds). Classifying psychopathology: mental kinds and natural kinds. Cambridge: MIT Press, 2014:65-74.
3. de Jonge P, Wardenaar KJ, Wichers M. What kind of thing is depression? Epidemiol Psychiatr Sci 2015;24:312-4.
4. Zachar P. Psychiatric disorders: natural kinds made by the world or practical kinds made by us? World Psychiatry 2015;14:288-90.
5. Cuthbert BM. The RDoC framework: facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. World Psychiatry 2014;13:28-35.

6. Stein DJ. An integrative approach to psychiatric diagnosis and research. World Psychiatry 2014;13:51-3.
7. Nesse RM, Stein DJ. Towards a genuinely medical model for psychiatric nosology. BMC Med 2012;10:5.
8. Olesen J. The international classification of headache disorders. Headache 2008;48:691-3.
9. Kendler KS. The dappled nature of causes of psychiatric illness: replacing the organic-functional/hardware-software dichotomy with empirically based pluralism. Mol Psychiatry 2012;17:377-88.
10. Stein DJ, Baldwin DS, Dolberg OT et al. Which factors predict placebo response in anxiety disorders and major depression? An analysis of placebo-controlled studies of escitalopram. J Clin Psychiatry 2006;67:1741-6.
11. Leucht S, Hierl S, Kissling W et al. Putting the efficacy of psychiatric and general medicine medication into perspective: review of meta-analyses. Br J Psychiatry 2012;200:97-106.
12. Guo JB, Zhu Y, Chen BL et al. Surgical versus non-surgical treatment for vertebral compression fracture with osteopenia: a systematic review and meta-analysis. PLoS One 2015;10:e0127145.
13. Stein DJ. The philosophy of psychopharmacology: happy pills, smart pills, pep pills. Cambridge: Cambridge University Press, 2008.
14. Klein DF, Ross DC. National Institute of Mental Health Treatment of Depression Collaborative Research Program General Effectiveness Report. Neuropsychopharmacology 1993;8:241-51.
15. Fountoulakis KN, Moller HJ. Efficacy of antidepressants: a re-analysis and re-interpretation of the Kirsch data. Int J Neuropsychopharmacol 2011;14:405-12.
16. Stein DJ, Mayberg H. Placebo: the best pill of all. CNS Spectr 2005;10:440-2.

DOI 10.1002/wps.20263

# Compulsion and "coercion" in mental health care

GEORGE SZMUKLER

Institute of Psychiatry, King's College London, London, UK

"Compulsion" in mental health care is a reasonably straightforward notion: the use of force, one hopes always governed by law, to make a person accept treatment that has been refused. The term "coercion" is usually taken to include "compulsion", but encompasses a broader range of practices. Sometimes it is used almost synonymously with treatment pressures, including "interpersonal leverage", and even "persuasion". I believe that, if we are to take our thinking – including research ideas – forward, we need a more precise understanding of "coercion" (1).

Most accepted is the definition proposed by Wertheimer (2), who includes "threats" as coercive. A "threat" is a conditional proposal ("if …, then …") that, if rejected by the person, leaves him/her worse off according to a "moral baseline" ("if you refuse the medication, you will be detained in hospital"). The moral baseline is that one is normally entitled not to be deprived of one's liberty. That is not to say that it can never be justified, but a special case needs to be made.

Wertheimer contrasts a "threat" with an "offer" (or inducement). An example: "if you take the prescribed medication, you will receive a payment" (3). Here a rejection of the proposal does not leave the person worse off, as he/she is not entitled to a payment. Nevertheless, such an inducement can be problematic, for example, by undermining the patient's sense of agency or through corrupting the value of the treatment (4).

I take coercion to cover both compulsion and threats. A further consideration is the difference between "objective" coercion and "perceived" coercion. The former follows the definitions given above. The latter is a person's perception of threat, even where no threat may be intended. A disquieting problem facing psychiatry is its "coercive shadow", the fear many patients have that non-compliance may lead to the use of compulsion. Patients may agree to treatment, including admission to hospital, "voluntarily" to avoid the humiliation and stigma of a compulsory order. Research shows this is very common, even though in most places threats are regarded as ethically unacceptable.

There has been little discussion of this topic, but I suggest that, given the uncertainty of whether a proposition is a threat or not, we might look at ways of "regulating" threats: for example, making them transparent – their only being made in "good faith" (that is, the threatener really means it)– and clarifying practice in codes of practice or professional ethics.

Apart from the problem of definition, research on coercion is dogged by the problem of context. There is a large variation in the rates of compulsory admission to hospital, both between countries (even without outliers, 3- to 4-fold) (5) and within countries (6). The use of seclusion, restraint and forced medication may vary hugely (7), even 10-fold from hospital to hospital in the same country (8).

The sources of variation can be attributed to different service configurations, different mental health laws, different social policies (for example, the rate and extent of bed reductions; the degree of emphasis on risk and public protection), and, crucially, culture. In some countries physical restraints are regarded as unacceptable and are rarely or not at all used; in others chemical restraints are thus regarded. Furthermore, the use of compulsion may change significantly over time according to changes in policy and practice. In England, there has been a doubling over the past 20 years (9).

Even if well-designed self-report or interview measures are used, ones that are interpreted similarly from place to place, the results of any one study on coercion will likely have limited generalizability. Thus, perhaps more than in any other field of health services research, international collaborative studies are needed. The EUNOMIA programme is a good example (10). Further points to be considered are where (in the community or in the hospital) and when (before discharge or after) the assessment is made, and by whom. Service user researchers may get different responses from conventional researchers. Variations here may lead to different results.

Research ethics committees often struggle with research in this area. It is sensitive, and there may be concerns about consent and the "voluntariness" of participation, which may lead to bias due to the exclusion of important subgroups of patients. With careful thought these problems can be overcome.

A huge challenge to involuntary treatment comes from the United Nations (UN) Convention on the Rights of Persons with Disabilities (11). By April 2015, 159 states were signatories. The elimination of discrimination by ensuring that rights may be enjoyed "on an equal basis with others" is a fundamental aim. Persons with serious mental illness are considered by the UN Committee for the Rights of Persons with Disabilities, the authoritative body set up by the UN to interpret and monitor compliance with the Convention, to fall under the characterization of "disability" (sometimes referred to as "psychosocial" disabilities).

Article 14 states that "the existence of a disability shall in no case justify a deprivation of liberty", meaning that "mental disorder" or "mental illness", even if it represents only one of a number of criteria for involuntary detention in a mental health law, renders such a law non-compliant with the

Convention. Article 12 recognizes that all persons enjoy "legal capacity" in all aspects of life on an "equal basis with others". The Committee, in a recent "General Comment" on this article, states that "substitute decision-making", where someone decides for the person with a disability (as opposed to "supported decision-making"), is non-compliant (12). Over twenty "concluding observations" made thus far by the Committee, following its monitoring of reports on progress from States in implementing the Convention, conclude that they must "take action to develop laws and policies to replace regimes of substitute decision-making by supported decision-making, which respects the person's autonomy, will and preferences" (13).

It is hard to imagine a society in which it would be seen as right that persons who are seriously incapable of exercising autonomy or expressing their will and preferences would be allowed to act so as to incur grave harms, including death. Where the UN Convention is valuable, apart from its clear articulation of a host of other rights for people with disabilities, is in making us scrutinize in depth our justifications for coercive interventions. Together with colleagues, we (14,15) have argued that conventional mental health law discriminates against persons with a mental disorder since it does not respect such persons' autonomy (or rights to self-determination or self-governance) in the same way as in the rest of medicine. In the latter, considerations such as impaired "decision-making capacity" and treatment needing to be in the person's "best interests" justify the over-riding of a treatment refusal. In the mental health field, a diagnosis of a "mental disorder" – usually vaguely defined – and the presence of some kind of risk to self or others comprise the criteria. The rules are entirely different.

Furthermore, the "protection of others" permits the preventive detention of persons with mental disorder on the basis of the risk they are deemed to pose before they have actually committed an offence. This group is unique in this regard. The many more persons without a mental disorder who are equally or more risky are not liable to such detention. In this regard, non-discrimination means either having generic "dangerousness" legislation equally applicable to all who present an unacceptable level of risk, or no preventive detention for anyone.

Thus we (14) have argued for a non-discriminatory, generic, "fusion law" that would apply to all persons, whatever their diagnosis – medical, surgical or psychiatric – and whatever the setting. Involuntary interventions would only be justified for those who lack decision-making capability (unable to understand and retain the relevant information, to appreciate its pertinence to their situation, to reason with it in the light of what is important to themselves, and to evidence a choice) and only where it would be in their "best interests" (essentially what that person would have chosen if he/she had retained capacity in the current circumstances). Advance statements or directives (see 16 in this issue of the journal) could play an important role here. Northern Ireland is currently proposing to legislate along these principles.

Bach and Kerzner (17), attentive to the "legal capacity" standard of the UN Convention, have proposed three levels of "decision-making capability". The first is "legally independent", having full decision-making ability as outlined above. The next level is where varying degrees of support – informal or formal – would be required to assist the person to arrive at a decision based on the person's will and preferences. The third level, "facilitated" decision-making, would represent a last resort and would be restricted to instances where it is impossible to arrive at a settled understanding or interpretation of the person's will and preferences and where decisions are made by another person. However, as part of this action, the facilitator would continue to work with the person to establish with time what are the person's will and preferences.

An approach that combines both of the above could be developed. "Decision-making capacity" and "best interests", both terms criticized by the UN Convention Committee, can be helpfully reconceptualized in terms of the person's "real" or "authentic" will and preferences (15).

The huge variation in rates of involuntary treatment suggests that in many countries there is considerable scope for a reduction. From an ethical point of view, a randomized controlled trial (RCT) of involuntary inpatient treatment is hardly possible. We accept that it can be morally justified, indeed obligatory, to treat people involuntarily under certain circumstances. However, there have been three RCTs of involuntary outpatient treatment (or community treatment orders). While each has its flaws, none has shown a clearly significant improvement in any of a range of outcomes (18). I have argued that an alternative approach, consistent with the "fusion" proposal, would conceive of community treatment orders in a different way and would look for different, individual, patient-preferred, outcomes (19).

There is reasonably consistent evidence, even when involuntary treatment has been authorized, that "perceived coercion" is less when the relationship between patient and clinicians is good, and when patients believe their "voice" has been heard (20).

A promising means of reducing the need for coercion at times of crisis, especially a relapse of illness, might be an advance directive, or the less legally formal "joint crisis plan" (see 16 and 21 in this issue of the journal). There is evidence in the case of the former that, when helped by a facilitator in drawing up the directive, in the short term at least, patients may experience their care as better (22). Joint crisis plans have been more extensively studied. An earlier, sizeable, RCT found a significant reduction in involuntary admissions when a joint crisis plan had been agreed between patient and clinical team. However, a much larger RCT involving 569 patients found no difference in involuntary admissions or any other outcome (23). A lack of treatment fidelity or clinician "buy-in", a problem for any multicentre complex intervention, may have been responsible. A joint crisis plan pilot study for patients who self-harm also found no benefit (24). However, 85% of patients

who had a joint crisis plan said they would recommend it to others. Perhaps this reflects the respect accorded to the patient's "voice" in the joint crisis plan negotiation.

In conclusion, there are considerable conceptual and practical difficulties in understanding and researching compulsion and coercion. Nevertheless, it is hugely important to our patients and, indeed, for the status of psychiatry that we do all that is possible to reduce recourse to these measures to a minimum.

## References


1. Szmukler G, Appelbaum P. Treatment pressures, leverage, coercion and compulsion in mental health care. J Ment Health 2008; 17:233-44.
2. Wertheimer A. A philosophical examination of coercion for mental health issues: some basic distinctions: analysis and justification. Behav Sci Law 1993;11:239-58.
3. Priebe S, Yeeles K, Bremner S et al. Effectiveness of financial incentives to improve adherence to maintenance treatment with antipsychotics: cluster randomised controlled trial. BMJ 2013; 347:f5847.
4. Sandel MJ. What money can't buy: the moral limits of markets. London: Penguin, 2012.
5. Salize HJ, Dressing H. Epidemiology of involuntary placement of mentally ill people across the European Union. Br J Psychiatry 2004;184:163-8.
6. Hoyer G. Involuntary hospitalization in contemporary mental health care. Some (still) unanswered questions. J Ment Health 2008;17:281-92.
7. Bak J, Aggernaes H. Coercion within Danish psychiatry compared with 10 other European countries. Nord J Psychiatry 2012; 66:297-302.
8. Husum TL, Bjorngaard JH, Finset A et al. A cross-sectional prospective study of seclusion, restraint and involuntary medication in acute psychiatric wards: patient, staff and ward characteristics. BMC Health Serv Res 2010;10:89.
9. Health and Social Care Information Centre. Inpatients formally detained in hospitals under the Mental Health Act 1983, and patients subject to supervised community treatment. www.hscic.gov.uk.
10. Raboch J, Kališová L, Nawka A et al. Use of coercive measures during involuntary hospitalization: findings from ten European countries. Psychiatr Serv 2010;61:1012-7.
11. United Nations. Convention on the rights of persons with disabilities. New York: United Nations, 2006.
12. UN Committee on Convention on the Rights of Persons with Disabilities. General Comment on Article 12: Equal recognition before the law, 2014. www.ohchr.org.
13. UN Committee on Convention on the Rights of Persons with Disabilities. Consideration of State reports. tbinternet.ohchr.org.
14. Dawson J, Szmukler G. Fusion of mental health and incapacity legislation. Br J Psychiatry 2006;188:504-9.
15. Szmukler G, Daw R, Callard F. Mental health law and the UN Convention on the Rights of Persons with Disabilities. Int J Law Psychiatry 2014;37:245-52.
16. Zelle H, Kemp K, Bonnie RJ. Advance directives in mental health care: evidence, challenges and promise. World Psychiatry 2015; 14:278-80.
17. Bach M, Kerzner L. A new paradigm for protecting autonomy and the right to legal capacity. Toronto: Law Commission of Ontario, 2010.
18. Burns T, Rugkasa J, Molodynski A et al. Community treatment orders for patients with psychosis (OCTET): a randomised controlled trial. Lancet 2013;381:1627-33.
19. Szmukler G. Is there a place for community treatment orders after the OCTET study? Acta Psychiatr Scand 2015;131:330-2.
20. Newton-Howes G, Mullen R. Coercion in psychiatric care: systematic review of correlates and themes. Psychiatr Serv 2011;62: 465-70.
21. Henderson C, Farrelly S, Moran P et al. Joint crisis planning in mental health care: the challenge of implementation in randomized trials and in routine care. World Psychiatry 2015;14:281-3.
22. Swanson JW, Swartz MS, Elbogen EB et al. Facilitated psychiatric advance directives: a randomized trial of an intervention to foster advance treatment planning among persons with severe mental illness. Am J Psychiatry 2006;163:1943-51.
23. Thornicroft G, Farrelly S, Szmukler G et al. Clinical outcomes of Joint Crisis Plans to reduce compulsory treatment for people with psychosis: a randomised controlled trial. Lancet 2013;381:1634-41.
24. Borschmann R, Barrett B, Hellier JM et al. Joint crisis plans for people with borderline personality disorder: feasibility and outcomes in a randomised controlled trial. Br J Psychiatry 2013;202:357-64.


DOI 10.1002/wps.20264

SPECIAL ARTICLE

# Developmental psychopathology: recent advances and future challenges

SETH D. POLLAK

Department of Psychology, University of Wisconsin-Madison, Madison, WI 53706, USA

*The integrative field of developmental psychopathology is having a huge impact on our understanding of human health and behavior. In this paper, I use the example of children's early stress exposure to illustrate how developmental psychopathologists now tend to deemphasize diagnostic categories and, instead, emphasize the social and biological contexts, events and circumstances that have created opportunities for maladaptive responses and health problems in youth. This example shows that developmental psychopathology is increasing understanding of how children develop the abilities that allow them to cope effectively with challenges and what leads to failures in development of these abilities. Integrating research about the neurobiology of learning may prove to be a powerful future direction to understand how the environment regulates behavior. Learning processes become increasingly intricate and fine-tuned as relevant neuroanatomical systems develop, and as the range, complexity and amount of environmental information increases for the developing child. A focus on these processes allows psychopathologists to formulate questions about which neural mechanisms children use to process information, how these mechanisms are themselves shaped by social context, why adverse social environments confer risk for children, and, perhaps, what sorts of neutrally informed interventions might remediate the deficits in self-regulation that underlie common psychopathologies.*

**Key words:** Developmental psychopathology, child stress, child psychiatry, child maltreatment, depression, child development, attention, learning

***(World Psychiatry 2015;14:262–269)***

Developmental psychopathology, as an integrative field of study and scientific approach, is just a few decades old. But it has already had a huge impact on our understanding of human health and behavior (1). The popularity and prominence of the approach has risen quite rapidly and has fostered connections between many fields of study, encompassing cross-cultural perspectives as well as new methods from the neurosciences. In this paper, I will illustrate some of the ways in which the developmental psychopathology perspective has shown utility for the field of psychiatry, highlighting recent trends and future challenges.

Psychiatrists have long been concerned with individual differences in how youth with behavioral problems manifest clinical symptoms, as well as differences between individuals in their responsiveness to treatments. But, at the time developmental psychopathology began to emerge, the field of psychiatry was very focused on diagnostic issues. Child psychiatry was concerned with topics that included formulations of taxonomies for mental disorders, the relations between those categories, and the bases for determining if an individual met criteria for a particular diagnosis. In more recent years, psychiatric research has begun to consider "biomarkers" associated with various forms of pathology as a way to reconcile diagnostic taxonomies with biological systems. But, despite the many publications correlating a diffuse array of biomarkers with various forms of psychopathology, no individual biomarker has yet emerged as a discrete entity that has been shown to account for a sufficient proportion of variance in behavior, or that is sensitive or specific to behavioral disorders.

It is in this regard that the developmental psychopathology approach holds promise. In isolation, markers (e.g., functional brain activities, hormonal assays, genetic markers, or cognitive test scores) are merely *correlates* of behavior problems. In contrast, developmental approaches seek to understand the *processes* by which these components have emerged and become integrated across biological, psychological and social contexts over the individual's life course. This approach leads to the dissolution of the distinction between mental and physical disorders, and can especially be seen with regard to understanding children's responses to stress. The approach has also manifested itself with a renewed emphasis among researchers on the interactions between persons and their environments. I will illustrate some of these principles using the phenomenon of child maltreatment as an example of severe early life stress exposure.

Child maltreatment predicts both unfavorable mental health outcomes as well as poor responsiveness to mental health treatment (2). Maltreated children are at risk for developing externalizing behavioral problems characterized by reactive aggression. For example, these children exhibited greater negative affect in response to an interpersonal stressor, which was subsequently associated with more aggressive behavior towards their peers. This relationship was mediated by children's allocation of attention to angry faces as measured by brain event-related potentials (3). These data suggest that physical maltreatment leads to inappropriate regulation of both negative affect and aggression, which likely places maltreated children at increased risk for the development and maintenance of externalizing behavior disorders.

Yet, child maltreatment is also associated with heightened risk for mood disorders, though not all individuals who experience maltreatment develop depression or anxiety. One clue about the ways in which the early experience

of maltreatment may lead to depression can be found in observations of maltreated children's attention bias for emotional cues (4). A recent study reported that maltreated children showed attentional biases to depression-relevant cues in certain conditions: first, after they had experienced a sad emotional state, and second, if they tended to have high levels of trait – or stable – cognitive patterns of rumination (5). These patterns may identify which maltreated children are at heightened risk for depression.

The phenomenon of rumination – passively and repetitively dwelling on and questioning negative feelings in response to distress – is a known risk factor for the development of psychopathology, especially depression (6). Recent research in a community sample of 9 to 14 year olds showed that it was common for youth to focus on an interpersonal stressor for a brief period of time after experiencing it; yet about 10% of the youth continued to ruminate for a long period of time after the stressor ended (7). Although most participants were able to disengage from this type of ruminative thinking, those individuals who continued to ruminate showed attentional biases away from positive stimuli (7). Thus, these children actively avoided environmental cues that might have helped them regain a positive mood state and recover from the stressful event. Consistent with this view, rumination in adolescents is associated with difficulty inhibiting negative information when switching from processing of negative to positive information (8). The ruminative process is difficult to stop once it has begun. But relatively straightforward interventions, such as brief periods of distraction or mindfulness, appear to be helpful in getting children out of ruminative states (9).

Of concern, however, is not just internalizing and externalizing psychopathology, but also sub-clinical problems that decrease children's quality of life, such as emotion regulatory difficulties, problems with social competence, factors that interfere with optimal school performance, as well as factors that affect physical health. Attention to these issues reflects the increasingly broad focus on the whole child, rather than psychiatric diagnoses in particular, within developmental psychopathology.

## A FOCUS ON DEVELOPMENTAL PROCESSES

Some developmental psychopathologists continue to professionally identify according to the diagnostic category they study (labeling themselves, for example, as "depression researchers" or "autism researchers"). But one noteworthy trend in the field is that, increasingly, younger generations of scholars are identifying themselves in terms of etiological and developmental mechanisms rather than discrete disorders. For example, these scientists may think of themselves as "stress researchers", "affective neuroscientists" or scholars of "mind-body interactions". In my view, such a change is not trivial and reflects a critical shift in emphasis among psychopathologists to link brain-behavior relationships to dimensions of maladaptive behaviors (see 10-16). While researchers continue to study issues such as dysregulation of mood, they increasingly construe their topics as perhaps broader than "anxiety" or "depression". And this reflects a major trend in the field to focus on maladaptive processes of change.

One reason for this change in emphasis is that it is now apparent that development is best characterized by probabilistic pathways rather than by linear causality. There has been no evidence that early adversity leads ineluctably to pathology. Rather, social and biological challenges initiate processes that may more likely lead to pathology if that maladaptive pathway continues to be supported. In this regard, developmental psychopathologists have become less focused on discrete causes of disorders. Instead, we are attempting to understand what places a child on one developmental pathway versus another, what constrains the individual's ability to alter these pathways, and during which developmental time periods, or circumstances, opportunities for change might be greatest.

## DISSOLVING DISTINCTIONS BETWEEN MENTAL AND PHYSICAL HEALTH

An unintended effect of focusing on processes versus discrete disorders has been a blurring, with developmental psychopathology, of traditional disciplinary boundaries. Methods and concepts from fields such as psychiatry, psychology and pediatrics have come into greater contact with those from internal medicine, immunology, endocrinology, epidemiology/population health, and genetics. For example, research on children's responses to trauma and stress still includes issues such as anxious and aggressive symptoms, but also includes foci such as sleep, physical growth and bone density, allergy/asthma, infectious disease, and cancer vulnerability (17-22). In other words, mental health problems are being understood and linked with indicators of physical health, eroding the distinction between mental and physical ailments.

It has now become apparent that early life stress can compromise development, with higher amounts of adversity linked to a diffuse array of developmental problems. There is evidence that an important facet of risk for mental illness can be understood as altered neural processing of social stimuli, which impairs regulatory processes. This research both informs our understanding of the emergence of health problems in children and adults, and also sheds light on principles of normative development. In this manner, we increase understanding of how is it that children's social experiences subsequently shape their thoughts, feelings, biology and behavior.

One lens for understanding the principle of development is the rubric of learning. The history of psychology is rich with examples of the immediacy and power of basic learning processes. For example, we need only become ill once to

create a strong food aversion, and changes in the frequency of reward schedules can quickly change behavior (23). Indeed, reward learning is currently a central topic of exploration among psychopathologists (see 24). Rodent studies have provided compelling evidence that learning theories can uncover rich information about the neurobiology of socio-emotional behavior. For example, experimental disruption of reward circuitry in the brain prevents mice pups from emitting vocalizations when removed from their mothers. Interfering with brain reward systems also prevents mice from showing a preference for their own mothers (25).

This association also works in the opposite direction: when attachment to the parent is disrupted, other aspects of the animals' reward systems are also affected. To illustrate this point, animals with disrupted attachments to their parents also have abnormal responses to novelty, altered appetitive conditioning, and unusually high sensitivity to dopamine antagonists and reactivity to other drug administrations (see 26,27). Such findings have also been extended to studies of children with disruptive behavioral disorders (28).

Similar types of effects are becoming evident with regard to the emotional development of school-aged children who have had adverse early experiences. Children who have suffered from physical abuse are exposed to inconsistent or poorly conveyed emotional signals in their environments. The adults who ought to be responsible for these children's care tend to vacillate between extreme emotional states of anger and social withdrawal (29). Yet, these social interactions with primary caregivers are the primary basis upon which these children begin to learn about their social environment. For this reason, greater understanding of the brain regions associated with learning reward or punishment is likely to help account for the effects of the environment on these children's interpersonal behavior.

Children who have been physically abused become adept at recognizing cues of anger and hostility (3,30,31). These patterns reflect ways in which children learn to direct their attention to salient and meaningful information in the environment. This type of attention to threat cues in the environment subsequently affects the way children come to construe their social worlds. As an illustration, 5-year old abused children tend to believe that almost any kind of interpersonal situation could result in an adult becoming angry; in contrast, most non-abused children see anger as likely only in particular interpersonal circumstances (32).

These types of data have raised new questions about how probabilistic information about other people's behaviors becomes instantiated in children's thinking. Given that children have a limited processing capacity and that there are limitless aspects of the world that can be attended to at any given moment, it may be the case that abused children prioritize negative social cues at the expense of positive cues. Consistent with this view, on a probabilistic reward task, most children responded quickly as their chances of winning a reward increased. In contrast, maltreated children were not sensitive to the likelihood of reward (33). And primate models also report that maltreated monkeys display less interest in rewards relative to control monkeys (34). A few candidate brain systems have emerged as potentially underlying these phenomena and provide clues about the development of psychopathology.

## CANDIDATE NEURAL SYSTEMS IN DEVELOPMENT AND STRESS

The brain areas that currently receive the most attention from developmental psychopathologists include the prefrontal cortex (a likely candidate because of its protracted period of postnatal development, as well as ties to behavioral regulation abilities such as impulse control and executive functions), the amygdala (because of ties to emotional regulation), and the basal ganglia and orbitofrontal cortex (which, together, seem to represent the outcomes of situations that the organism has experienced) (see 35).

Much current research has been focused on the role of stress on children's cognitive abilities, specifically executive functioning, dependent on the prefrontal cortex. While descriptive studies in children and adults who have experienced specific types of maltreatment are important and informative, many research groups have begun to focus on the idea that it is not specific experiences, such as physical abuse, that affects biobehavioral development, but rather more generally stress and/or instability in children's lives (e.g., 36). A powerful example of this comes from the study by Hanson et al (37), who found that adolescents with high levels of cumulative life stress tend to have smaller volumes in the prefrontal cortex, specifically prefrontal gray and white matter between the anterior cingulate and the frontal poles. Moving beyond simple correlative analyses, this work also revealed that individual differences in prefrontal volumes accounted for the association between cumulative life stress and spatial working memory.

There has also been much research attention, but just as much inconsistency in findings, regarding the amygdala and its role in emotional dysregulation. The divergence in findings may stem from methodological factors, heterogeneous samples of at-risk children, nonlinear effects of life stress, or a combination of all three. To address some of these issues, Hanson et al (38) completed rigorous hand-tracing of the amygdala in samples of children who experienced different forms of early stress, including physical abuse, early neglect or extreme family poverty. They found smaller amygdala volumes in children exposed to these different forms of stress, with brain development associated with both greater cumulative stress exposure and the emergence of child behavioral problems. These data suggest that early and severe life stress may be associated with increased excitation and cell death, reflected in reductions in brain volume. However, caution must be used when inferring developmental patterns from cross-sectional studies; only longitudinal research can truly validate such a model of amygdala

development after early stress exposure. Structural and functional alterations in the amygdala may help us understand individual differences in risk and resilience to behavioral problems as related to toxic stress.

The basal ganglia is a diverse network of subcortical structures that work in concert to orchestrate and execute planned, motivated behaviors that require integration of movement, thinking and feeling (39). The orbitofrontal cortex is a rapidly flexible associative-learning area that is crucial for signaling outcome expectancies such as reward/punishment and the regulation of flexible behavior (40). Current thinking is that the basal ganglia guide learning based on assessments of the probability of a positive outcome, while the orbitofrontal cortex represents gain-loss information and, together, these systems provide a robust way for the organism to learn from and adapt to the environment (41). As expected, impairments in these systems are associated with poor learning from environmental cues. It is especially interesting that orbitofrontal cortex neurons do not stop firing in response to the reward after learning, suggesting that these neurons support predictions on the basis of afferent input and anticipation prior to other emotion-processing regions such as the amygdala (42).

Consistent with this view, damage to the orbitofrontal cortex causes deficits in reversal learning, reduces the speed of reward learning, and is activated in humans during processes such as regret and counterfactual reasoning (43-45). Common to these examples is the need to signal, in real-time, information about outcomes predicted by circumstances in the environment. Some emerging evidence suggests functional changes in the orbitofrontal cortex and basal ganglia during reward processing in adolescents. This further suggests that these systems are a source of developmental changes in social behavior (46).

There is also some evidence that functioning of these systems may account, in part, for how early life stressors confer pervasive lifetime risks for children. Many kinds of early life stressors (maternal separation, social defeat, chronic stress exposure, abuse) appear to alter neurotransmitters and receptors in the basal ganglia that are subsequently associated with impairments in learning (47). Child maltreatment has been associated with lower basal ganglia recruitment during a reward task (48), and children who experienced early life stress have smaller orbitofrontal cortex volumes (49).

What developmental processes might link these components of neural circuitry? One well-understood system is the hypothalamic-pituitary-adrenal (HPA) axis, which is central for understanding the negative effects of stress and trauma on children. When an individual encounters a stressor, corticotropin releasing hormone (CRH) is secreted by the hypothalamus. This hormone acts on the pituitary gland, causing it to release adrenocorticotropic hormone (ACTH). ACTH then acts upon the adrenal gland, resulting in the production of cortisol. Cortisol binds with glucocorticoid receptors in the hippocampus to regulate the HPA axis and inhibit further release of CRH. Similarly, cortisol released in response to stress binds with glucocorticoid receptors at the cellular level to regulate the immune system (50). This system promotes adaptation in response to normative stressors. Toxic or extreme levels of early life stress exposure may impair this system (51).

Other hormone systems also hold potential for understanding how early life adversity affects subsequent social behavior. For example, a recent study examined functioning of the neuropeptide oxytocin in children aged 8-11 years following a social stressor. Girls with histories of physical abuse showed higher levels of urinary oxytocin and lower levels of salivary cortisol following the stressor when compared to controls (52). Abused and control boys, however, did not differ in their hormonal responses. These data suggest that early adversity may disrupt the development of the stress regulation system in girls by middle childhood. Disruptions of this system have implications not only for children's successful regulation of emotion, but also for aspects of comforting behaviors such as the establishment of stable and secure interpersonal relationships.

From a developmental perspective, it is important to emphasize that enhanced threat detection (as well as the myriad systems that children use to promote self-regulation and comforting) are critical for children living in contexts that do not provide adequate protection. Thus, hormone systems such as glucocorticoids and oxytocin that play a role in coordinating these responses (53) may be important targets for interventions aimed at improving children's adjustment.

Accordingly, one of the most promising advances has been the use of epigenetic approaches to understand emotion regulatory processes. Epigenetics may well provide new traction in understanding etiological processes in a range of psychological disorders. We used to think of inheritance in terms of the letters of the DNA code passed from parents' egg and sperm. But now we know that there is another path: parental behavior can write information onto DNA completely bypassing egg and sperm. This adds a level of flexibility to extend a fixed DNA code. This biological flexibility seems quite logical: through experience, individuals use information about the world they are growing up in, changing DNA to cope with the environment.

Not only might actual characteristics of the environment affect gene functioning. It is also possible that children's interpretations and subjective perceptions of their experience is enough to trigger epigenetic changes (54). Given that the behavioral problems of maltreated children are largely accounted for by experiential rather than genetic risk factors (55), this dovetails with observations that maltreated children overly attend to threat/hostility in their environments. Such attentional processes may reflect short-term adaptation to hostile environments, but carry long-term risk for health and behavior.

Although the mechanisms through which these effects are achieved likely involve diverse cellular and molecular pathways, there is emerging evidence supporting the hypothesis that epigenetic changes, such as DNA methylation and

histone modifications, may mediate the effects of early life variations in the social interactions between mothers and infants. Moreover, there may be plasticity within these epigenetic pathways at later developmental time points, such that the social experiences of juveniles and adults may also induce epigenetic change (see 56). These findings have implications for understanding the emergence of behavior problems in early childhood (such as emotion regulation problems) as well as distal problems in adulthood (such as cancer and cardiovascular disease). These data also highlight the dynamic interactions occurring between genes and environments during the course of development.

Recently, epigenetic changes in the glucocorticoid receptor gene were examined by Romens et al (57) in whole blood from children aged 11-14 years. The promoter region of the gene is the sequence needed to turn the gene on and off. It is usually found near the beginning of a gene, and has binding sites for enzymes that make RNA. In the study by Romens et al, abused children had more methylation on several sites within exon 1F of the promoter region of the NR3C1 gene, especially CpG site 3, which may have important implications for brain development, given that it is the binding site for nerve growth factor (58).

These results highlight molecular mechanisms linking childhood stress with biological changes that may lead to mental and physical disorders. Consistent findings across both rodent and human studies suggest that better parental care decreases methylation of the glucocorticoid receptor promoter, increasing the expression of the receptor. Increased expression of the glucocorticoid receptor in the hippocampus reduces stress responsiveness. Though this is an oversimplified explanation (other factors are involved, such as chromatin and histones), the general idea is that methyl inhibits gene transcription and can be thought of as a useful framework for understanding the complexities of gene expression.

But translation across species is difficult. The current glucocorticoid receptor epigenetic data are consistent with the view that genes can be turned on and off; yet such studies in humans cannot infer causality and are limited in terms of specificity of the cellular processes occurring in the brains of living children. They also do not reflect gene expression. What the animal studies can do is to control for confounding variables that are not possible to account for in studies of humans, where we need to be opportunistic in our research.

One clear link between the controlled animal studies and peripheral measurement of epigenetic changes in humans concerns effects of early stress on immune system competence. Indeed, consistent with peripheral changes in methylation of the glucocorticoid receptor gene, children with early stress exposure show deficits in immune functions (17,22).

## CONCLUSIONS

Recent research in developmental psychopathology has increased our understanding of how individuals develop the array of capacities that allow them to cope effectively with challenges posed by each developmental period. This approach is also uncovering new insights into what leads to failures in development of these abilities. In this paper, I have used the example of children's early stress exposure to demonstrate how developmental psychopathologists now tend to deemphasize diagnostic categories and, instead, emphasize the social and biological contexts, events and circumstances that have created opportunities for maladaptive responses and health problems in youth.

Developmental psychopathologists have been less focused on causes of psychopathology and have tried to excavate processes of change. What leads an individual to adopt one pathway of development versus another? From this corpora of scholarship, two useful heuristics emerge. The first is that risk for psychopathology is cumulative. We now understand that aberrant early development of relatively simple skills early in life creates a weak foundation for more complex, later-emerging skills. Similarly, early challenges are likely to build and accumulate over the life course, increasing the burden on an individual and leading to increasingly taxing demands on coping strategies (59).

The second heuristic concerns situating biological development within an environmental context, sometimes called "biological embedding". This reflects an interest in how social contexts "get under the skin" to change biological processes. It is clear that epigenetic changes represent one such possibility (54). Other candidate mechanisms include changes in the neuroendocrine system (53) and altered neural processing of social cues (60-63).

Integrating research about the neurobiology of learning may prove to be a powerful way to test novel hypotheses about how the environment comes to regulate behavior. This is because successful social adaptation reflects children's ability to learn from complex and varied interpersonal experiences. Children need to discern factors including cues for approach versus withdrawal, actions that lead to punishments versus rewards, and which behaviors lead to success in having needs and desires met. These processes become increasingly intricate and fine-tuned as relevant neuroanatomical systems develop, and as the range, complexity and amount of social information increases for the developing child.

A focus on developmental processes allows us to formulate questions about which neural mechanisms we use to process socio-emotional information, how these mechanisms are themselves shaped by social context, why adverse social environments confer risk for children, and, perhaps, what sorts of neutrally informed interventions might remediate deficits in self-regulation.

### Issues for future directions

A number of issues are likely to be the focus of increased interest in the near future. First, it is not yet clear whether it

is most fruitful to focus on specific stressors (such as parent psychopathology or physical abuse), or if a broader conceptualization of the effects of stress on children's development is sufficient (37). For example, Hanson et al (64) found that infants from very low-income families had lower volumes of gray matter, a tissue critical for processing of information and execution of actions. Differences in brain growth were found to vary with socioeconomic status, with children from lower-income households having slower trajectories of growth during infancy and early childhood. These volumetric differences were associated with the emergence of disruptive behavioral problems (64).

Similar cognitive and neurobiological differences have been reported in children who experienced early neglect, especially children raised in institutionalized settings (65, 66). For example, children who suffered early neglect showed developmental deficits in prefrontal white matter microstructure, consistent with more diffuse organization, and this was related to neurocognitive deficits (67). Thus, a broad range of stressful early life experiences may be associated with similar developmental responses.

Another challenge is how to conceptualize stress in children. We do not yet fully understand when stress exposure will be developmentally inconsequential versus harmful. The next wave of research in the field of developmental psychopathology will need to address questions about what kinds of circumstances are necessary for environmental experience to sustain a long-term impact on behavior. We need to better understand thresholds for when issues such as stress move from tolerable to toxic, and to identify the central differences between individual's responses to adversity. Such questions are in the service of leveraging this understanding into treatments that are effective and appropriate for individuals at different phases of development.

## Clinical implications

An elucidation of developmental processes includes understanding adaptation as well as maladaptation. Therefore, a key aspect of developmentally appropriate interventions requires contextualizing a child's behavior in terms of how it may have been useful to the child in the past. It appears that some cognitive, affective and behavioral patterns that emerge in stress-exposed children may have allowed these children to cope with aberrant life circumstances. As an example, in a psychiatric context, we construe anxiety as a disadvantage. Indeed, anxiety is problematic for individuals living in low-danger, highly consistent environments. But if danger or uncertainty is high, then keeping a low profile and responding quickly to possible threat may be useful. For this reason, it is important to view symptoms within the child's life context rather than solely within their present circumstances. If a child is continuing to live in a family context that is unstable, where threat is high, it may well be harmful to reduce the child's anxiety or vigilance to threat. Even at high cost, children need the supports to cope with the realities of their lives.

As clinicians and researchers begin to develop new and effective treatments for children, a challenge will involve learning how to tailor interventions for given individuals based on those individuals' specific biological and environmental circumstances. At present, many treatments for children remain somewhat generic, with popular approaches such as cognitive behavioral, mindfulness or attachment-oriented therapies being applied similarly across a range of mental health conditions, ages and individual differences. In addition, intervention studies tend to focus on very broad, non-specific behavioral outcome measures, such as ratings or interviews of overt symptomatology, school achievement, or observed ratings of behavior. But our behavioral constructs have not yet evolved to have the same level of mechanistic specificity as newer biological measures. More sensitive and specific behavioral measures will be necessary to truly discern the processes underlying mental health issues.

There is hope for effective interventions. Although data suggest that social experiences can alter human physiology, these changes are not necessarily permanent. For example, there is some evidence for epigenetic reversibility from rodents within the glucocorticoid receptor system (68). Such advances will require not only that we discover ways to target and change biobehavioral processes, but that we are able to personalize treatments based on the nature and timing of a child's experience and the individual child's sensitivity/reactivity to those experiences.

If the hypothesis is true that the early life experiences believed to precipitate psychological problems for young people also undermine their lifelong physical health, this would imply that the burden of adult and late-life diseases could also be reduced by successfully improving the psychological health of children. This will be the challenge for the next decade of developmental psychopathology.

## Acknowledgements

The U.S. National Institute of Mental Health, through grant number R01-MH61285, supported the writing of this paper. The author thanks A. Bechner for assistance with the preparation of the manuscript.

## References

1. Cicchetti D (ed). 25th anniversary special issue – Development and psychopathology: a vision realized. Develop Psychopathol 2013;25:1199-653.
2. Nanni V, Uher R, Danese A. Childhood maltreatment predicts unfavorable course of illness and treatment outcome in depression: a meta-analysis. Am J Psychiatry 2012;169:141-51.
3. Shackman J, Pollak SD. Impact of physical maltreatment on the regulation of negative affect and aggression. Develop Psychopathol 2014;26:1021-33.

4. Pollak SD, Cicchetti D, Hornung K et al. Recognizing emotion in faces: developmental effects of child abuse and neglect. Develop Psychol 2000;36:679-88.
5. Romens SE, Pollak SD. Emotion regulation predicts attention bias in maltreated children at-risk for depression. J Child Psychol Psychiatry 2012;53:120-7.
6. Nolen-Hoeksema S, Wisco BE, Lyubomirsky S. Rethinking rumination. Perspect Psychol Sci 2008;3:400-24.
7. Hilt LM, Pollak SD. Characterizing the ruminative process in young adolescents. J Clin Child Adolesc Psychol 2013;42:519-40.
8. Hilt LM, Leitzke BT, Pollak SD. Cognitive control in high- versus low-ruminating youth: the importance of emotion. J Exp Psychopathol 2014;5:302-13.
9. Hilt LM, Pollak SD. Getting out of rumination: comparison of three brief interventions in an adolescent sample. J Abnorm Child Psychol 2012;40:1157-65.
10. Cuthbert BN. The RDoC framework: facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. World Psychiatry 2014;13:28-35.
11. Weinberger DR, Goldberg TE. RDoCs redux. World Psychiatry 2014;13:36-8.
12. Wakefield JC. Wittgenstein's nightmare: why the RDoC grid needs a conceptual dimension. World Psychiatry 2014;13:38-40.
13. Tamminga CA. Approaching a human neuroscience for disease understanding. World Psychiatry 2014;13:41-3.
14. Jablensky A, Waters F. RDoC: a roadmap to pathogenesis? World Psychiatry 2014;13:43-4.
15. Keshavan MS, Ongur D. The journey from RDC/DSM diagnoses toward RDoC dimensions. World Psychiatry 2014;13:44-6.
16. Maj M. Keeping an open attitude towards the RDoC. World Psychiatry 2014;13:1-3.
17. Shirtcliff EA, Coe CL, Pollak SD. Early childhood stress is associated with elevated antibody levels to herpes simplex virus type 1. Proc Natl Acad Sci USA 2009;106:2963-7.
18. Kiecolt-Glaser JK, Gouin JP, Weng NP et al. Childhood adversity heightens the impact of later-life caregiving stress on telomere length and inflammation. Psychosom Med 2011;73:16-22.
19. Miller GE, Chen E, Parker KJ. Psychological stress in childhood and susceptibility to the chronic diseases of aging: moving toward a model of behavioral and biological mechanisms. Psychol Bull 2011;137:959-97.
20. Taylor SE, Way BM, Seeman TE. Early adversity and adult health outcomes. Develop Psychopathol 2011;23:939-54.
21. Shonkoff JP, Garner AS, Committee on Psychosocial, Developmental, and Behavioral Pediatrics. The lifelong effects of childhood adversity and toxic stress. Pediatrics 2012;129:e232-46.
22. Danese A. Developmental psychoneuroimmunology: from bench to bedside. Brain Behav Immun 2014;36:27-8.
23. Gallistel CR, Gibbon J. Time, rate, and conditioning. Psychol Rev 2000;107:289-344.
24. Forbes EE, Goodman SH. Reward function: a promising but (still) underexamined dimension in developmental psychopathology. J Abnorm Psychol 2014;123:310-3.
25. Moles A, Kieffer B, D'Amato F. Deficit in attachment behavior in mice lacking the mu-opioid receptor gene. Science 2004;304:1983-6.
26. Bakermans-Kranenburg MJ, van Ijzendoorn MH. Differential susceptibility to rearing environment depending on dopamine-related genes: new evidence and a meta-analysis. Develop Psychopathol 2011;23:39-52.
27. Matthews K, Robbins T. Early experience as a determinant of adult behavioural responses to reward: the effects of repeated maternal separation in the rat. Neurosci Biobehav Rev 2003;27:45-55.
28. Finger E, Marsh A, Blair K et al. Disrupted reinforcement signaling in the orbitofrontal cortex and caudate in youths with conduct disorder or oppositional defiant disorder and a high level of psychopathic traits. Am J Psychiatry 2011;168:152-62.
29. Shackman J, Fatani S, Camras L et al. Emotion expression among abusive mothers is associated with their children's emotion processing and problem behaviours. Cogn Emot 2010;24:1421-30.
30. Pollak SD, Vardi S, Putzer Bechner A et al. Physically abused children's regulation of attention in response to hostility. Child Dev 2005;76:968-77.
31. Shackman J, Shackman A, Pollak S. Physical abuse amplifies attention to threat and increases anxiety in children. Emotion 2007;7:838-52.
32. Perlman S, Kalish C, Pollak SD. The role of maltreatment experience in children's understanding of the antecedents of emotion. Cogn Emot 2008;22:651-70.
33. Guyer A, Kaufman J, Hodgdon H et al. Behavioral alterations in reward system function: the role of childhood maltreatment and psychopathology. J Am Acad Child Adolesc Psychiatry 2006;45:1059-67.
34. Pryce C, Dettling A, Spengler M et al. Deprivation of parenting disrupts development of homeostatic and reward systems in marmoset monkey offspring. Biol Psychiatry 2004;56:72-9.
35. McCrory E, De Brito SA, Viding E. The link between child abuse and psychopathology: a review of neurobiological and genetic research. J Roy Soc Med 2012;105:151-6.
36. Evans GW, Li D, Whipple SS. Cumulative risk and child development. Psychol Bull 2013;139:1342-96.
37. Hanson JL, Chung MK, Avants BB et al. Structural variations in prefrontal cortex mediates the relationship between early childhood stress and spatial working memory. J Neurosci 2012;32:7917-25.
38. Hanson JL, Nacewicz BM, Sutterer MJ et al. Behavioral problems after early life stress: contributions of the hippocampus and amygdala. Biol Psychiatry 2014;77:314-23.
39. Haber S. The primate basal ganglia: parallel and integrative networks. J Chem Neuroanat 2003;26:317-30.
40. Kringelbach M, Rolls E. The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. Prog Neurobiol 2004;72:341-72.
41. Frank M, Claus E. Anatomy of a decision: striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. Psychol Rev 2006;113:300-26.
42. Schoenbaum G, Roesch MR, Stalnaker TA et al. A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. Nat Rev Neurosci 2009;10:885-92.
43. Honey CJ, Kotter R, Breakspear M et al. Network structure of cerebral cortex shapes functional connectivity on multiple time scales. Proc Natl Acad Sci USA 2007;104:10240-5.
44. Murray EA, Wise SP, Rhodes SEV. What can different brains do with reward? In: Gottfried JA (ed). The neurobiology of sensation and reward. Boca Raton: Taylor & Francis, 2011:61-98.
45. Passingham RE, Stephan KE, Kotter R. The anatomical basis of functional localization in the cortex. Nat Rev Neurosci 2002;3:606-16.
46. Galvan A. Earlier development of the accumbens relative to orbitofrontal cortex might underlie risk-taking behavior in adolescents. J Neurosci 2006;26:6885-92.
47. Desteno D, Schmauss C. A role for dopamine D2 receptors in reversal learning. Neuroscience 2009;162:118-27.
48. Mehta M, Gore-Langton E, Golembo N et al. Hyporesponsive reward anticipation in the basal ganglia following severe institutional deprivation early in life. J Cogn Neurosci 2010;22:2316-25.
49. Hanson J, Chung M, Avants B et al. Early stress is associated with alterations in the orbitofrontal cortex: a tensor-based morphometry investigation of brain structure and behavioral risk. J Neurosci 2010;30:7466-72.
50. Lupien SJ, McEwen BS, Gunnar MR et al. Effects of stress throughout the lifespan on the brain, behavior and cognition. Nat Rev Neurosci 2009;10:434-45.
51. Koss KJ, Hostinar CE, Donzella B et al. Social deprivation and the HPA axis in early development. Psychoneuroendocrinology 2014;50:1-13.

52. Seltzer LJ, Ziegler T, Connolly MJ et al. Stress-induced elevation of oxytocin in maltreated children: evolution, neurodevelopment, and social behavior. Child Dev 2014;85:501-12.
53. Hostinar CE, Gunnar MR. The developmental effects of early life stress: an overview of current theoretical frameworks. Curr Dir Psychol Sci 2013;22:400-6.
54. Slavich GM, Cole SW. The emerging field of human social genomics. Clin Psychol Sci 2013;1:331-48.
55. Jaffee SR, Caspi A, Moffitt TE et al. Physical maltreatment victim to antisocial child: evidence of an environmentally mediated process. J Abnorm Psychol 2004;113:44-55.
56. Champagne FA, Curley JP. Epigenetic influence of the social environment. In: Petronis A, Mill J (eds). Brain, behavior and epigenetics. New York: Springer, 2011:185-208.
57. Romens SE, MacDonald J, Svaren J et al. Associations between early life stress and gene methylation in children. Child Dev 2014;86:303-9.
58. Weaver ICG, 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.
59. Essex MJ, Klein MH, Cho E et al. Maternal stress beginning in infancy may sensitize children to later stress exposure: effects on cortisol and behavior. Biol Psychiatry 2002;52:776-84.
60. Hankin BL, Gibb BE, Abela J et al. Selective attention to affective stimuli and clinical depression among youths: role of anxiety and specificity of emotion. J Abnorm Psychol 2010;119:491-501.
61. Pollak SD. The role of parenting in the emergence of human emotion: new approaches to the old nature-nurture debate. Parent Sci Pract 2012;12:232-42.
62. Reeb-Sutherland BC, Williams LR, Degnan KA et al. Identification of emotional facial expressions among behaviorally inhibited adolescents with lifetime anxiety disorders. Cogn Emot 2015;29: 372-82.
63. Briggs-Gowan MJ, Pollak SD, Grasso D et al. Attention bias and anxiety in young children exposed to family violence. J Child Psychol Psychiatry (in press).
64. Hanson JL, Hair N, Shen DG et al. Family poverty affects the rate of human infant brain growth. PLoS One 2013;8:e80954.
65. Pollak SD, Nelson CA, Schlaak MF et al. Neurodevelopmental effects of early deprivation in postinstitutionalized children. Child Dev 2010;81:224-36.
66. Sheridan MA, Fox NA, Zeanah CH et al. Variation in neural development as a result of exposure to institutionalization early in childhood. Proc Natl Acad Sci USA 2012;109:12927-32.
67. Hanson JL, Adiuru N, Chung MK et al. Early neglect is associated with alterations in white matter integrity and cognitive functioning. Child Develop 2013;84:1566-78.
68. Weaver ICG, Champagne FA, Brown SE et al. Reversal of maternal programming of stress responses in adult offspring through methyl supplementation: altering epigenetic marking later in life. J Neurosci 2005;25:11045-54.

DOI 10.1002/wps.20237

SPECIAL ARTICLE

# How important are the common factors in psychotherapy? An update

**Bruce E. Wampold**

Department of Counseling Psychology, University of Wisconsin, Madison, WI, USA; Modum Bad Psychiatric Center, Vikersund, Norway

*The common factors have a long history in the field of psychotherapy theory, research and practice. To understand the evidence supporting them as important therapeutic elements, the contextual model of psychotherapy is outlined. Then the evidence, primarily from meta-analyses, is presented for particular common factors, including alliance, empathy, expectations, cultural adaptation, and therapist differences. Then the evidence for four factors related to specificity, including treatment differences, specific ingredients, adherence, and competence, is presented. The evidence supports the conclusion that the common factors are important for producing the benefits of psychotherapy.*

**Key words:** Common factors, contextual model, psychotherapy, alliance, empathy, expectations, cultural adaptation, therapist differences, specific ingredients

***(World Psychiatry 2015;14:270–277)***

The so-called common factors have a long history in psychiatry, originating with a seminal article by S. Rosenzweig in 1936 (1) and popularized by J. Frank in the various editions of his book *Persuasion and Healing* (2-4). During this period, the common factors have been both embraced and dismissed, creating some tension (5-9). The purpose of this paper is not to review or discuss the debate, but to provide an update, summarizing the evidence related to these factors.

To understand the evidence for the common factors, it is important to keep in mind that these factors are more than a set of therapeutic elements that are common to all or most psychotherapies. They collectively shape a theoretical model about the mechanisms of change in psychotherapy.

A particular common factor model, called the contextual model, has been recently proposed (8,10). Although there are other common factor models (e.g., 4,11), based on different theoretical propositions, the predictions made about the importance of various common factors are similar and the choice of the model does not affect conclusions about the impact of these factors. The contextual model is presented below, followed by a review of the evidence for the common factors imbedded in the model.

## THE CONTEXTUAL MODEL

The contextual model posits that there are three pathways through which psychotherapy produces benefits. That is, psychotherapy does not have a unitary influence on patients, but rather works through various mechanisms. The mechanisms underlying the three pathways entail evolved characteristics of humans as the ultimate social species; as such, psychotherapy is a special case of a social healing practice.

Thus, the contextual model provides an alternative explanation for the benefits of psychotherapy to ones that emphasize specific ingredients that are purportedly beneficial for particular disorders due to remediation of an identifiable deficit (8).

The three pathways of the contextual model involve: a) the real relationship, b) the creation of expectations through explanation of disorder and the treatment involved, and c) the enactment of health promoting actions. Before these pathways can be activated, an initial therapeutic relationship must be established.

### Initial therapeutic relationship

Before the work of therapy can begin, an initial bond between therapist and patient needs to be created. E. Bordin stated in 1979 that "some basic level of trust surely marks all varieties of therapeutic relationships, but when attention is directed toward the more protected recesses of inner experience, deeper bonds of trust and attachment are required and developed" (12, p. 254). The initial meeting of patient and therapist is essentially the meeting of two strangers, with the patient making a determination of whether the therapist is trustworthy, has the necessary expertise, and will take the time and effort to understand both the problem and the context in which the patient and the problem are situated.

The formation of the initial bond is a combination of bottom-up and top-down processing. Humans make very rapid determination (within 100 ms), based on viewing the face of another human, of whether the other person is trustworthy or not (13), suggesting that patients make very rapid judgments about whether they can trust their therapist. More than likely, patients make rapid judgments about the dress of the therapist, the arrangement and decorations of the room (e.g., diplomas on the wall), and other features of the therapeutic setting (14). However, patients come to therapy with expectations about the nature of psychotherapy as

well, due to prior experiences, recommendations of intimate or influential others, cultural beliefs, and so forth. The initial interaction between patient and therapist is critical, it seems, because more patients prematurely terminate from therapy after the first session than at any other point (15).

### Pathway 1: The real relationship

The real relationship, defined psychodynamically, is "the personal relationship between therapist and patient marked by the extent to which each is genuine with the other and perceives/experiences the other in ways that befit the other" (16, p. 119). Although the psychotherapeutic relationship is influenced by general social processes, it is an unusual social relationship in that: a) the interaction is confidential, with some statutory limits (e.g., child abuse reporting), and b) disclosure of difficult material (e.g., of infidelity to a spouse, of shameful affect, and so forth) does not disrupt the social bond. Indeed, in psychotherapy, the patient is able to talk about difficult material without the threat that the therapist will terminate the relationship.

The importance of human connection has been discussed for decades, whether is it called attachment (17), belongingness (18), social support (19), or the lack of loneliness (20,21). In fact, perceived loneliness is a significant risk factor for mortality, equal to or exceeding smoking, obesity, not exercising (for those with chronic cardiac disease or for healthy individuals), environmental pollution, or excessive drinking (22-24). Psychotherapy provides the patient a human connection with an empathic and caring individual, which should be health promoting, especially for patients who have impoverished or chaotic social relations.

### Pathway 2: Expectations

Research in a number of areas documents that expectations have a strong influence on experience (25). Indeed, the purported price of a bottle of wine influences rating of pleasantness as well as neural representations (26). The burgeoning research on the effects of placebos documents the importance of expectations, as placebos have robustly shown to alter reported experience as well as demonstrating physiological and neural mechanisms (27,28).

Expectations in psychotherapy work in several possible ways. Frank (4) discussed how patients present to psychotherapy demoralized not only because of their distress, but also because they have attempted many times and in many ways to overcome their problems, always unsuccessfully. Participating in psychotherapy appears to be a form of remoralization.

However, therapy has more specific effects on expectations than simple remoralization. According to the contextual model, patients come to therapy with an explanation for their distress, formed from their own psychological beliefs, which is sometimes called "folk psychology" (29-31). These beliefs, which are influenced by cultural conceptualizations of mental disorder but also are idiosyncratic, are typically not adaptive, in the sense that they do not allow for solutions. Psychotherapy provides an explanation for the patient's difficulties that is adaptive, in the sense that it provides a means to overcome or cope with the difficulties. The patient comes to believe that participating in and successfully completing the tasks of therapy, whatever they may be, will be helpful in coping with his or her problems, which then furthers for the patient the expectation that he or she has ability to enact what is needed. The belief that one can do what is necessary to solve his or her problem has been discussed in various ways, including discussions of mastery (4,32), self-efficacy (33), or response expectancies (25).

Critical to the expectation pathway is that patients believe that the explanation provided and the concomitant treatment actions will be remedial for their problems. Consequently, the patient and therapist will need to be in agreement about the goals of therapy as well as the tasks, which are two critical components of the therapeutic alliance (34,35). Hatcher and Barends described the alliance as "the degree to which the therapy dyad is engaged in collaborative, purposive work" (36, p. 293). Creating expectations in psychotherapy depends on a cogent theoretical explanation, which is provided to the patient and which is accepted by the patient, as well as on therapeutic activities that are consistent with the explanation, and that the patient believes will lead to control over his or her problems. A strong alliance indicates that the patient accepts the treatment and is working together with the therapist, creating confidence in the patient that the treatment will be successful.

### Pathway 3: Specific ingredients

The contextual model stipulates that there exists a treatment, particularly one that the patient finds acceptable and that he or she thinks will be remedial for his or her problems, creating the necessary expectations that the patient will experience less distress. Every treatment that meets the conditions of the contextual model will have specific ingredients, that is, each cogent treatment contains certain well-specified therapeutic actions.

The question is how the specific ingredients work to produce the benefits of psychotherapy. Advocates of specific treatments argue that these ingredients are needed to remediate a particular psychological deficit. The contextual model posits that the specific ingredients not only create expectations (pathway 2), but universally produce some salubrious actions. That is, the therapist induces the patient to enact some healthy actions, whether that may be thinking about the world in less maladaptive ways and relying less on dysfunctional schemas (cognitive-behavioral treatments), improving interpersonal relations (interpersonal psychotherapy and some dynamic therapies), being more accepting

of one's self (self-compassion therapies, acceptance and commitment therapy), expressing difficult emotions (emotion-focused and dynamic therapies), taking the perspective of others (mentalization therapies), and so forth. The effect of lifestyle variables on mental health has been understated (37). A strong alliance is necessary for the third pathway as well as the second, as without a strong collaborative work, particularly agreement about the tasks of therapy, the patient will not likely enact the healthy actions.

According to the contextual model, if the treatment elicits healthy patient actions, it will be effective, whereas proponents of specific ingredients as remedial for psychological deficits predict that some treatments – those with the most potent specific ingredients – will be more effective than others (8).

## EVIDENCE FOR VARIOUS COMMON FACTORS

Now that the contextual model has been briefly presented, attention is turned toward an update of the evidence for the common factors. Each factor reviewed is imbedded in the contextual model, although each of them is more generically considered atheoretically as an important one. As will be apparent, many of the common factors are not theoretically or empirically distinct.

To present the evidence succinctly and with as little bias and error as possible, we rely on meta-analyses of primary studies. Studies that examine the association of levels of a common factor and outcome are typically reported by some type of correlation statistic (such as Pearson's product-moment correlation), whereas studies that experimentally manipulate and compare conditions typically report some standardized mean difference (such as Cohen's d). For comparison purposes, correlational statistics are converted to Cohen's d. All meta-analyses reported aggregate statistics, corrected for bias, based on the effects of individual studies appropriated weighted. To understand the importance of effects, Cohen (38) classified a d of 0.2 as small, 0.5 as medium, and 0.8 as large. The evidence is summarized in Figure 1, where the effects of various common factors are compared to those of various specific factors.

### Alliance

The alliance is composed of three components: the bond, the agreement about the goals of therapy, and the agreement about the tasks of therapy (12). As discussed above, alliance is a critical common factor, instrumental in both pathway 2 and pathway 3.

Alliance is the most researched common factor. Typically the alliance is measured early in therapy (at session 3 or 4) and correlated with final outcome. The most recent meta-analysis of the alliance included nearly 200 studies involving over 14,000 patients and found that the aggregate correlation between alliance and outcome was about .27, which is equivalent to a Cohen's d of 0.57 (39), surpassing the threshold for a medium sized effect.

There have been a number of criticisms of the conclusion that alliance is an important factor in psychotherapy (40), most of which have focused on the correlational nature of alliance research. However, each of the criticisms has been considered and has been found not to attenuate the importance of the alliance (see 8).

First, it could well be that early symptom relief causes a strong alliance at the third or fourth session – that is, early responders report better alliances and have better outcomes. To address this threat, early therapy progress must be statistically controlled or longitudinal research is needed to examine the association of alliance and symptoms over the course of therapy. The studies that have examined this question have found evidence to support either interpretation, but the better designed and more sophisticated studies are converging on the conclusion that the alliance predicts future change in symptoms after controlling for already occurring change.

Second, it could be that the correlation between alliance and outcome is due to the patients' contributions to the alliance. According to this line of thinking, some patients may come to therapy well prepared to form a strong alliance and it is these patients who also have a better prognosis, so the alliance-outcome association is due to the characteristics of the patients rather than to something that therapists provide to the patients. Disentangling the patient and therapist contributions involves the use of multilevel modeling. Recently, Baldwin et al (41) performed such an analysis and found that it was the therapist contribution which was important: more effective therapists were able to form a strong alliance across a range of patients. Patients' contribution did not predict outcome: patients who are able to form better alliances, perhaps because they have secure attachment histories, do not have better prognoses. Indeed, patients with poor attachment histories and chaotic interpersonal relationships may well benefit from a therapist who is able to form alliances with difficult patients. These results have been corroborated by meta-analyses (42).

Third, there may be a halo effect if the patient rates both the alliance and the outcome. However, meta-analyses have shown that the alliance-outcome association is robust even when alliance and outcome are rated by different people. It also appears that the alliance is equally strong for cognitive-behavioral therapies as it is for experiential or dynamic treatments, whether a manual is used to guide treatment or not, and whether the outcomes are targeted symptoms or more global measures.

There are other threats to validity of the alliance as a potent therapeutic factor, but the evidence for each of them is nonexistent or weak (8). The research evidence, by and large, supports the importance of the alliance as an important aspect of psychotherapy, as predicted by the contextual model.

As mentioned above, distinctions between certain common factors are difficult to make. A distinction has been

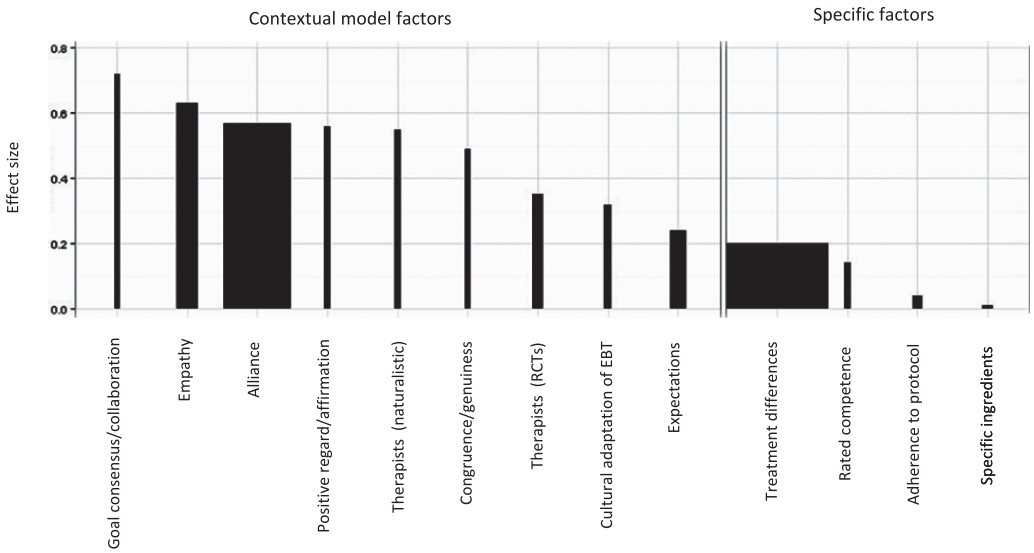


**Figure 1** Effect sizes for common factors of the contextual model and specific factors. Width of bars is proportional to number of studies on which effect is based. RCTs – randomized controlled trials, EBT – evidence-based treatments

made between the bond, as defined as a component of the alliance, which is related to purposeful work, and the real relationship, which is focused on the transference-free genuine relationship (8,16). There is some evidence that the real relationship is related to outcome, after controlling for the alliance (16), and, although the evidence is not strong, it does support the first pathway of the contextual model.

A second construct related for the alliance is labeled goal consensus/collaboration. Although related to agreement about the goals and tasks for therapy, goal consensus/collaboration is measured with different instruments. As shown in Figure 1, the effect for goal consensus and collaboration is strong (d=0.72), based on a meta-analysis of 15 studies (43).

## Empathy and related constructs

Empathy, a complex process by which an individual can be affected by and share the emotional state of another, assess the reasons for another's state, and identify with the other by adopting his or her perspective, is thought to be necessary for the cooperation, goal sharing, and regulation of social interaction. Such capacities are critical to infant and child rearing, as children, who are unable to care for themselves, signal to the caregiver that care is needed, a process that is then put to use to manage social relations among communities of adult individuals. Therapist expressed empathy is a primary common factor, critical to pathway 1 of the contextual model, but which also augments the effect of expectations.

The power of the empathy in healing was beautifully revealed in a study of placebo acupuncture for patients with irritable bowel syndrome (44). Patients with this syndrome were randomly assigned to a limited interaction condition, an augmented relationship condition, or treatment as usual (waiting list for acupuncture). In the limited interaction condition, the acupuncturist met with the patient briefly, but was not allowed to converse with him or her, and administered the sham acupuncture (a device that gives the sensation of having needles pierce the skin, but they do not). In the augmented relationship condition, the practitioner conversed with the patient about the symptoms, the relevance of lifestyle and relationships to irritable bowel syndrome, as well as the patient's understanding of the cause and meaning of her disorder. All this was done in a warm and friendly manner, using active listening, appropriate silences for reflection, and a communication of confidence and positive expectation. For the four dependent variables (global improvement, adequate relief, symptom severity, and quality of life), the two sham acupuncture conditions were superior to treatment as usual. However, the augmented relationship condition was superior to the limited interaction condition, particularly for quality of life.

The above study is noteworthy because it was an experimental demonstration of the importance of a warm, caring, empathic interaction within a healing setting. Unfortunately, experimental manipulation of empathy in psychotherapy studies is not possible, for design and ethical reasons. Nonetheless, there have been numerous studies (n=59) that have correlated ratings of therapist empathy with outcome, which have been meta-analytically summarized (45), resulting in a relatively large effect (d=0.63; see Figure 1). Constructs related to empathy have also been meta-analyzed and found to be related to outcome, including positive regard/affirmation (d=0.56, n=18; see Figure 1) (46) and congruence/genuineness (d=0.49, n=18; see Figure 1) (47).

It should be recognized that several of the threats to validity for the alliance are also present with regard to empathy. For example, it is clearly easier for a therapist to be warm and

caring toward a motivated, disclosing and cooperative patient than to one who is interpersonally aggressive, and the former types of patients will most likely have better outcomes than the latter, making the empathy/outcome correlation an artifact of patient characteristics rather than therapist action. Unfortunately, studies such as the ones conducted to rule out these threats to validity for the alliance have not been conducted for empathy and related constructs.

## Expectations

Examining the role of expectations in psychotherapy is difficult. In medicine, expectations can be induced verbally and then physicochemical agents or procedures can be administered or not, making the two components (creation of expectations and the treatment) independent. In psychotherapy, creating the expectations, through explanation of the patient's disorder, presenting the rationale for the treatment, and participating in the therapeutic actions, is part of therapy. It is difficult to design experimental studies of expectations in psychotherapy (not impossible, but not yet accomplished in any important manner).

The typical way to assess the effect of expectations in psychotherapy is to correlate patient ratings of their expectations with outcomes, but we have seen that such correlational studies produce threats to validity. Furthermore, in many studies, expectations are measured *prior* to when the rationale for the treatment is provided to the patient, when it is the explanation given to the patient that is supposed to create the expectations. Assessing expectations after the explanation has been given (i.e., during the course of treatment) is also problematic, as those patients who have made significant progress in therapy will naturally respond that they think therapy will be helpful.

Despite the difficulties with investigating expectations in psychotherapy, this is a topic of much interest (48-50). Recently, a meta-analysis of expectations showed that there was a relatively small, but statistically significant, relationship between rated expectations and outcome ($d=0.24$, $n=46$; see Figure 1) (49). The best evidence for expectations in the context of healing is derived from studies of the placebo effect, where exquisite care has been taken to experimentally manipulate variables of interest and to control for threats to validity, by using physiological and neurological variables as well as subjective reports. A summary of this literature is beyond the scope of this article, but many excellent reviews are available (8,27,28).

## Cultural adaptation of evidence-based treatments

The contextual model emphasizes that the explanation given for the patient's distress and the therapy actions must be acceptable to the patient. Acceptance is partly a function of consistency of the treatment with the patient's beliefs, particularly beliefs about the nature of mental illness and how to cope with the effects of the illness. This suggests that evidence-based treatments that are culturally adapted will be more effective for members of the cultural group for which the adapted treatment is designed. There are many ways to adapt treatments, including those involving language, cultural congruence of therapist and patient, cultural rituals, and explanations adapted to the "myth" of the group.

A recent meta-analysis demonstrated that adapting evidence-based treatments by using an explanation congruent with the cultural group's beliefs (i.e., using the cultural "myth" as the explanation) was more effective than unadapted evidence-based treatments, although the effect was modest ($d=0.32$, $n=21$; see Figure 1) (51).

## Therapist effects

Therapist effects are said to exist if some therapists consistently achieve better outcomes with their patients than other therapists, regardless of the nature of the patients or the treatment delivered. Therapist effects have been studied in clinical trials and in naturalistic settings. In both designs, the measure of therapist effects is an intraclass correlation coefficient. Technically, this coefficient indexes the degree to which two patients from the same therapist have similar outcomes relative to two patients from two different therapists. To compare therapist effects to other common factors, the intraclass correlation coefficient is converted to Cohen's d.

The contextual model predicts that there will be differences among therapists *within* a treatment. That is, even though the therapists are delivering the same specific ingredients, some therapists will do so more skillfully and therefore achieve better outcomes than other therapists delivering the same treatment. Evidence for this conjecture is found in clinical trials. A meta-analysis of therapist effects in clinical trials found modest therapist effects ($d=0.35$, $n=29$; see Figure 1) (52). Keep in mind that the therapists in clinical trials generally are included because of their competence and then they are given extra training, supervised, and monitored. Moreover, the patients in such trials are homogeneous, as they have a designated diagnosis and are selected based on various inclusionary/exclusionary criteria. In such designs, patients are randomly assigned to therapists. Consequently, consistent differences among therapists in such trials, although modest, are instructive.

Not surprisingly, therapist effects in naturalistic settings are greater than in clinical trials. In the former settings, therapists are more heterogeneous, patients may not be randomly assigned to therapists, patients are heterogeneous, and so forth. A meta-analysis of therapist effects in such settings found a relatively large effect ($d=0.55$, $n=17$; see Figure 1) (52).

The finding of robust therapist effects raises the question about what are the characteristics or actions of more effective therapists. Recent research has begun to address this

question. Studies have shown that effective therapists (vis-à-vis less effective therapists) are able to form stronger alliances across a range of patients, have a greater level of facilitative interpersonal skills, express more professional self-doubt, and engage in more time outside of the actual therapy practicing various therapy skills (8).

## SPECIFIC EFFECTS

Evidence for the common factors is also collected by examining the evidence for specific aspects of psychotherapy. The contextual model makes several predictions about specific effects, which will be discussed as each specific effect is considered.

### Treatment differences

When pathway 3 of the contextual model was discussed earlier, it was emphasized that the model contends that all therapies with structure, given by empathic and caring therapists, and which facilitate the patient's engagement in behaviors that are salubrious, will have approximately equal effects. That is, the specific ingredients, discussed in pathway 3, are not critical because they remediate some psychological deficit.

The question of whether some treatments are superior to others has long been debated, with origins at the very beginning of the practice of psychotherapy (think about the disagreements amongst Freud, Adler and Jung, for example). Today, there are claims that some treatments, in general or for specific disorders, are more effective than others. Others, however, claim that there are no differences among psychotherapies, in terms of their outcomes.

The literature addressing this issue is immense and summarizing the results of relative efficacy is not possible. Nevertheless, the various meta-analyses for psychotherapies in general or for specific disorders, if they do find differences among various types of treatment, typically find at most differences of approximately d=0.20, the value shown in Figure 1.

### Specific effects from dismantling studies

To many, the dismantling design is the most valid way to identify the effects of specific ingredients. In this design, a specific ingredient is removed from a treatment to determine how much more effective the treatment is in total compared to the treatment without the ingredient that is purportedly remedial for the psychological deficit.

Two meta-analyses have examined dismantling designs and both found minimal differences between the total treatment and the treatment without one or more critical ingredients (d=0.01, n=30, see Figure 1) (53,54). The most recent of these meta-analyses did find that adding an ingredient to an existing treatment increased the effect for targeted variables by a small amount (d=0.28) (53).

### Adherence and competence

In clinical trials, it is required that adherence to the protocol and the competence at delivering the treatment are rated. This makes sense: if the goal is to make inferences about a particular treatment, then it is necessary to ensure that the treatment was delivered with the necessary components and not with extraneous components (i.e., with *adherence* to the protocol) and that the treatment components were delivered skillfully (i.e., given *competently*).

It would seem logical theoretically that adherence to the protocol and competence would be related to outcome. That is, for cases where the therapist followed the protocol and did so skillfully, there should be better outcomes. However, this is not the case. In a meta-analysis of adherence and competence (55), effects were small (d=0.04, n=28 for adherence; d=0.14, n=18 for competence; see Figure 1).

The results for adherence and competence demand further explanation. If the specific ingredients of a treatment are critical, then adherence should make a difference – actually delivering those ingredients should be related to outcome. There is evidence that rigid adherence to a protocol can attenuate the alliance and increase resistance to the treatment (i.e., failing to accept the treatment, a contextual model tenet) (8), and that flexibility in adherence is related to better outcomes (56), results consistent with prediction of the contextual model.

The findings for competence are a bit more difficult to understand. Competence in these trials typically is rated by experts in the treatment being given, based on watching therapy sessions. Why can't experts differentiate between "good" therapy and "bad" therapy? If this were indicative of experts' abilities to judge competence, then the notion of psychotherapy supervision would be turned upside down, because what is observed and evaluated would have no relation to outcomes – how could the supervisor then make a case for providing input to the supervisee? But the clue to the resolution of this mystery is found in the definition of competence. Most psychotherapy trials rate the competence *for a specific treatment*. That is, what is rated is the skill in providing the elements of the treatment protocol, rather than common factors, such as empathy, alliance, affirmation, and so forth – aspects of therapy that do predict outcome and seem to differentiate more effective therapists from less effective therapists.

## CONCLUSIONS

Although the common factors have been discussed for almost a century, the focus of psychotherapy is typically on the development and dissemination of treatment models. If

not discounted, then the common factors are thought of as perhaps necessary, but clearly not sufficient. The evidence, however, strongly suggests that the common factors must be considered therapeutic and attention must be given to them, in terms of theory, research and practice.

One of the criticisms of the common factors is that they are an atheoretical collection of commonalities. In this paper, the contextual model was presented to convey a theoretical basis for these factors.

## References


1. Rosenzweig S. Some implicit common factors in diverse methods of psychotherapy: "At last the Dodo said, 'Everybody has won and all must have prizes'". Am J Orthopsychiatry 1936;6:412-5.
2. Frank JD. Persuasion and healing: a comparative study of psychotherapy. Baltimore: Johns Hopkins University Press, 1961.
3. Frank JD. Persuasion and healing: a comparative study of psychotherapy, 2nd ed. Baltimore: Johns Hopkins University Press, 1973.
4. Frank JD, Frank JB. Persuasion and healing: a comparative study of psychotherapy, 3rd ed. Baltimore: Johns Hopkins University Press, 1991.
5. Barlow DH. Psychological treatments. Am Psychol 2004;59:869-78.
6. Laska KM, Gurman AS, Wampold BE. Expanding the lens of evidence-based practice in psychotherapy: a common factors perspective. Psychotherapy 2014;51:467-81.
7. Wampold BE. The great psychotherapy debate: model, methods, and findings. Mahwah: Lawrence Erlbaum Associates, 2001.
8. Wampold BE, Imel ZE. The great psychotherapy debate: the research evidence for what works in psychotherapy, 2nd ed. New York: Routledge, 2015.
9. Baker TB, McFall RM, Shoham V. Current status and future prospects of clinical psychology: toward a scientifically principled approach to mental and behavioral health care. Psychol Sci Publ Int 2008;9:67-103.
10. Wampold BE, Budge SL. The 2011 Leona Tyler Award address: the relationship – and its relationship to the common and specific factors of psychotherapy. Couns Psychol 2012;40:601-23.
11. Orlinsky DE, Howard KI. Process and outcome in psychotherapy. In: Garfield SL, Bergin AE (eds). Handbook of psychotherapy and behavior change. New York: Wiley, 1986:311-81.
12. Bordin ES. The generalizability of the psychoanalytic concept of the working alliance. Psychotherapy: Theory, Research and Practice 1979;16:252-60.
13. Willis J, Todorov A. First impressions: making up your mind after a 100-ms exposure to a face. Psychol Sci 2006;17:592-8.
14. Heppner PP, Claiborn CD. Social influence research in counseling: a review and critique. J Couns Psychol 1989;36:365-87.
15. Connell J, Grant S, Mullin T. Client initiated termination of therapy at NHS primary care counselling services. Couns Psychother Res 2006;6:60-7.
16. Gelso C. A tripartite model of the therapeutic relationship: theory, research, and practice. Psychother Res 2014;24:117-31.
17. Bowlby J. Attachment and loss. New York: Basic Books, 1980.
18. Baumeister RF. The cultural animal: human nature, meaning, and social life. New York: Oxford University Press, 2005.
19. Cohen S, Syme SL. Social support and health. San Diego: Academic Press, 1985.
20. Cacioppo S, Cacioppo JT. Decoding the invisible forces of social connections. Front Integr Neurosci 2012;6:51.
21. Lieberman MD. Social: why our brains are wired to connect. New York: Crown Publishing Group, 2013.
22. Holt-Lunstad J, Smith TB, Baker M et al. Loneliness and social isolation as risk factors for mortality: a meta-analytic review. Perspect Psychol Sci 2015;10:227-37.
23. Holt-Lunstad J, Smith TB, Layton JB. Social relationships and mortality risk: a meta-analytic review. PLoS Med 2010;7:e1000316.
24. Luo Y, Hawkley LC, Waite LJ et al. Loneliness, health, and mortality in old age: a national longitudinal study. Soc Sci Med 2012; 74:907-14.
25. Kirsch I. How expectancies shape experience. Washington: American Psychological Association, 1999.
26. Plassmann H, O'Doherty J, Shiv B et al. Marketing actions can modulate neural representations of experienced pleasantness. Proc Natl Acad Sci USA 2008;105:1050-4.
27. Price DP, Finniss DG, Benedetti F. A comprehensive review of the placebo effect: recent advances and current thought. Annu Rev Psychol 2008;59:565-90.
28. Benedetti F. Placebo effects: understanding the mechanisms in health and disease, 2nd ed. New York: Oxford University Press, 2014.
29. Boyer P, Barrett HC. Domain specificity and intuitive ontologies. In: Buss DM (ed). The handbook of evolutionary psychology. Hoboken: Wiley, 2005:96-118.
30. Molden DC, Dweck CS. Finding 'meaning' in psychology: a lay theories approach to self-regulation, social perception, and social development. Am Psychol 2006;61:192-203.
31. Thomas RM. Folk psychologies across cultures. Thousand Oaks: Sage, 2001.
32. Liberman BL. The role of mastery in psychotherapy: maintenance of improvement and prescriptive change. In: Frank JD, Hoehn-Saric R, Imber SD et al (eds). Effective ingredients of successful psychotherapy. New York: Brunner/Mazel, 1978:35-72.
33. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. In: Baumeister RF (ed). The self in social psychology. New York: Psychology Press, 1999:285-98.
34. Crits-Christoph P, Connolly Gibbons MB, Crits-Christoph K et al. Can therapists be trained to improve their alliances? A preliminary study of alliance-fostering psychotherapy. Psychother Res 2006;16:268-81.
35. Horvath AO. The alliance in context: accomplishments, challenges, and future directions. Psychotherapy: Theory, Research, Practice, Training 2006;43:258-63.
36. Hatcher RL, Barends AW. How a return to theory could help alliance research. Psychotherapy: Theory, Research, Practice, Training 2006;43:292-9.
37. Walsh R. Lifestyle and mental health. Am Psychol 2011;66:579-92.
38. Cohen J. Statistical power analysis for the behavioral sciences, 2nd ed. Hillsdale: Erlbaum, 1988.
39. Horvath AO, Del Re AC, Fluckiger C et al. Alliance in individual psychotherapy. Psychotherapy 2011;48:9-16.
40. DeRubeis RJ, Brotman MA, Gibbons CJ. A conceptual and methodological analysis of the nonspecifics argument. Clin Psychol Sci Pract 2005;12:174-83.
41. Baldwin SA, Wampold BE, Imel ZE. Untangling the alliance-outcome correlation: exploring the relative importance of therapist and patient variability in the alliance. J Consult Clin Psychol 2007;75:842-52.
42. Del Re AC, Flückiger C, Horvath AO et al. Therapist effects in the therapeutic alliance-outcome relationship: a restricted-maximum likelihood meta-analysis. Clin Psychol Rev 2012;32:642-9.
43. Tryon GS, Winograd G. Goal consensus and collaboration. In: Norcross JC (ed). Psychotherapy relationships that work: evidence-based responsiveness, 2nd ed. New York: Oxford University Press, 2011:153-67.
44. Kaptchuk TJ, Kelley JM, Conboy LA et al. Components of placebo effect: randomised controlled trial in patients with irritable bowel syndrome. BMJ 2008;336:999-1003.

45. Elliott R, Bohart AC, Watson JC et al. Empathy. Psychotherapy 2011;48:43-9.
46. Farber BA, Doolin EM. Positive regard. Psychotherapy 2011;48:58-64.
47. Kolden GG, Klein MH, Wang CC et al. Congruence/genuineness. Psychotherapy 2011;48:65-71.
48. Greenberg RP, Constantino MJ, Bruce N. Are patient expectations still relevant for psychotherapy process and outcome? Clin Psychol Rev 2006;26:657-78.
49. Constantino MJ, Arnkoff DB, Glass CR et al. Expectations. J Clin Psychol 2011;67:184-92.
50. Arnkoff DB, Glass CR, Shapiro SJ. Expectations and preferences. In: Norcross JC (ed). Psychotherapy relationships that work: therapist contributions and responsiveness to patients. Oxford: Oxford University Press, 2002:335-56.
51. Benish SG, Quintana S, Wampold BE. Culturally adapted psychotherapy and the legitimacy of myth: a direct-comparison meta-analysis. J Couns Psychol 2011;58:279-89.
52. Baldwin SA, Imel ZE. Therapist effects: finding and methods. In: Lambert MJ (ed). Bergin and Garfield's handbook of psychotherapy and behavior change. New York: Wiley, 2013:258-97.
53. Bell EC, Marcus DK, Goodlad JK. Are the parts as good as the whole? A meta-analysis of component treatment studies. J Consult Clin Psychol 2013;81:722-36.
54. Ahn H, Wampold BE. Where oh where are the specific ingredients? A meta-analysis of component studies in counseling and psychotherapy. J Couns Psychol 2001;48:251-7.
55. Webb CA, DeRubeis RJ, Barber JP. Therapist adherence/competence and treatment outcome: a meta-analytic review. J Consult Clin Psychol 2010;78:200-11.
56. Owen J, Hilsenroth MJ. Treatment adherence: the importance of therapist flexibility in relation to therapy outcomes. J Couns Psychol 2014;61:280-8.

DOI 10.1002/wps.20238

# Advance directives in mental health care: evidence, challenges and promise

Heather Zelle[1], Kathleen Kemp[2], Richard J. Bonnie[1]

[1]University of Virginia School of Law, Charlottesville, VA 22903, USA; [2]Department of Child and Adolescent Psychiatry, Rhode Island Hospital, and Warren Alpert Medical School, Brown University, Providence, RI 02903, USA

Psychiatric advance directives (PADs) are written documents or oral statements that allow adults with decision-making capacity to declare their treatment preferences and/or to designate proxy decision makers to act on their behalf should they be deemed incapable in the future of making informed choices on their own.

In the U.S., the Patient Self-Determination Act (1) created momentum for recovery-oriented care, which has led to the enactment of mental health-related advance planning legislation in about two-thirds of the states (2,3). Internationally, increasing attention to such tools is found in the U.K., Ireland, Germany, Belgium, Canada, New Zealand, Australia and India.

## EVIDENCE

A recent theoretical framework (4) discerned from the existing literature three complementary facets of the PAD intervention process: a) enhancement of consumer autonomy; b) improvement of consumer and treatment provider therapeutic alliance; c) integration of care through system partnerships.

### Enhancement of consumer autonomy

PADs improve psychiatric and recovery-oriented outcomes by empowering consumers with serious mental illness to take an active role in their own care (5), choosing among high-quality, evidence-based treatments in the least restrictive setting possible.

PADs are thought to embody a recovery-oriented philosophy by encouraging consumers to preselect their treatments for times of future crises. Research has shown that consumers who have executed PADs endorse feelings of self-determination, autonomy, and empowerment (6-9).

### Improvement of consumer and treatment provider therapeutic alliance

PADs also facilitate communication between providers and consumers about future treatment choices, and these discussions improve therapeutic relationships (7) as well as provide clinically relevant treatment information (10,11). In fact, research suggests that 95% of PADs are rated both clinically useful and consistent with clinical treatment standards (7,10).

In the context of completing PADs, facilitation refers to a collaborative process between a consumer and a provider that informs the consumer about PADs, engages the consumer in a discussion of past treatment experiences, and helps the consumer work through the process of documenting future treatment preferences and instructions.

Clinician- or treatment provider-facilitated PADs may also improve consumer uptake of PADs. Up to three quarters of consumers indicate they would complete a PAD if provided the choice and support (6,7,12). Thus far, the facilitation process has significantly reduced barriers to PAD completion, with increases in completion of almost 30 times compared to non-facilitated PAD models (7,13).

PADs may also reduce negative coercive treatment experiences. Compared to consumers without PADs, consumers with facilitated PADs were approximately half as likely to require a coercive intervention during a mental health crisis over a 24-month follow-up period (12). This is particularly important because consumers' fear of coercive treatment interventions reduce their willingness to interact with the mental health system and engage in treatment (14).

### Integration of care through system partnerships

Despite these positive signs, mixed or even no evidence exists about the impact of PADs on primary outcomes such as psychiatric admissions, compliance with treatment, harm to self or others, or treatment utilization. Henderson et al (15) demonstrated a reduced number of involuntary psychiatric admissions for PAD completers when facilitated by the individual's clinician; however, similar outcome research showed no effect on psychiatric admissions with non-clinician facilitated PADs (16).

Similarly, there is a lack of research and evidence on the use of PADs to coordinate care across providers/institutions. There is mixed evidence, though, about the thoughts and practices of providers within single institutions (e.g., 17).

## CHALLENGES IN IMPLEMENTATION

Experience has demonstrated that many barriers interfere with implementation and use of PADs. Opponents and

proponents alike acknowledge the low usage rates of PADs, which fall below the usage rates of advance directives focused on only end-of-life care (18). Low usage rates are not attributable to a lack of interest, however, as the same study that showed usage rates of 4-13% across five cities also found that 66-77% of consumers reported interest in PADs when told about them (18).

An important recent advance in the consideration of barriers is the use of taxonomies. Barriers can be identified by the intervention stage at which they occur: intervention design, PAD completion, or PAD access and honoring (4,19). Barriers can also be identified by the level at which they occur: system level, agency level, and individual level (19,20). Arguably, barriers begin even before PAD services are created, as many stakeholders continue to hold misperceptions or conflicting perceptions about PADs and their use (e.g., 21-23).

Once implementation is undertaken, system-level barriers include legal impediments (e.g., unauthorized practice of law, misunderstanding of legal duties and ramifications) and obstacles to communication (e.g., lack of cross-system collaboration) (19,20).

Agency-level barriers include difficulties in integrating a new practice into existing agency culture, need for training, lack of resources (e.g., overworked staff, lack of payment for facilitation services), and impediments to coordinating services (e.g., creating a referral system, engaging doctors as needed for portions of PADs, electronic health record integration).

Individual-level barriers can include engaging clients (both initially and over time, because advance care planning is a process), understanding difficult material, communicating with one's providers and loved ones, and taking steps to ensure that the PAD will be readily accessible (19,20).

Finally, it is well worth noting that, although low- and middle-income countries may be expected to face additional barriers, recent research suggests that completion of PADs is feasible in those countries (24).

## PROMISE

The continuing appeal of PADs in the face of many challenges is likely based on several factors, one of which is the growing attention to patient autonomy across health care systems in several countries (25) and treatment ideologies that advance such moral principles – namely, recovery-oriented models (26-30).

As noted earlier, the U.S. increased its attention to patient autonomy beginning in the early 1990s, with additional developments such as the New Freedom Commission on Mental Health report that prompted national administrative attention to recovery (31). The last decade and a half has seen similar policy and practice developments in the United Nations (32); European countries, such as Ireland, U.K. and Belgium (33,34); Australasian countries, such as Australia and New Zealand (35,36), and India (24,37).

In the U.S., the Commonwealth of Virginia has enacted a particularly forward-thinking revision to its health care decision laws: mental health care was woven into the language of the general Health Care Decisions Act, thus treating it on par with other major domains of health care about which an individual can document decisions (38). Virginia also adopted a presumption that all adults have capacity to make legally binding advance directives, and that a determination of incapacity cannot be based upon diagnosis alone (38). Another innovation that expands individuals' ability to make treatment decisions is Virginia's full inclusion in its law of a "Ulysses Clause" (the person authorizes the doctor in advance to ignore him/her, during future crises, when he/she is saying "No" to treatment) (38).

On the international stage, the United Nations Convention on the Rights of Persons with Disabilities supported a similar approach to individuals' right to autonomy and control over treatment (32).

The fact that PADs instantiate several desirable principles and concepts of care also lends to their appeal. The many facets of PADs may appeal differentially to various user groups: health care consumers benefit from the advancement of autonomy; consumers and clinicians benefit from improved working alliance; and consumers, providers and care systems benefit from coordination of care.

Some individuals and cultures value independence highly, so there is a natural draw to the self-determination that PADs can create. In comparison, some other individuals and cultures value family or group dynamics more highly, in which case PADs are also desirable because they allow for decision making among loved ones and/or for an individual to take a burden off of loved ones by planning ahead (e.g., 39). Thus, PADs have the ability to appeal to multiple audiences simultaneously (4,19).

## CONCLUSIONS

As the many challenges noted above suggest, implementation of PADs has been difficult despite their intuitive appeal. A PAD is a single tool embodying multiple principles and care concepts meant to be used in different ways by several types of stakeholders across multiple providers in what are typically disjointed health care systems (19,20).

Efforts to embed use of PADs in routine mental health care can benefit from research on strategies for increasing their usage and a burgeoning literature on dissemination and implementation of health care innovations (e.g., 40-43), as well as from studies on health behavior change (44-46).

## References

1. Patient Self-Determination Act of 1990. 42 U.S. Code §§ 1395cc, 1396a.

2. Srebnik D, Fond JL. Advance directives for mental health treatment. Psychiatr Serv 1999;50:919-25.
3. Menninger KA. Advance directives for medical and psychiatric care. Vol. 102, 3rd ed. New York: Thomson Reuters, Lawyers Cooperative Publishing, 2014.
4. Nicaise P, Lorant V, Dubois V. Psychiatric advance directives as a complex and multistage intervention: a realist systematic review. Health Soc Care Comm 2013;21:1-14.
5. Swanson JW, Tepper M, Backlar P et al. Psychiatric advance directives: an alternative to coercive treatment? Psychiatry 2000; 63:160-72.
6. Srebnik D, Brodoff L. Implementing psychiatric advance directives: service provider issues and answers. J Behav Health Serv Res 2003;30:253-68.
7. Swanson JW, Swartz M, Ferron J et al. Psychiatric advance directives among public mental health consumers in five U.S. cities: prevalence, demand, and correlates. J Am Acad Psychol Law 2006;34:43-57.
8. Elbogen EB, Swanson JW, Appelbaum P et al. Competence to complete psychiatric advance directives: effects of facilitated decision making. Law Hum Behav 2007;31:275-89.
9. Henderson C, Lee R, Herman D et al. From psychiatric advance directives to joint crisis plan. Psychiatr Serv 2009;60:1390-1.
10. Srebnik D, Rutherford L, Peto T et al. The content and clinical utility of psychiatric advance directives. Psychiatr Serv 2005;56:592-8.
11. Srebnik D, Russo J. Use of psychiatric advance directives during psychiatric crisis events. Adm Policy Ment Health 2008;35:272-82.
12. Swanson JW, Swartz M, Elbogen EB et al. Psychiatric advance directives and reduction of coercive crisis interventions. J Ment Health 2008;17:255-67.
13. Van Dorn RA, Swanson JW, Swartz MS et al. Reducing barriers to completing psychiatric advance directives. Adm Policy Ment Health 2008;35:440-8.
14. Joshi K. Psychiatric advance directives. J Psychiatr Pract 2003;9: 303-6.
15. Henderson C, Flood C, Leese M et al. Effect of joint crisis plans on use of compulsory treatment in psychiatry: single blind randomised controlled trial. BMJ 2004;329:136.
16. Campbell LA, Kisely SR. Advance treatment directives for people with severe mental illness. Cochrane Database Syst Rev 2009;1: CD005963.
17. Kim MM, Van Dorn RA, Scheyett AM et al. Understanding the personal and clinical utility of psychiatric advance directives: a qualitative perspective. Psychiatry 2007;70:19-29.
18. Swanson JW, Swartz M, Elbogen EB et al. Facilitated psychiatric advance directives: a randomized trial of an intervention to foster advance treatment planning among persons with severe mental illness. Am J Psychiatry 2006;163:1943-51.
19. Zelle H, Oliver JE, Bonnie RJ et al. Implementing advance directives in mental health services: a manual for providers and advocates. Charlottesville: University of Virginia Institute of Law, Psychiatry, and Public Policy, 2015.
20. Kemp K, Zelle H, Bonnie RJ. Embedding advance directives in routine care for persons with serious mental illness: the challenge of implementation. Psychiatr Serv 2015;66:10-4.
21. Elbogen E, Swartz M, Dorn RV et al. Clinical decision making and views about psychiatric advance directives. Psychiatr Serv 2006;57:350-5.
22. Van Dorn RA, Swartz MS, Elbogen EB et al. Clinicians' attitudes regarding barriers to the implementation of psychiatric advance directives. Adm Policy Ment Health 2006;33:449-60.
23. Wilder CM, Swanson JW, Bonnie RJ et al. A survey of stakeholder knowledge, experience, and opinions of advance directives for mental health in Virginia. Adm Policy Ment Health 2013;40:232-9.
24. Ramesh Kumar TC, John S, Gopal S et al. Psychiatric advance statements: an Indian experience. Int J Soc Psychol 2012;59:531-4.
25. Beauchamp TL, Childress JF. Principles of biomedical ethics, 7th ed. New York: Oxford University Press, 2013.
26. Davidson L, O'Connell MJ, Tondora J et al. Recovery in serious mental illness: a new wine or just a new bottle? Prof Psychol Res Pract 2005;36:480-7.
27. Jacobson N, Greenley D. What is recovery? A conceptual model and explication. Psychiatr Serv 2001;52:482-5.
28. Onken SJ, Craig CM, Ridgway P et al. An analysis of the definitions and elements of recovery: a review of the literature. Psychiatr Rehabil J 2007;31:9-22.
29. U.S. Substance Abuse and Mental Health Services Administration. National consensus statement on mental health recovery (SMA05-4129). Rockville: U.S. Substance Abuse and Mental Health Services Administration, 2004.
30. Slade M, Amering M, Farkas M et al. Uses and abuses of recovery: implementing recovery-oriented practices in mental health systems. World Psychiatry 2014;13:12-20.
31. New Freedom Commission on Mental Health. Achieving the promise: transforming mental health care in America. Rockville: U.S. Substance Abuse and Mental Health Services Administration, 2003.
32. Morrissey F. The United Nations Convention on the Rights of Persons with Disabilities: a new approach to decision-making in mental health law. Eur J Health Law 2012;19:423-40.
33. Morrissey F. Advance directives in mental health care: hearing the voice of the mentally ill. Med Leg J Ireland 2010;16:21.
34. Nicaise P, Soto VE, Dubois V et al. Users' and health professionals' values in relation to a psychiatric intervention: the case of psychiatric advance directives. Adm Policy Ment Health 2015;42: 384-93.
35. Saraf S. Advance statements in the new Victorian Mental Health Act. Australas Psychiatry 2015;23:230-2.
36. Weller P. Psychiatric advance directives and human rights. Psychiatry Psychol Law 2010;17:218-29.
37. Sarin A, Murthy P, Chatterjee S. Psychiatric advance directives: potential challenges in India. Indian J Med Ethics 2012;9:104-7.
38. Virginia Health Care Decisions Act of 1992. VA Code §§ 54.1-2981 through 54.1-2996.
39. Van Dorn RA, Swanson JW, Swartz MS. Preferences for psychiatric advance directives among Latinos: views on advance care planning for mental health. Psychiatr Serv 2009;60:1383-5.
40. Henderson C, Jackson C, Slade M et al. How should we implement psychiatric advance directives? Views of consumers, caregivers, mental health providers and researchers. Adm Policy Ment Health 2010;37:447-58.
41. Aarons GA, Horowitz JD, Dlugosz LR et al. The role of organizational processes in dissemination and implementation research. In: Brownson RC, Colditz GA, Proctor EK (eds). Dissemination and implementation research in health: translating science to practice. New York: Oxford University Press, 2012:128-53.
42. Berwick DM. Disseminating innovations in health care. JAMA 2003;289:1969-75.
43. Durlak JA, DuPre EP. Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. Am J Commun Psychol 2008;41:327-50.
44. Bandura A. Self-efficacy: the exercise of control. New York: W.H. Freeman, 1997.
45. Deci EL, Ryan RM. The "what" and "why" of goal pursuits: human needs and the self-determination of behavior. Psychol Inq 2000;11:227-68.
46. Rothschild ML. Carrots, sticks, and promises: a conceptual framework for the management of public health and social issue behaviors. J Marketing 1999;63:24-37.

DOI 10.1002/wps.20268

# Joint crisis planning in mental health care: the challenge of implementation in randomized trials and in routine care

Claire Henderson[1], Simone Farrelly[1], Paul Moran[1], Rohan Borschmann[2], Graham Thornicroft[1], Max Birchwood[3], The Crimson[1,3-5] and Joshua[1,2] Study Groups

[1]Health Service and Population Research Department, King's College London, Institute of Psychiatry, Psychology and Neuroscience, London, UK; [2]Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Australia; [3]Division of Mental Health and Wellbeing, University of Warwick, Warwick, UK; [4]Division of Psychiatry, School of Medicine, University of Manchester, Manchester, UK; [5]Community Based Medicine, University of Manchester, Manchester, UK

Joint crisis planning produces a plan for use during a future mental health crisis or relapse. Its distinguishing feature is facilitation by a mental health professional external to the treatment team, who engages a mental health service user and members of his/her treatment team in a process of shared decision making.

To date, there have been three trials of joint crisis plans, producing two key findings. First, the process of producing and using a joint crisis plan is highly appreciated by service users, can improve therapeutic relationships and reduce the rate of involuntary measures, and is likely to be cost-effective. Second, joint crisis plans are challenging to produce and use, exemplifying the widespread difficulty within medicine of adopting shared decision making.

In this context, the aim of this paper is to consider whether repeated emphasis on individualized crisis planning in policy documents will be sufficient to bring about the adoption of shared decision making in mental health care. Experience from the above-mentioned three trials provides indications of what other measures may help.

## HOW IS JOINT CRISIS PLANNING DIFFERENT FROM ROUTINE CARE?

Written treatment plans are routine in many community mental health services and many contain an action plan for crisis or relapse. Their chief goal is to ensure timely, co-ordinated and effective care.

In England, the Care Programme Approach (CPA, 1) provides a framework for care of the most vulnerable mental health service users, including those at risk for suicide and self-harm and people with a history of relapses requiring urgent intervention. Further guidance (2,3) has re-emphasized the need to undertake detailed crisis planning, and the Mental Capacity Act 2005 (4) provides for advance refusals of treatment in a crisis.

In the U.S., supporting people to create a psychiatric advance directive is viewed as a component of recovery-oriented treatment planning (5). Psychiatric advance directives promote consumer choice and prioritize the goal of autonomy.

Routine treatment plans lie at the other, more paternalistic, end of the crisis planning spectrum, as they may be produced without service user involvement, although by consensus this is not seen as good practice. Most routine crisis plans in England remain stubbornly "one size fits all" (6). Within the National Health Service organizations participating in the CRIMSON multisite randomized controlled trial of joint crisis plans (7), at baseline only 15% of participants had a crisis plan containing any information specific to that individual (6). The inference is that most community mental health teams do not consider individualized crisis plans a priority.

Joint crisis planning lies toward the centre of the above spectrum, as an application of the shared decision making model (8,9). To achieve this, it employs an external facilitator to complete the crisis plan, instead of the service user's care co-ordinator or case manager. The facilitator aims to engage the service user and treating mental health professionals during formulation of the joint crisis plan. Developed after consultation with service user groups (10), this process aims to empower service users whilst facilitating early detection and treatment of relapse. Held by the service user, a joint crisis plan contains his/her treatment preferences for any future psychiatric emergency using first person language.

## WHAT DIFFERENCE DOES THE JOINT CRISIS PLANNING PROCESS MAKE IN COMPARISON TO ROUTINE TREATMENT PLANNING?

Results published in 2004 of a single site randomized controlled trial of joint crisis plans for people with psychotic or bipolar illness showed reduced rate of involuntary hospitalization associated with their use (11) and generally positive views of the plan among service users and mental health professionals (12). Similarly, in 2006, a U.S. study of facilitated psychiatric advance directives showed an improvement in working alliance at one month (13). A more recent randomized controlled trial in the Netherlands found that crisis planning was associated with a reduction in court-ordered admission to hospital (14), but not other forms of

involuntary admission. However, this intervention did not involve an external facilitator.

The CRIMSON multisite trial (N=569) sought to provide definitive evidence on the effectiveness of joint crisis plans delivered in routine practice (7). No significant treatment effect was seen for the primary outcome of involuntary hospitalization or secondary outcomes of overall psychiatric hospital admissions, length of stay, perceived coercion and engagement with services. However, there was a positive effect on service user-rated therapeutic relationships, consistent with the 2004 trial (11) and the trial of facilitated psychiatric advance directives (13). Qualitative trial data (15) supported the improvement in therapeutic relationships when clinicians engaged well in the discussion. Service users reported that the facilitator helped to address power imbalances and that clinicians listened more and were more reasonable.

However, lack of engagement amongst some clinicians may have undermined the potential effect of planning (for instance, psychiatrists' lack of attendance or engagement at the planning meeting, or lack of awareness of the joint crisis plan on the part of subsequent clinicians following staff turnover) (16). Moreover, while some clinicians believed the external facilitator was necessary for empowering service users, others feared potential interference. Finally, many clinicians believed that they already engaged in joint crisis planning, or that crisis planning was a bureaucratic exercise of little value due to lack of service user choice.

While the main outcomes from CRIMSON might support some of these views, other evidence from this trial does not. Contrary to the assertion that the joint crisis plan adds little to routine practice, an audit of routine crisis plans of the trial participants showed that individualization was infrequent (6). Further, content analysis of the joint crisis plans showed a wide range of service user choices, that were on the whole clinically reasonable, including efforts to self-manage early warning signs of relapse and some requests for hospitalization (17). Finally, while clinicians endorsed shared decision making approaches and believed that they were enacting it in routine care, reports from service users contradicted this view (15). It seems that more needs to be done to convince clinicians of the potential benefits of the approach.

## ARE JOINT CRISIS PLANS RELEVANT AND HELPFUL FOR SERVICE USER GROUPS OTHER THAN THOSE WITH PSYCHOSIS?

The single site JOSHUA randomized controlled trial (18) was set up to develop and provide a preliminary test of the effectiveness of joint crisis plans for people with borderline personality disorder, who are especially vulnerable to the experience of crises and their adverse consequences, particularly in terms of self-harm. Again, participants' views were generally strongly positive: joint crisis plans were used both during (74%) and between (44%) crises, and approximately half of intervention participants reported experiencing a greater sense of control over their mental health problems and an improved relationship with their mental health team at follow-up (19).

Nevertheless, the trial failed to demonstrate superiority for the primary outcome, self-reported self-harm, and also for all secondary outcomes. This was despite an excellent level and rate of joint crisis plan production, although subsequent problems in adherence to the contents may have reduced its effectiveness. For this trial, the production process excluded treating psychiatrists as a response to service user preference. The trial under-recruited, thus, the absence of positive significant findings in favour of joint crisis plans may partly have been explained by type II error.

## THE ECONOMICS OF CRISIS PLANNING

The provision of facilitators to ensure high quality crisis planning may appear prohibitively costly. However, the 2004 trial of joint crisis plans showed that their use was cost-effective relative to the control condition (available non-individualized treatment information plus routine care planning) (20). Likewise, the JOSHUA randomized controlled trial showed that there was at least an 80% probability that the joint crisis plan plus treatment as usual was more cost-effective than treatment as usual (19).

The economic evaluation of CRIMSON (21) showed no evidence for the total sample of cost-effectiveness of the joint crisis plan. However, analysis by ethnic subgroup showed there is at least a 90% probability of the joint crisis plan intervention being more cost-effective than treatment as usual in the Black ethnic group.

## CONCLUSIONS

Joint crisis plans may be cost-effective for Black people with psychotic or bipolar illness (21) and people with borderline personality disorder (19), two groups for whom mental health services have tended to provide the least satisfactory care. This suggests that any future study of joint crisis plans should target service users whom the clinical team are particularly struggling to engage in collaborative working.

In England, those who are poorly engaged with services are likely to be subject to a paternalistic approach in the form of a community treatment order. This has not been shown to be effective in reducing involuntary admissions or any other outcomes (22,23). Interventions such as the joint crisis plan are welcomed by service users when the clinical team engages with the process, and this may improve therapeutic alliance (12,13,15). However, although many clinicians endorse the general idea of shared decision making (24), the variability of adoption reflects a mixed response to this method of operationalizing it.

To date, external facilitation has not been adopted in the UK. However, without the facilitator, the application of shared decision making to crisis planning is likely to continue to be variable. One way to resolve this dilemma would be to train care co-ordinators/case managers to provide external facilitation for other teams as part of a reciprocal arrangement among teams, thus adding to their own skills in encouraging shared decision making. Whether addressing the barrier to adoption in this way leads to positive outcomes in routine care remains to be seen.


## Acknowledgements

The first joint crisis plan trial and the CRIMSON and JOSHUA trials were all funded by the UK Medical Research Council. The first trial was also funded through a Medical Research Council/London Region fellowship awarded to C. Henderson. The JOSHUA trial was supported by a Medical Research Council trial platform grant (ID: 85397). M. Birchwood is in part funded by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Healthcare and Research (West Midlands, UK). The authors also acknowledge financial support from the UK Department of Health via the NIHR Biomedical Research Centre and Dementia Unit awarded to South London and Maudsley NHS Foundation Trust in partnership with King's College London and King's College Hospital NHS Foundation Trust.


## References


1. Department of Health. The care programme approach. London: Department of Health, 1991.
2. Department of Health. National service framework for mental health. Modern standards and service models. London: Department of Health, 1999.
3. Department of Health. Refocusing the Care Programme Approach: policy and positive practice guidance. London: Department of Health, 2008.
4. Department of Constitutional Affairs. Mental Capacity Act – Code of Practice. Norwich: The Stationery Office, 2005.
5. Deegan PE, Drake R. Shared decision making and medication management in the recovery process. Psychiatr Serv 2006;57:1636-9.
6. Farrelly S, Szmukler G, Henderson C et al. Individualisation in crisis planning for people with psychotic disorders. Epidemiol Psychiatr Sci 2014;23:353-9.
7. Thornicroft G, Farrelly S, Szmukler G et al. Clinical outcomes of joint crisis plans to reduce compulsory treatment for people with psychosis: a randomised controlled trial. Lancet 2013;381:1634-41.
8. Charles C, Gafni A, Whelan T. Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango). Soc Sci Med 1997;44:681-92.
9. Henderson C, Swanson JW, Szmukler G et al. A typology of advance statements in mental health care. Psychiatr Serv 2008; 59:63-71.
10. Sutherby K, Szmukler GI, Halpern A et al. A study of 'crisis cards' in a community psychiatric service. Acta Psychiatr Scand 1999; 100:56-61.
11. Henderson C, Flood C, Leese M et al. Effect of joint crisis plans on use of compulsory treatment in psychiatry: single blind randomised controlled trial. BMJ 2004;329:136.
12. Henderson C, Flood C, Leese M et al. Views of service users and providers on joint crisis plans: single blind randomized controlled trial. Soc Psychiatry Psychiatr Epidemiol 2009;44:369-76.
13. Swanson JW, Swartz MS, Elbogen EB et al. Facilitated psychiatric advance directives: a randomized trial of an intervention to foster advance treatment planning among persons with severe mental illness. Am J Psychiatry 2006;163:1943-51.
14. Ruchlewska A, Wierdsma AI, Kamperman AM et al. Effect of crisis plans on admissions and emergency visits: a randomized controlled trial. PLoS One 2014;9:e91882.
15. Farrelly S, Lester, H, Rose D et al. Improving therapeutic relationships: joint crisis planning for individuals with psychotic disorders. Qual Health Res (in press).
16. Farrelly S, Lester H, Rose D et al. Barriers to shared decision making: qualitative study of the joint crisis plan for psychosis. Health Expect (in press).
17. Farrelly S, Brown G, Rose D et al. What service users with psychotic disorders want in a mental health crisis or relapse: thematic analysis of joint crisis plans. Soc Psychiatry Psychiatr Epidemiol 2014;49:1609-17.
18. Moran P, Borschmann R, Flach C et al. The effectiveness of joint crisis plans for people with borderline personality disorder: protocol for an exploratory randomised controlled trial. Trials 2010;11: 18.
19. Borschmann R, Barrett B, Hellier JM et al. Joint crisis plans for people with borderline personality disorder: feasibility and outcomes in a randomised controlled trial. Br J Psychiatry 2013;202: 357-64.
20. Flood C, Byford S, Henderson C et al. Joint crisis plans for people with psychosis: economic evaluation of a randomised controlled trial. BMJ 2006;333:729.
21. Barrett B, Waheed W, Farrelly S et al. Randomised controlled trial of joint crisis plans to reduce compulsory treatment for people with psychosis: economic outcomes. PLoS One 2013;8:e74210.
22. Burns T, Rugkasa J, Molodynski A et al. Community treatment orders for patients with psychosis (OCTET): a randomised controlled trial. Lancet 2013;381:1627-33.
23. Rugkasa J, Molodynski A, Yeeles K et al. Community treatment orders: clinical and social outcomes, and a subgroup analysis from the OCTET RCT. Acta Psychiatr Scand 2015;131:321-9.
24. Seale C, Chaplin R, Lelliot P. Sharing decisions in consultations involving anti-psychotic medication: a qualitative study of psychiatrists' experiences. Soc Sci Med 2006;62:2861-73.


DOI 10.1002/wps.20256

# Differential diagnosis and current polythetic classification

JOSEF PARNAS

Psychiatric Center Hvidovre, Faculty of Health and Medical Sciences, and Center for Subjectivity Research, Faculty of Humanities, University of Copenhagen, Copenhagen, Denmark

The introduction of polythetic diagnostic classification (DSM-III and ICD-10) in psychiatry was anticipated to improve the reliability of psychiatric diagnoses, facilitate research, and eventually allow the then future DSM-IV to become anchored in objective, etiological criteria. However, the preparations and release of DSM-IV and DSM-5 highlighted the fact that the etiological promise has not materialized. Psychiatric classifications will continue in the foreseeable future to be based mainly on clinical descriptions.

This has stimulated a broad range of reflections and critiques of psychiatric nosology (e.g., 1,2). Yet, the criticism is typically confined to concrete, technical issues, e.g. discussing a necessity for novel categories, modification of existing criteria, correcting definitions and misunderstandings, etc.. The polythetic-operational *foundation* of current classification remains largely unchallenged (3). Thus, despite a nearly universal consensus about the etiological stalemate, psychiatrists continue to believe that the polythetic system is epistemologically adequate and that it has indeed broadly improved clinical diagnostic reliability.

I wish to question the alleged improvement of reliability and to challenge the epistemological adequacy of the polythetic approach. The issue of differential diagnosis will serve as a concrete clinical embodiment of this critique. Needless to say, a full discussion of the theoretical and clinical ramifications of these topics (e.g., the issue of "comorbidity") is beyond the scope of this article.

## IMPROVED RELIABILITY

Diagnostic reliability is typically reported as interrater agreement for selected disorders in the so-called "field trials", accompanying the construction of diagnostic criteria, or in research studies. The data behind such reports stem from somewhat artificially constructed measurement contexts that formalize what actually happens in the ordinary, everyday clinical practice. Such reports also tend to embellish or even inflate the presented reliability levels (4).

We have no data on the general quality (reliability and concurrent validity) of contemporary clinical diagnostic practices, or data comparing the general utility of successive diagnostic systems. We lack anthropologically oriented research, emphasizing ecological aspects of reliability, i.e. examining the actual reliability of working psychiatrists as they assign a diagnosis, situated in their daily environment. Instead, we may use other sources of information that, at least indirectly, may shed some light upon these issues. Several, disparate signs jointly indicate a still existing and serious problem with diagnostic reliability.

First, we witness the epidemic-like explosion of certain diagnoses (e.g., autistic spectrum, attention-deficit/hyperactivity disorder). Such dramatic increases may be due to an actual and *true* rise in the incidence, changes in treatment seeking behavior, or availability of novel and more efficacious treatments. However, common clinical experience suggests that quite often such "popularities" emerge because physicians become unduly impressed by newly circulated checklists targeting specific disorders while failing to perform a comprehensive diagnostic assessment, or because physicians are unaware of or simply ignore the diagnostic rules.

Thus, a study of referrals to a mental health center in Netherlands (5) found that, among 242 first-contact patients reporting at least one unequivocal psychotic symptom, only 44% were diagnosed with psychosis, whereas 56% received a non-psychosis diagnosis or no diagnosis at all. In another study of patients discharged with a diagnosis of schizoaffective disorder from two Danish university clinics (6), only 10% of cases actually fulfilled operational criteria for that disorder, whereas the remainder suffered from schizophrenia or bipolar disorder.

Finally, the forensic-psychiatric odyssey of the Norwegian mass-murderer A. Breivik, independently assessed by two teams of psychiatric experts with the resulting ICD-10 diagnoses of paranoid schizophrenia and personality disorder respectively, does not testify to a dramatic improvement of reliability (7,8).

## THE POLYTHETIC-OPERATIONAL DIAGNOSTIC SYSTEM

A polythetic diagnostic category of current DSM/ICD is based on a list of symptoms and signs believed to be characteristic for the diagnosis in question. Typically, a certain number of diagnostically equivalent symptoms or signs from a given list is sufficient to arrive at a diagnosis. These "diagnostic criteria" are, contrary to a widespread belief, not "operational" in any epistemological or scientific sense. They are just briefly described in an ordinary non-technical lay language at "the lowest order of inference" (3).

Two issues deserve attention here. First, the symptoms/signs *shared* by two or several disorders tend to be omitted from the diagnostic lists in order to strengthen the clinical distinctiveness of the categories (e.g., depressed mood and anxiety are exclusively listed in the context of mood and anxiety disorders). Second, the simplification of the psychopathological descriptions to brief, lay language statements converts the symptoms and signs into phenomenological primitives or homogeneous elementals. There is only one kind of delusion (i.e., it is assumed that all delusions share the same phenomenological structure), one kind of anxiety, one kind of auditory verbal hallucination, etc.. Consequently the syndromes, solely constituted by aggregates of such elementals, lose their characteristic salience, and their boundaries become blurred. A recent study using a network model of DSM-IV symptoms demonstrated that half of the symptoms are connected with short paths. The individual disorders are therefore also mutually proximate, accounting for the high levels of empirically observed comorbidity (9).

The narrative, conceptual and phenomenological descriptions of pre-DSM-III psychopathology were eliminated from the contemporary diagnostic manuals. Those descriptions contained a discussion of the characteristic prototypes of mental disorders, their phenomenological structures and the interdependency of their constituent features (e.g., in the manic syndrome, the potential relations between the global "volatility" of the manic gestalt, increased mood, vitality, psychomotor speed, and grandiosity). They also contained a consideration of the phenomenological structure of the individual symptoms and signs, their relations of implication or entailment, and their context dependence. Such information no longer exists in the diagnostic manuals and is largely gone into oblivion.

For example, a reader of DSM-IV or DSM-5 is told that schizophrenia is a mixture of positive and negative symptoms that happens to satisfy certain inclusion and exclusion criteria. This definition says more about what schizophrenia *is not* (e.g., non-organic, non-affective) than *what it is* (10), i.e. what kind of validity is behind this category (11), what is its characteristic gestalt that constitutes its difference from other potentially similar mixtures of positive and negative symptoms (12), what justifies schizophrenia's dominating diagnostic rank in the taxonomic hierarchy, or why it is risky to expose a patient with schizophrenia to an orthodox psychoanalysis.

## MAKING DIAGNOSIS AND DEFINING CONCEPTS: PROTOTYPES AND GESTALTS

The process of differential diagnosis in the pre-DSM-III era was framed by *prototypical* considerations. Although such considerations still take place or, more exactly, *cannot avoid* to take place in any diagnostic situation (including somatic medicine), in psychiatry they only operate on an implicit, un-reflected level, because they are un-anchored and incompatible with the philosophy underlying the polythetic-operational classification.

A prototype is a central example of a given category (a sparrow is more typical of the category "bird" than is a penguin or an ostrich), with a graded dilution of typicality towards its borders, where it eventually overlaps with neighboring prototypes. Thus, the prototypical categories exhibit an intrinsic dimensionality (13). However, a prototype *is not just an example* (exemplar), but contains condensed information on its internal configuration of properties and its relations to neighboring prototypes (14). The concept of prototype/gestalt is fit for description of single symptoms and signs as well as larger entities such as diagnostic categories. One can use the concept of prototype-gestalt in a narrow or a wide sense, neither one limited to perception but also involving complex cognitive-affective operations.

In a narrow sense, a gestalt is a unity or organization of phenomenal aspects, that emerges from the interactions among its component features (part-whole relations). The whole is irreducible to a mere aggregate, because it is more than a sum of its parts. In a diagnostic process there are reciprocal dependencies between the whole and its single features. The clinical whole confers on its constitutive features their characteristic diagnostic significance. Conversely, the single clinical features, by instantiating the gestalt, imbue it with clinical concreteness and rootedness (12).

In a wider sense, the notion of gestalt entails an interplay of factors that extend beyond the subject to include not only a mental state, but also the patient's engagements with the environment and others. For instance, detecting a delusion involves taking into account not only the patient's verbal contents but also his experiences, way of arguing, relational style and relevant historical information. To use the concept of delusion competently, a psychiatrist must master plenty of other prototypes and concepts (e.g., psychosis, rationality, reality, hallucination, etc.) (8).

The argument for a prototype-based diagnosis is fundamentally linked to the fact that perception is always apperceptively (conceptually) informed: perceiving something is to perceive it *as a something*, as a token of a certain type. A perceptual or cognitive object is always given as a certain gestalt. The unfamiliar is perceived in terms of the familiar, i.e. in terms of the general type or gestalt that is "activated in the particular perception" (15). This process is called *typification* and is intrinsic to all human cognition and hence to the diagnostic process as well.

The natural unfolding of a comprehensive semi-structured prototype-based diagnostic assessment involves *reflective and critical questioning* of typifications, which become supported, weakened or discarded by explicitly elicited diagnostic information on symptoms, their evolution, social history etc., progressively limiting the number of diagnostic options (16). Typification as such can never be eliminated because it is an automatic aspect of cognition. A recent review of mechanisms involved in concept formation, use and understanding suggests that *concepts* (e.g., psychiatric categories)

are not constituted by a list of criteria, but are organized around prototypes/gestalts (17): "Theory of concepts must be primarily prototype-based…, within a broader knowledge representation scheme in which the concept is positioned both within a hierarchy and within a theoretical framework(s) appropriate to that domain" (14, p. 488). It follows that the more knowledgeable and experienced is the psychiatrist, the more refined is the diagnostic repertoire.

Finally, it needs to be emphasized that prototypically defined and described nosological categories may be enriched and supplemented by lists of criteria. This was, in fact, the original but, unfortunately, unrealized intention behind the DSM-III (3).

## CLINICAL REALITY OF DIFFERENTIAL DIAGNOSIS IN A POLYTHETIC SYSTEM

All diagnosis is an instance of differential diagnosis: the task is to pick up, from a larger catalogue of potential options, the one that most adequately fits the patient.

Let us then imagine a young clinician in an open outpatient facility, trained with the DSM/ICD manuals as her exclusive source of psychopathological knowledge. She encounters a self-referred male in his early 20-ies, sitting on the floor of the waiting room in a lotus position, mumbling, and occasionally laughing to himself in a silly manner. How should she proceed after the initial greetings?

Since her diagnostic-cognitive field lacks a prototypical-conceptual grid, she is exposed to what in cognitive science is known as a "frame problem", i.e. the issue of how to decide what is relevant, indeed what is even the relevant overall context within which to approach a given problem (16). Theoretically, she would therefore need to explore the inclusion and exclusion criteria for nearly all disorders (the number will vary with the degree of diagnostic hierarchy). That is, obviously, not feasible in practice. Instead, she may imitate a digital computer and use a structured interview. Such interview is essentially constructed as a binary decision tree with mandatory probing questions and suggested cut-off points. The epistemological problems and the quite meager pragmatic utility of structured interviews have been amply addressed (16,18). Here, it is important to note that the very nature of structured questioning confers a limited diagnostic utility on the interview, because of the low sensitivity and specificity of the responses. Responding with a "no" or "yes" to the question of "feeling down" neither excludes nor strongly supports any specific diagnosis.

Most likely, our clinician will conduct a so-called "clinical interview", a conversation starting with the patient's complaints and reasons for seeking help, and assisted by various symptom checklists locally in use. In this process, the patient may be diagnosed with major depression if he answers affirmatively to five or six criteria of this diagnosis. In other words, for a psychiatrist untrained to impose a conceptual-psychopathological grid on the diagnostic information, the patient's initial behavior (suggestive of schizophrenia) may easily fail to display a relevant clinical salience and hence fail to enter into the diagnostic considerations.

Thus, a young psychiatrist, unfamiliar with the prototypical structure of psychopathology, finds herself exposed to a myriad of chaotic, unconnected data, where each individual feature is equally worthy of attention and may therefore become a pivot of a potential diagnostic class. With growing experience, this clinician will invariably acquire her own *private* prototypes, shaped by the local ideologies and habits and by personal inclinations, i.e., in an implicit way that *is not exposed to an academic, rigorous, and peer-shared reflection*. "Private" prototypes become easily activated by single, popping up clinical features that happen to evoke a single aspect of a contingent diagnostic category. Here, a decisive role is often played by the very first verbalizations of the complaint. If a patient mentions a habit of cutting herself, a "borderline" diagnosis will be likely considered.

## CONCLUSIONS

The distinctions and concepts in the realm of experience and behavior play now, and will continue to play, a decisive role in psychiatric classifications. These distinctions do not function with the simplicity of causal referents, as it is often the case with signs and symptoms of somatic medicine (e.g. jaundice → bilirubine cycle). Rather, they exhibit a phenomenological-empirical and theoretical complexity, which cannot be adequately represented through the simplifying, reductive approach of the operational-polythetic system.

The differential diagnostic process is not (only) a matter of a digitalized decision tree, but involves context dependencies and complex pattern recognitions. These empirical, phenomenological and theoretical issues constitute the domain of the *science of psychopathology*. In recent decades, research, study and training in psychopathology have been seen as largely redundant, because the polythetic manuals seemed to offer all that was needed for research and practice. These assumptions have proven to be false.

## References


1. Hyman SE. The diagnosis of mental disorders: the problem of reification. Annu Rev Clin Psychol 2010;6:155-79.
2. Frances AJ, Widiger T. Psychiatric diagnosis: lessons from the DSM-IV past and cautions for the DSM-5 future. Annu Rev Clin Psychol 2012;8:109-30.
3. Parnas J, Bovet P. Psychiatry made easy: operation(al)ism and some of its consequences. In: Kendler KS, Parnas J (eds). Philosophical issues in psychiatry III: The nature and sources of historical change. Oxford: Oxford University Press, 2014:190-212.
4. Vanheule S. Diagnosis and the DSM. A critical review. Hampshire: Palgrave Macmillan, 2014.
5. Boonstra N, Wuderink L, Sytema S et al. Detection of psychosis by mental health services; a naturalistic cohort study. Clin Pract Epidemiol Ment Health 2008;4:29.

6. Vollmer-Larsen A, Jacobsen TB, Hemmingsen R et al. Schizoaffective disorder – the reliability of its clinical diagnostic use. Acta Psychiatr Scand 2006;11:402-7.
7. Melle I. The Breivik case and what psychiatrists can learn from it. World Psychiatry 2013;12:16-21.
8. Parnas J. The Breivik case and "conditio psychiatrica". World Psychiatry 2013;12:21-2.
9. Borsboom D, Cramer AOJ, Schmittmann VD et al. The small world of psychopathology. PLoS One 2011;6:e27407.
10. Maj M. A critique of DSM-IV operational criteria for schizophrenia. Br J Psychiatry 1998;172:458-60.
11. Parnas J. A disappearing heritage: the clinical core of schizophrenia. Schizophr Bull 2011;37:1121-30.
12. Parnas J. The core Gestalt of schizophrenia. World Psychiatry 2012;11:67-9.
13. Westen D. Prototype diagnosis of psychiatric syndromes. World Psychiatry 2012;11:16-21.
14. Murphy GL. The big book of concepts. Cambridge: MIT Press, 2002.
15. Mishara A, Schwartz MA. Who's on first? Mental disorders by any other name? Assoc Adv Philos Psychiatry Bull 2010;17:60-5.
16. Nordgaard J, Sass LA, Parnas J. The psychiatric interview: validity, structure, and subjectivity. Eur Arch Psychiatry Clin Neurosci 2013;263:353-64.
17. Machery E. Doing without concepts. New York: Oxford University Press, 2009.
18. Nordgaard J, Revsbech R, Saebye D et al. Assessing the diagnostic validity of a structured psychiatric interview in a first-admission hospital sample. World Psychiatry 2012;11:181-5.

DOI 10.1002/wps.20239

# Psychiatric disorders: natural kinds made by the world or practical kinds made by us?

PETER ZACHAR

Department of Psychology, Auburn University Montgomery, Montgomery, AL 36117, USA

The concept of natural kind, a term of art in philosophy, is being increasingly appropriated by mental health professionals (1-4). First introduced in the 19th century, the notion of a natural kind has benefited from sustained philosophical attention over the past forty years (5-7). Newly developed ways of thinking about the concept are worth taking note of in psychiatry.

Typical examples of natural kinds are chemical elements such as gold, biological species such as tiger, and infectious diseases such as tuberculosis. These all: a) are naturally occurring as opposed to artificial; b) have clearly demarcated boundaries separating members of the natural kind from non-members; c) possess observable features that are causally produced by internal properties; and d) these causal properties can be used to objectively validate category membership. Also, studying what instances of a kind have in common allows us to know what to expect of the kind in general.

Few would claim that currently available psychiatric taxonomies classify natural kinds. Diagnostic co-occurrence and use of the *not elsewhere* and *not otherwise specified* codes are widespread, underlying pathological processes shared by all cases have not been discovered, and no diagnostically reliable biomarkers have been identified. In addition, treatments are frustratingly non-specific. A psychiatric taxonomy of natural kinds is at present only an aspirational ideal (8,9).

## THE ESSENTIALIST BIAS AND TAXONOMIC THINKING

In an essentialist framework for species taxonomies, there are tigers and lions, but no ligers. What makes something a "real" lion is a set of hidden properties – called the essence or nature of the species. To be an instance of a natural kind is to possess the essence of the kind.

Essentialism retains perennial importance because humans are readily disposed to think about biological categories in terms of essences (10). In fact, developmental psychologists have discovered that children begin to adopt essentialist assumptions about category membership in preschool. Children see category membership as fixed, rooted in hidden, unchanging causal properties, and more useful than appearances for making inferences about expected behavior (11,12).

The introduction of empiricism by thinkers such as J. Locke initiated a long and productive critique of essentialist metaphysics in modern philosophy (13-15). The empiricists contended that the notion of an essence is an empty abstraction. Locke was inspired not only by the scientific revolution and 17th century English politics, but also by his own work as a physician. He learned the craft as a collaborator of T. Sydenham. Together, they believed that medical classifications should be based upon observable natural histories of diseases rather than theorizing about hidden causes/essences (16,17).

The most philosophically important rejection of essentialism among scientists is found in Darwin's theory of evolution. According to Darwin, rather than being a fixed type in which every member of a species shares the same essence, a species is a population of individuals that vary. In fact, many phenomena in nature contradict our essentialist assumptions – including the production of tiger/lion hybrids such as ligers and tigons (18).

Scientific taxonomies are useful simplifying devices. Information not contained in a taxonomic category is minimized or ignored – and confusing variation is thus reduced. In obtaining a basic scientific education in chemistry and biology, students are taught simplifying taxonomies. Such taxonomies cohere with student's pre-existing essentialist assumptions and further reinforce those assumptions. These simplifying taxonomies are afterwards taken to be scientific ideals. When a domain such as psychiatry is subsequently encountered, attempts to taxonomize it are slotted into this customary framework and essentialist assumptions begin to function as a cognitive bias – an *essentialist bias*. Features that make all instances of a category the same are emphasized. Gaining expertise in a domain makes the variation within categories more noticeable, but the pull of essentialism in taxonomic thinking remains hard to resist.

## A NON-ESSENTIALIST VIEW OF NATURAL KINDS

In the 1970s some philosophers began to argue that species categories should not be viewed as natural kinds (19,20). To keep the concept of natural kind relevant for species taxonomies, R. Boyd extended its boundaries to encompass an alternative non-essentialist view called the *homeostatic property cluster* concept (21,22). According to this view, a natural kind represents a set of co-occurring features that reliably cluster together because of shared causal processes, but there is no set of features that all members of the natural kind must possess. For example, certain anatomical structures, body

type, and predatory behaviors form a homeostatic property cluster called tiger about which we can make inferences.

Such kinds are natural because they are produced by similarity-generating causal mechanisms (23). The relevant mechanisms that maintain the cluster as a whole may be internal (e.g., a genome), but they can also be external (e.g., availability of mates). Variations in the relevant causal process (a lion parent) will create variations in the outcome. As a result, there may be individuals who are subject to some but not all of the usual causal processes, and whether or not they should be considered members of the species is indeterminate.

Given the possibility of indeterminacy in classifying species, one should not be surprised that similar difficulties arise in the classification of psychiatric disorders (24). For instance, consider the difference between intense grief and mild depression. Although we can conceptually distinguish between the two, there are borderline cases that share some but not all features of both. In practice, making a differential diagnosis requires a judgement call. If clustering is imperfect due to variation in the causes, additional background considerations are needed to inform diagnostic decisions. For instance, a past history of depression might shift an indeterminate case in one direction or another.

Complicated cases also contradict the typical essentialist picture. Such cases can manifest symptoms from the depression, anxiety and somatic symptom clusters, the obsessive-compulsive spectrum, the domain of personality disorder, and occasionally the psychoses. The symptom configurations for these cases evolve over time, with certain symptoms coming into the foreground, and then receding into the background as other symptoms take their place. Interactions between symptoms can also generate new symptoms not on the usual criteria lists (25). Viewing a complicated symptom network as an assortment of distinct disorders is probably reifying ICD and DSM categories more than is justified.

Despite the availability of this liberalized view of natural kinds, it is likely that the simplifying assumptions of essentialism will continue to serve as aspirational ideals in psychiatric thinking. The ambition to definitely categorize what disorder a patient "really" has is stronger when essentialist assumptions are activated. Indeed, we can expect each new cohort of students to enter psychiatry with essentialist biases (26,27). As students are taught to think about patients in terms of psychiatric categories, they will be disposed to see the categories as more invariant across cases than they actually may be, and to assign extra "metaphysical" relevance to hidden causal properties.

The homeostatic property cluster model, however, better coheres with clinical expertise and deserves to be actively promulgated in psychiatric education as an alternative to our instinctive essentialism. The task is not hopeless. Psychological/mental concepts are typically less subject to essentializing than biological concepts (28) and essentialist inferences about taxonomic categories can be attenuated with clinical experience (29). With attention to these issues during training, professionals might be less likely to become cynical about classification after clinical experience makes the inadequacies of essentialist expectations more evident.

## PRACTICAL KINDS AND TAXONOMIC DEVELOPMENT

Natural kind concepts are supposed to represent what exists independent of our classifications, but in application, concepts for disorders become subject to our goals and interests. The clinical goals of practitioners and patients, the various scientific goals of researchers, philosophical theories about the nature of disorders, the priorities of health service administrators and social policy analysts, and commercial interests, for better or worse, have all played a role in how constructs for psychiatric disorders are developed. No one would consider this situation scientifically ideal, but the complexity of psychiatric phenomena makes it hard to avoid.

When the development of a classification requires a balance between competing background assumptions and goals, psychiatric constructs are better thought of as *practical kinds*. The homeostatic property cluster model recognizes situations where classification can be indeterminate as exceptions to a rule, but says little about the role of background assumptions and goals in selecting "good" classifications. In psychiatry, indeterminacy is more than an occasional exception. It exists at the boundary of the normal and the abnormal, and between conventionally recognized symptom configurations and a more extensive, interconnected symptom space (30,31). Such is the inspiration behind the claim that psychiatric disorders are practical kinds.

What have philosophers learned about the kinds that should be taken note of in psychiatry? H. Putnam has observed that to ask whether kinds are made by the world or made by us is too black-and-white a question (32). As tools that we use in our work, concepts are what Locke called the workmanship of human understanding. Concepts for psychiatric disorders are constituted by discoveries *and* decisions. There is an interaction between what the world produces and what we find useful to notice. The *concept* of natural kind orients us to regularities in psychiatric phenomena that exist irrespective of our wishes or preferences; they are the result of causal processes that scientists seek to discover. The *concept* of practical kind orients us to the variety of the decisions we make in order to classify an indeterminate world.

### Acknowledgement

The author thanks A. Kostko, S. Waterman, S. Steingard and A. Solomon for helpful comments.

## References


1. Wakefield JC. The myth of open concepts: Meehl's analysis of construct meaning versus black box essentialism. Appl Prevent Psychol 2004;11:77-82.
2. Hyman SE. Diagnosis of mental disorders in light of modern genetics. In: Regier DA, Narrow WE, Kuhl EA et al (eds). The conceptual evolution of DSM-5. Arlington: American Psychiatric Publishing, 2011:3-17.
3. McNally RJ. The ontology of posttraumatic stress disorder: natural kind, social construction, or causal system? Clin Psychol Sci Pract 2012;19:220-8.
4. Paris J, Phillips J. Making the DSM-5: concepts and controversies. New York: Springer, 2013.
5. Kripke S. Naming and necessity. Dordrecht: Reigel, 1972.
6. Hacking I. Natural kinds: rosy dawn, scholastic twilight. R Inst Philos Suppl 2007;82:203-39.
7. Cooper R. Natural kinds. In: Fulford KWM, Davies M, Gipps R et al (eds). The Oxford handbook of philosophy and psychiatry. Oxford: Oxford University Press, 2013:960-5.
8. Meehl PE. Bootstraps taxometrics: solving the classification problem in psychopathology. Am Psychol 1995;50:266-75.
9. Hyman SE. The diagnosis of mental disorders: the problem of reification. Annu Rev Clin Psychol 2010;6:155-79.
10. Gelman SA, Hirschfeld LA. How biological is essentialism. In: Medin DL, Atran S (eds). Folkbiology. Cambridge: MIT Press, 1999:403-46.
11. Gelman SA. Psychological essentialism in children. Trends Cogn Sci 2004;8:404-9.
12. Sobel DM, Yoachim CM, Gopnik A et al. The blicket within: preschoolers' inferences about insides and essences. J Cogn Devel 2007;8:159-82.
13. Hume D. A treatise of human nature. Oxford: Oxford University Press, 2000/1739.
14. Mill JS. A system of logic. Toronto: University of Toronto Press, 1973/1843.
15. Popper K. Conjectures and refutations: the growth of scientific knowledge. London: Routledge, 1963.
16. Meynell GG. John Locke and the preface to Thomas Sydenham's Observationes medicae. Med History 2006;50:93-110.
17. Gaukroger S. The collapse of modernism and the rise of sensibility. Oxford: Oxford University Press, 2010.
18. Short RV. An introduction to mammalian interspecific hybrids. J Heredity 1997;88:355-7.
19. Ghiselin MT. A radical solution to species problem. Syst Zool 1974;23:536-44.
20. Hull DL. Are species really individuals? Syst Zool 1976;25:174-91.
21. Boyd R. What realism implies and what it does not. Dialectica 1989;43:5-29.
22. Boyd R. Realism, anti-foundationalism and the enthusiasm for natural kinds. Philos Stud 1991;61:127-48.
23. Craver CF. Mechanisms and natural kinds. Philos Psychol 2009; 22:575-94.
24. Kendler KS, Zachar P, Craver C. What kinds of things are psychiatric disorders. Psychol Med 2011;41:1143-50.
25. Keeley JW, Chmielewski MS, Bagby RM. Interaction effects in comorbid psychopathology. Compr Psychiatry 2015;60:35-9.
26. Haslam N, Ernst D. Essentialist beliefs about mental disorders. J Soc Clin Psychol 2002;21:628-44.
27. Adriaens PR, De Block A. Why we essentialize mental disorders. J Med Philos 2013;38:107-27.
28. Heyman GD, Gelman SA. Beliefs about the origins of human psychological traits. Dev Psychol 2000;36:663-78.
29. Ahn W-K, Flanagan EH, Marsh JK et al. Beliefs about essences and the reality of mental disorders. Psychol Sci 2006;17:759-66.
30. Zachar P, McNally RJ. Psychiatric classification and the sorites problem: the case of PTSD. In: Keil G, Keuck L, Hauswald R (eds). Gradualist approaches to health and disease. Oxford: Oxford University Press (in press).
31. Zachar P. A metaphysics of psychopathology. Cambridge: MIT Press, 2014.
32. Putnam H. Realism with a human face. Cambridge: Harvard University Press, 1990.


DOI 10.1002/wps.20240

# What have we learned from the Psychiatric Genomics Consortium

MICHAEL C. O'DONOVAN

Medical Research Council Centre for Psychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University School of Medicine, Cardiff CF24 4HQ, UK

Decades of research in the pre-molecular genetics era firmly established that major psychiatric disorders are highly to moderately heritable, but only with the emergence of molecular genetic technology about 35 years ago was it possible to envisage identifying the specific pathogenic genes responsible.

In psychiatry, the opportunity to probe pathophysiology using DNA seemed particularly attractive given that other biomedical approaches had been frustrated by the complexity of the brain, the challenges in obtaining access to fresh tissue, and the extensive potential for reverse causal associations due to the many environmental and behavioural consequences of the disorders.

Simple in concept, translating heritability to pathophysiology has proven arduous. This paper provides a perspective on that process, how some of the obstacles have been overcome, and some of the implications of the current findings. The focus is the work of the Psychiatric Genomics Consortium (PGC), whose main impacts relate to common rather than rare genetic variation. This reflects the data available rather than any ideological position that rare genetic variation is unimportant in psychiatry.

## FROM MENDELIAN TO MULTIFACTORIAL POLYGENIC INHERITANCE

Early studies were predicated on the hypothesis of Mendelian transmission, where high penetrance mutations are sufficient to cause disease. With the exception of neurodegenerative diseases and some forms of autism, this proved a dead end in psychiatry. Although the possibility that Mendelian alleles act in a small proportion of cases cannot be excluded, the vast majority of psychiatric illness do not conform to this mode of inheritance.

Accordingly, the focus switched to oligogenic, polygenic or multifactorial threshold models and the concept of "susceptibility alleles" which only modestly increase liability of disorder. How modestly was ultimately laid bare by studies using genome-wide association study (GWAS) technology, notably that of the Wellcome Trust Case Control Consortium (1), but also early studies of schizophrenia (2) and bipolar disorder (3).

The conclusions were that common risk alleles typically confer effects with odds ratios (OR) less than 1.1 and that the sample sizes required to detect them were beyond those available to individual groups, or even existing psychiatric consortia. These considerations led to the formation of the Psychiatric Genome Wide Association Consortium (4), now known as the Psychiatric Genomics Consortium (PGC).

## THE PSYCHIATRIC GENOMICS CONSORTIUM

Initially focussing on attention-deficit/hyperactivity disorder (ADHD), autism, bipolar disorder, major depression and schizophrenia, the PGC has expanded to include anorexia nervosa, obsessive-compulsive disorder/Tourette syndrome, post-traumatic stress disorders, and substance use disorder. With a dynamic membership currently comprising over 800 investigators from 36 countries, the PGC actively welcomes additional investigators (see http://pgc.unc.edu).

From the perspective of genome-wide significant results, schizophrenia has enjoyed the greatest success (5), followed by bipolar disorder (6). ADHD, major depressive disorder, and autism spectrum disorder (ASD) are yet to leave the starting blocks, but studies by the PGC (and others) have shown that common risk alleles do indeed contribute to these disorders (7,8) and success likely reflects the relative sample sizes studied.

The impact of numbers is clear. In 2011, with a schizophrenia discovery sample of 9,394 cases, the PGC reported only five novel findings (9), yet within three years, data from around 35,500 cases resulted in 128 independent genetic associations (5). Overall, the pattern was of minimal progress until a breakthrough threshold of about 13,000 cases was attained, after which the rate of new findings increased rapidly by about four independent associations per 1,000 new cases.

Published sample sizes for the other founder PGC phenotypes are still below the breakthrough point for schizophrenia (for ADHD and ASD, less than 5000 cases; for major depressive disorder, less than 10,000), but inspired by schizophrenia, equivalent (or larger) samples will be available in the next couple of years.

As yet unknown differences in the genetic architectures between disorders may mean that both the breakthrough threshold and subsequent ratio of discovery to sample size may differ across disorders. In particular, for major depressive disorder, early findings suggest that the population variance contributed by each individual allele may be particularly small, and that alternative approaches may be required for defining more homogeneous – and heritable – phenotypes (10).

## PLEIOTROPY

Pleiotropy denotes the influence of a genetic variant on multiple apparently unrelated phenotypes. Observed in pre-PGC GWAS studies (2,11), this phenomenon has been more fully explored by the PGC. Using novel methods that allow patterns of allele sharing across disorders to be estimated at a genome-wide level, and the degree of shared genetic risk to be quantified, the Cross Disorder Group of the PGC reported substantial overlaps between common alleles influencing risk of schizophrenia and bipolar disorder, and between those influencing risk of major depressive disorder and each of schizophrenia, bipolar disorder and, most surprising of all, ADHD (7).

These findings complement studies of rare genetic variation showing that identical rare mutations can increase risk of schizophrenia, ASD, intellectual disability, and ADHD (12,13). Moreover, the recent PGC schizophrenia study (5) found that loci defined by common allelic associations were enriched for genes carrying rare mutations in intellectual disability and autism.

Thus, pleiotropic effects in psychiatry appear to be the rule rather than the exception. An alternative view is that apparent pleiotropy merely reflects deficiencies in the pathophysiological validity of our classification system and that the distinctive phenomenological states enshrined in categorical diagnosis do not define discrete pathophysiological disorders (see 14).

Together with analogous findings in non-genetic research, pleiotropy has provided much of the impetus for calls for psychiatric research to move beyond diagnostic categories and consider alternative measures such as domains of psychopathology, or other non-clinical features (e.g., cognitive measures), that might map better onto underlying biology (14).

## PATHOPHYSIOLOGY AND THERAPEUTICS

Genetics is still short of delivering clear insights into disease mechanisms. While each of the 128 independent genetic associations in schizophrenia have the potential to generate new insights into the disorder, achieving this requires associations to be linked to changes in the function of specific genes, a step not yet unequivocally taken for any common variant association. Nevertheless, some general clues about disorder-related biology are emerging, particularly when the common variant work of the PGC is considered together with findings from studies of rare genetic variation.

At the most general level, schizophrenia associations are enriched at elements that regulate gene expression in brain, and possibly immune tissues (5). That schizophrenia is emerging as (largely) a brain disorder is in one sense trivial, but in the context of historically highly polarized opinions about its origins, such empirical findings are important. More specifically, there is increasing evidence that common (5) and rare variant associations (15-17) in schizophrenia show a tendency to converge upon genes encoding functionally related proteins, for example multiple calcium channels and post-synaptic proteins complexes of glutamatergic synapses, including NMDA, AMPA and metabotropic receptors.

In other disorders, the data are sparse, and the patterns less clear. Nevertheless, in bipolar disorder as in schizophrenia, the findings point to perturbation of function at calcium channels (6). PGC studies exploiting the genetic correlation between schizophrenia, major depressive disorder and bipolar disorder further suggest a shared involvement across these disorders for genes implicated in histone methylation, a process involved in regulating gene expression, and in immune pathways (18).

These and other findings are finally allowing the development of new molecular models of psychiatric pathophysiology based upon, for example, synaptic plasticity (19). The models are crude and yet to be tested experimentally, and even if they do reflect aspects of pathophysiology, they are unlikely to represent the full picture. Nevertheless, the findings suggest that continued acquisition of genetic data will provide increasingly better insights into novel disease mechanisms, and in doing so, novel therapeutic options.

This journey is long haul, but it has been argued (20) that associations in schizophrenia spanning genes encoding the dopamine D2 receptor (the target for all known effective antipsychotic drugs) and a number of glutamate receptors (existing targets of interest among pharmaceutical companies) suggest that other genes within GWAS associated regions may provide fast-track targets for developing treatments. One example is the kindling of interest in the application of calcium channel blockers in bipolar disorder.

## RISK PROFILES AND PATIENT STRATIFICATION

Using approaches introduced by the International Schizophrenia Consortium (11), the most recent PGC study (5) calculated that, in schizophrenia, a composite genetic risk profile score derived from all independent nominally significant associated alleles ($p < 0.05$) captures about 7% of total liability for the disorder in people of European ancestry, though somewhat less in people of non-European ancestry. Viewed from the perspective of effect size, those in the top decile of risk profile scores had approximately 4-5 fold higher risk than average. This degree of risk prediction is not clinically useful but, as more genetic variance is captured by larger studies, and risk profile scores are perhaps combined with other forms of data, it may become so in the future.

Beyond risk prediction, the potential applications for using risk profile scores to stratifying patients are extensive. By way of illustration, studies are underway to test the possibility that high risk profile scores might predict chronicity or treatment resistance and the need for early introduction of clozapine.

The availability of risk profile scores as a marker of trait liability is also providing a new, and increasing widely used,

research tool. Applications include selecting individuals on the basis of those scores rather than affected status to investigate the neurobiological basis of schizophrenia liability. Furthermore, researchers with a developmental perspective have initiated epidemiological studies of children aiming to identify the cognitive and behavioural correlates of genetic risk that predate clinical disorders, and may even mediate the link between risk and disorder and be open to therapeutic intervention.

## CONCLUSIONS

Recent years have brought considerable advances in psychiatric genetics and, in the arena of common genetic variation, the PGC is now the major driving force. Funding aside, future success is critically dependent upon the continued donation of biological samples from individuals (almost half a million have done so already) and on the willingness of more researchers to contribute what is often their life's work of data acquisition, frequently in the face of a perception of risk to self-interest.

If this continues, given an increasingly global reach and expanding membership and sample base, there are reasons to expect that progress in schizophrenia will accelerate, and that the momentum achieved by that disorder will transmit to the full spectrum of psychiatric disorders. In doing so, we believe that the discipline will justify the faith the early genetics pioneers placed in it to provide the fundamental insights into aetiology that will fuel the accelerating phase of mechanistic research marking progress in other areas of medicine, thus potentially transforming the outlook for patients with these disorders.


## Acknowledgements

The author is funded by Medical Research Council (MRC) Centre (G0800509) and Program (G0801418) grants, and the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 279227.


## References


1. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661-78.
2. O'Donovan MC, Craddock N, Norton N et al. Identification of loci associated with schizophrenia by genome-wide association and follow-up. Nat Genet 2008;40:1053-5.
3. Ferreira MA, O'Donovan MC, Meng YA et al. Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. Nat Genet 2008;40:1056-8.
4. Sullivan PF. The psychiatric GWAS consortium: big science comes to psychiatry. Neuron 2010;68:182-6.
5. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature 2014;511:421-7.
6. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nat Genet 2011;43:977-83.
7. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet 2013;45:984-94.
8. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet 2013;381:1371-9.
9. Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. Genome-wide association study identifies five new schizophrenia loci. Nat Genet 2011;43:969-76.
10. Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium. A mega-analysis of genome-wide association studies for major depressive disorder. Mol Psychiatry 2013;18:497-511.
11. International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 2009;460:748-52.
12. Kirov G, Rees E, Walters JT et al. The penetrance of copy number variations for schizophrenia and developmental delay. Biol Psychiatry 2014;75:378-85.
13. Malhotra D, Sebat J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. Cell 2012;148:1223-41.
14. Owen MJ. New approaches to psychiatric diagnostic classification. Neuron 2014;84:564-71.
15. Kirov G, Pocklington AJ, Holmans P et al. De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. Mol Psychiatry 2012;17:142-53.
16. Fromer M, Pocklington AJ, Kavanagh DH et al. De novo mutations in schizophrenia implicate synaptic networks. Nature 2014;506:179-84.
17. Purcell SM, Moran JL, Fromer M et al. A polygenic burden of rare disruptive mutations in schizophrenia. Nature 2014;506:185-90.
18. Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. Nat Neurosci 2015;18:199-209.
19. Hall J, Trent S, Thomas KL et al. Genetic risk for schizophrenia: convergence on synaptic pathways involved in plasticity. Biol Psychiatry 2015;77:52-8.
20. Lencz T, Malhotra AK. Targeting the schizophrenia genome: a fast track strategy from GWAS to clinic. Mol Psychiatry 2015;20:820-6.


DOI 10.1002/wps.20270

# Antidepressants versus placebo in major depression: an overview

Arif Khan[1,2], Walter A. Brown[3]

[1]Northwest Clinical Research Center, Bellevue, WA, USA; [2]Department of Psychiatry and Behavioral Science, Duke University School of Medicine, Durham, NC, USA; [3]Department of Psychiatry and Human Behavior, Brown University, Providence, RI, USA

*Although the early antidepressant trials which included severely ill and hospitalized patients showed substantial drug-placebo differences, these robust differences have not held up in the trials of the past couple of decades, whether sponsored by pharmaceutical companies or non-profit agencies. This narrowing of the drug-placebo difference has been attributed to a number of changes in the conduct of clinical trials. First, the advent of DSM-III and the broadening of the definition of major depression have led to the inclusion of mildly to moderately ill patients into antidepressant trials. These patients may experience a smaller magnitude of antidepressant-placebo differences. Second, drug development regulators, such as the U.S. Food and Drug Administration and the European Medicines Agency, have had a significant, albeit underappreciated, role in determining how modern antidepressant clinical trials are designed and conducted. Their concerns about possible false positive results have led to trial designs that are poor, difficult to conduct, and complicated to analyze. Attempts at better design and patient selection for antidepressant trials have not yielded the expected results. As of now, antidepressant clinical trials have an effect size of 0.30, which, although similar to the effects of treatments for many other chronic illnesses, such as hypertension, asthma and diabetes, is less than impressive.*

**Key words:** Major depression, antidepressants, placebo, clinical trials, expectation bias, drug development regulators

***(World Psychiatry 2015;14:294-300)***

Twenty years ago we believed that antidepressants worked in 70% of depressed patients and placebo in 30% of them, as stated in the U.S. Department of Health and Human Services report on treatment of major depression (1). This notion, however, has undergone a major revision in the past two decades.

Kuhn's original report describing the therapeutic effects of imipramine was based on clinical vignettes (2). As is the case with most disorders, it was evident even in this first report that not all depressed patients responded to the new drug. Kuhn pointed out that patients with endogenous or vital depression were most likely to respond.

A considerable body of research subsequently explored which depressed patients responded to select antidepressants such as imipramine and phenelzine compared to electroconvulsive therapy (ECT) (e.g., 3). As part of this development, the need arose to quantify the depressive syndrome, and pioneers like M. Hamilton designed depression rating scales (4).

In the U.S., Klerman and Cole produced a detailed review of trials evaluating the effectiveness of imipramine (5). Consistent with Kuhn's findings, they reported that hospitalized depressed patients with a melancholic pattern of symptoms were most likely to respond to the drug. Much of the wisdom about the magnitude of antidepressant and placebo response was based on these early clinical trials of tricyclic antidepressants, and these data carried well into the early 1990s (6).

However, in the 1970s and 1980s, several important changes were occurring in psychiatry. Most significant was the advent of DSM-III. Using an atheoretical approach, this diagnostic system minimized differences between subtypes of depression and conceptualized a broad syndrome called major depressive disorder, characterized by a single or recurrent bouts named major depressive episodes. Such a "uniform" diagnosis now included millions of patients and became an attractive target for the pharmaceutical industry. Thus, the American Psychiatric Association, sponsor of the DSM-III, unintentionally expanded the market for antidepressants.

Not surprisingly, a plethora of new drugs were developed, and almost all of the trials for the new compounds, such as selective serotonin reuptake inhibitors (SSRIs) and serotonin and norepinephrine reuptake inhibitors (SNRIs), included depressed patients meeting the DSM-III generic criteria for "major depressive episode". Some attempts were made to recruit in antidepressant trials the more classical "endogenous" or "melancholic" subtypes of patients. However, these attempts were often half-hearted and criteria were not always strictly followed.

So, when we accessed the public domain data from the U.S. Food and Drug Administration (FDA) archives for the antidepressants approved between 1985 and 1997 (7), it quickly became apparent that many of the assumptions about the relative potency of antidepressants compared to placebo were not based on data from the contemporary trials but from an earlier era. Specifically, it became evident that the magnitude of symptom reduction was about 40% with antidepressants and about 30% with placebo.

The U.S. FDA public domain reports used symptom reduction as a measure of improvement and did not include therapeutic response rates. Even with this caveat, however, it was evident that the conventional wisdom of 70% response with antidepressants was at best an overestimate.

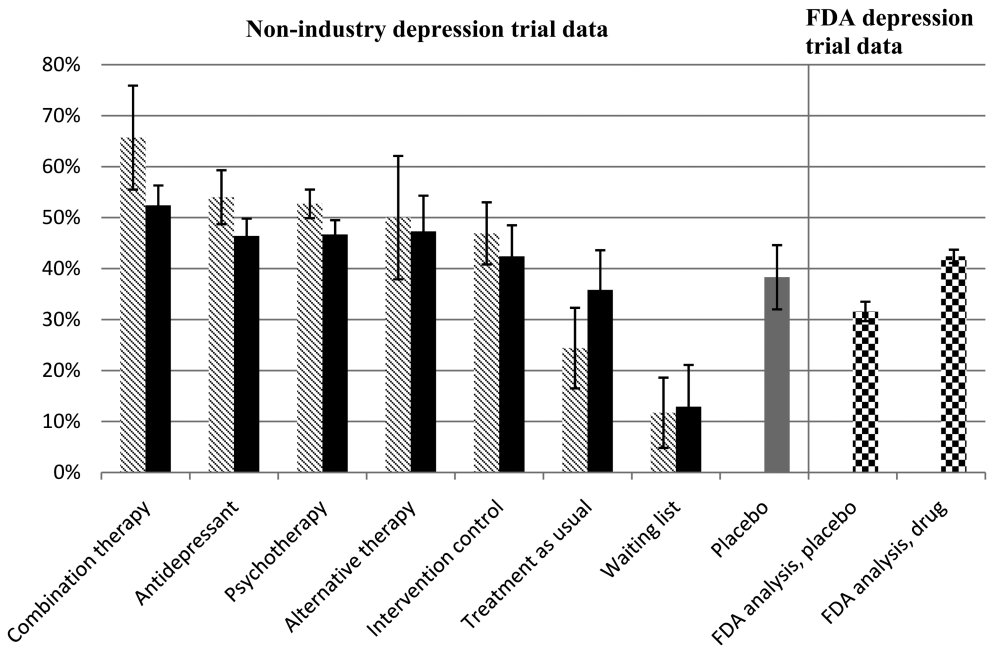


**Figure 1** Mean percentage symptom reduction in unblinded and blinded treatment arms from published depression trials compared to data from pivotal registration depression trials as reported by the U.S. Food and Drug Administration (FDA) (adapted from 13). Striped bars represent unblinded trial arms; black bars represent blinded trial arms; the grey bar represents placebo control arms from published non-registration trials; checkered bars represent data from pivotal registration trials. The mean percentage symptom reduction was weighted by the number of assigned patients. Error bars represent 95% confidence intervals

Not surprisingly, Walsh et al (8) also noted that the magnitude of symptom reduction with placebo had been increasing in the past three decades, based on an analysis of published antidepressant clinical reports. This publication prompted considerable attention and speculation from a number of investigators.

The effectiveness of modern antidepressants was not only questioned by placebo-controlled clinical trials, but also by trials based on a clinical practice model that did not include placebo. An experiment about antidepressant effectiveness started by J. Rush in Texas became a large scale national effort, supported by the U.S. National Institute of Mental Health, known as the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) project (9). This project showed that antidepressants such as citalopram led to a therapeutic response in only about 4 out of 10 depressed outpatients.

These challenges to the assumptions about the effectiveness of antidepressants brought about close scrutiny of the clinical trial data provided by the pharmaceutical companies, since the development, manufacturing and marketing of antidepressants is obviously a commercial venture. Specifically, criticism was raised by both academics and the general public as to the integrity of antidepressant clinical data generated by the industry (e.g., 10,11).

As a reaction, JAMA Network editors refused to accept data analyses completed by pharmaceutical companies (12). Instead, they insisted that they would only consider industry papers for publication if the original clinical trial data were independently reviewed by academic, non-industry statisticians.

## THE IMPACT OF EXPECTATION BIAS

Given such an acrimonious situation with potential conflicts, we compared depression clinical trial data from non-pharmaceutical industry sources to antidepressant clinical data from the FDA Freedom of Information Act (FOIA) sources. In this analysis, we evaluated the magnitude of symptom reduction with all of the acknowledged depression treatments as well as their active or passive controls, including placebo (13).

This rather complex set of data, illustrated in Figure 1, contains several significant findings. On the left side of the figure (striped bars, black bars, grey bar) are data from non-pharmaceutical company sources, and on the right side (checkered bars) are data based on FDA reports. The magnitude of symptom reduction with placebo pill is higher in the non-pharmaceutical industry depression trials (grey bar) compared to pivotal registration trials (checkered bar).

Of even more interest is the pattern of response among the non-pharmaceutical industry double- or triple-blinded depression trials. The striped bars indicate the magnitude of depressive symptom reduction in trials where the investigators and their staff were aware of the design and expectations of the study. The black bars indicate the magnitude of symptom reduction when the investigators and raters were "blinded" to the design and execution of the study.

Clearly, investigator and rater bias influences the magnitude of symptom reduction with all treatments, whether they are approved treatments, active controls, passive controls, sham

treatments, treatment as usual, waiting list, or placebo. For example, the magnitude of symptom reduction where the design of the trials was known to the investigators and raters (striped bars) followed the pattern of accepted expectations. The combined pharmacotherapy and psychotherapy had the best outcome, followed by antidepressants alone, known forms of psychotherapy (e.g., cognitive behavioral therapy), alternative therapies such as acupuncture or exercise, intervention controls (e.g., sham acupuncture, control psychotherapies such as educational sessions), and placebo. Not unexpectedly, "treatment as usual" fared worse than placebo, and waiting list had the smallest improvement.

On the other hand, the pattern was quite different if the investigators and their staff were blinded to the design and execution of the trials. Under these circumstances, the symptom reduction with each treatment was of smaller magnitude and the differences among the various treatments and controls were also smaller. The depressed patients assigned to all the treatments (active or control) – antidepressants, psychotherapy, acupuncture, exercise, sham acupuncture, sham psychotherapy and "treatment as usual" – experienced a symptom reduction that was comparable to that observed with placebo. In other words, when the level of blinding was high and it was difficult for the investigators, their staff and depressed patients to guess treatment assignment, the differences between these treatments, controls and placebo became quite small.

For two of the treatment paradigms, it is difficult to blind both clinicians and patients completely. One of these paradigms is combined pharmacotherapy and psychotherapy and the other is the waiting list. Not surprisingly, the magnitude of symptom reduction compared to placebo was significantly different using these two paradigms. The combined pharmacotherapy and psychotherapy showed a superior treatment response compared to placebo, and the waiting list an inferior treatment response than placebo. Simply put, clinicians and depressed patients continued to fulfill expectations of each treatment based on prior assumptions.

The effect of expectation bias is well illustrated by other experimental data. Sinyor et al (14) reported that, if there was no placebo control in an antidepressant trial comparing two antidepressants, the magnitude of symptom reduction was 65.7%. If the trial included two antidepressant treatments and one placebo arm (33% placebo exposure risk), the magnitude of symptom reduction with the antidepressants was 57.7%, while that with placebo was 44.6%. If the antidepressant trial included one antidepressant arm and one placebo arm (50% placebo exposure risk), the magnitude of symptom reduction with antidepressant was 51.7% and that with placebo was 34.3%.

In short, the apparent therapeutic effect of antidepressants is related to the risk of exposure to placebo, when this is known to clinicians and depressed patients from the consent form. If you lower the risk of exposure to placebo, then the apparent therapeutic effect with the antidepressants and placebo is greater.

These data from antidepressant clinical trials are applicable to clinical practice. First and foremost, it is critical to note that patients with mild to moderate depression are prone to nonspecific therapeutic effects. The comments made by Brown (15) regarding the experience of patients assigned to placebo are pertinent. He states: "The capsule they receive is pharmacologically inert, but hardly inert with respect to its symbolic value and its power as a conditioned stimulus. In addition, placebo-treated patients receive all the components of the treatment situation common to any treatment, i.e., a thorough evaluation; an explanation for distress; an expert healer: a plausible treatment; a healer's commitment, enthusiasm, and positive regard; an opportunity to verbalize their distress".

Indeed, Frank has argued that these elements of the treatment situation are the active ingredients of all the psychotherapies (16). Since antidepressant clinical trials involve extensive evaluations, long visits, many experts and "new and exotic treatments", it is not surprising that, under such conditions, the differences between active treatments and inactive treatments including sham acupuncture and placebo are, at best, small.

Although considerable attention has been paid to the magnitude of placebo response in depression and the small antidepressant-placebo differences, this phenomenon is not unique to depressive disorders. Illnesses that are chronic, have a fluctuating course and are associated with subjective distress are prone to placebo response. The following are some disorders that show the same pattern as depression.

Among patients with irritable bowel syndrome, treatment response occurs in 56% of cases, whereas the response rate to placebo is 46% (17). Thirty-six percent of patients with ulcerative colitis experience a therapeutic response with 5-aminosalicylic acid, whereas the response rate among those assigned to placebo is 20% (18).

A therapeutic response to one of six different anti-hypertensive agents was observed in 58% of patients with hypertension, while the response rate with placebo was 30% (19). The magnitude of change in one-second forced expiratory volume was 7% with bronchodilators compared to 4% with placebo (20).

In patients with osteoarthritis, the frequency of therapeutic response measured after arthroscopic lavage and debridement is lower than with sham procedures (21). Parkinson's disease patients are also prone to placebo response: the reduction of symptoms with selegiline is 12%, while that with placebo is 10% (22).

Lastly, non-pharmacological somatic treatments for depression such as ECT and vagal nerve stimulation (VNS), under controlled clinical trial conditions, also show the same pattern. For example, sham ECT can result in 30% of severely depressed patients experiencing a therapeutic effect (23). Similarly, the implant of an "inactive" VNS pacemaker results in a 10% treatment response, while the

response rate to "active" VNS is 15% among patients with chronic and treatment resistant depression (24).

These data clearly suggest that a high magnitude of placebo response is not unique to depressed patients, but inherent in an experimental paradigm. Thirty years ago, Quitkin et al (25) noted that the placebo response has an early onset and a fluctuating course, and it was assumed that depressed patients who respond to placebo relapse quickly back into depression. However, there is now evidence that, once patients respond to placebo, they remain well for a considerable period of time.

In a select sample of nine antidepressant trials, depressed patients who responded to either the investigational antidepressant or placebo during the double-blind trial continued on the same treatment assignment for six months or longer (26). Seventy-nine percent (333/420) of the depressed patients assigned to placebo did not relapse, compared to 93% (1074/1154) of the depressed patients assigned to antidepressants. In other words, four out of five depressed patients who improved with placebo remained well without relapse for six months or longer.

Mayberg et al (27) noted that clinical improvement with either fluoxetine or placebo was associated with cerebral glucose metabolism increases in depressed patients. Such a potential biological basis for placebo response is further supported by similar studies in pain and Parkinson's disease (28).

In summary, depressed patients are prone to non-specific treatment effects, in particular when receiving placebo. Expectations by both patients and clinicians play a significant role in the magnitude of treatment effects in depression clinical trials. Once set, placebo response tends to persist and there are sufficient data to suggest that this is associated to changes in brain glucose metabolism.

## THE IMPACT OF REGULATORY DECISIONS

The decisions of the FDA have strongly influenced what has happened to the design and execution of antidepressant trials in the past three decades. It is important to note that some of these decisions were taken by the regulators based on their assessment of prevailing wisdom and knowledge. Their ultimate intent was to reassure themselves and the public that pharmaceutical companies must demonstrate that their antidepressant is consistently superior to placebo before the drug is approved for marketing. This is part of the public health mandate being enforced by the FDA.

Although many of these regulatory decisions have a major impact on the design, execution and interpretation of antidepressant clinical trials, this fact is not well understood. As an illustration, although the concept of therapeutic response is easy to grasp, the FDA staff has never accepted this as a valid method. Actually, the counting of the number of depressed patients who responded versus those who did not was abandoned by the FDA after the approval of the antidepressant amitriptyline (29).

The rationale is as follows. A single measure may focus on factors that are not related to the specific disorder. As an example, opiates may produce a sense of well-being and "be therapeutic" globally for patients with malignancies, but they have little or no effect on the disease itself. Thus, the documentation of the impact of a drug on a disorder such as depression, as defined by the prevailing wisdom (in this instance, that of the DSM-III), requires a syndromal improvement, rather than a global feeling of well-being.

Hence, the FDA has considered rating scales such as the Hamilton Depression Rating Scale (HAM-D) (4) or the Montgomery-Asberg Depression Rating Scale (MADRS) (30) as surrogate markers to indicate a syndromal improvement for clinical depression. Interestingly, the FDA has accepted that the total score on these scales (that leads to a single number) is a valid method to assess improvement.

Not surprisingly, the variability produced in an antidepressant clinical trial is inherently influenced by this key decision. Specifically, the Clinical Global Impression (CGI) score can only be between 1 and 7, a rather narrow range, while the maximum total score for the HAM-D can be as high as 54 and as low as 0, and the maximum total score for the MADRS can be as high as 60 and as low as 0. This potential scatter has been seen by the FDA staff as an advantage in reducing the odds of a false positive result. However, not surprisingly, the use of CGI almost always leads to a better antidepressant-placebo separation.

Besides using that conservative outcome method, FDA also adopts very stringent criteria for data analyses. The FDA staff has considered the last observation carried forward (LOCF) method of analysis as the optimal one. In this model, if a depressed patient quits participating in a trial, the last known total HAM-D or MADRS score is replicated for the rest of the measurement points. Since the onset of response to placebo is early (31) and response to antidepressants occurs later, this acts to minimize antidepressant-placebo differences.

The FDA has recently accepted the concept of mixed-effect model repeated measure (MMRM) analysis, which consists of substituting missing data with a computational statistical model based on the overall pattern of the outcome measures. Although this method may be better than the LOCF (not proven yet), it is still mired in statistical concepts and not easily translated for interpretation, and certainly is not designed to favor outcome with antidepressants.

To complicate matters further, the European Medicines Agency (EMA) uses alternate models in evaluating new antidepressants. For example, it requires a relapse prevention model, in which depressed patients are treated with the new antidepressant and only those who respond to it are randomized into an experimental paradigm. In a double-blind manner, a segment of the responders continue to be assigned to the new antidepressant and another segment to placebo. Depressed patients are followed for approximately six months and the numbers of patients relapsing into another depressive episode testify to the effectiveness of the

new antidepressant compared to placebo. As a rule, the differences between the two groups are larger than in the acute, parallel design models. For example, Geddes et al (32) showed that the relapse rate using this model was 41% for depressed patients assigned to placebo compared to 18% for those assigned to antidepressants.

The FDA does not accept such models to approve a new antidepressant and thus pharmaceutical companies are forced to come up with multiple models that are not complementary and leave both the clinician and the researcher confused, making it difficult to transfer common sense ideas into clinical practice. Such a conundrum can be used in a masterly way by marketers or be cynically dismissed as a marketing ploy.

This major hobbling of antidepressant clinical trials and the fact that the results of these trials need to be interpreted with caution is neither appreciated nor heeded by researchers or clinicians (10), nor by the media, which need sensational stories (11).

As these data about the weaknesses of the double-blind placebo-controlled antidepressant trials were gathered, several attempts have been made to address this situation. As noticed in a recent review (33), the best way to show antidepressant-placebo differences is to reduce the number of investigative sites, say to ten to twelve. This fact is currently ignored, as most multicenter pharmaceutical industry antidepressant trials include an average of 60 sites, some studies going to 120 sites worldwide.

Another major factor, as we emphasized, is the placebo risk exposure. Simply put, a two treatment option (with a placebo risk of 50%) has the best chance of keeping placebo response to a minimum. However, most contemporary antidepressant clinical trials have a minimum of three treatment arms, with a considerable number having four or more. This approach is significantly influenced by regulatory agencies such as the FDA. Specifically, the FDA requires that trials attempt to show a dose-response relationship for new antidepressants. In other words, it requests the use of doses of the new antidepressant that may not be effective, so that the minimum effective dose can be identified. For example, a dose of 10 mg of fluoxetine has to be consistently shown not to be superior to placebo, so that FDA staff can consider the next higher dose being the possibly lowest effective one. Thus, several studies are conducted in futility that simply make the results of antidepressant trials look worse than they are. In this context, it is important to note that no clear dose-response relationship has been established to date for most of new antidepressants.

Another regulatory burden, although not universally required by the FDA, is the use of an active control – i.e., an approved antidepressant such as fluoxetine – to show what is technically termed "assay sensitivity". Such a paradigm is not only likely to increase the magnitude of placebo response, as the placebo exposure risk goes down, but also leads to many trials showing that the active comparator is not superior to placebo, adding more confusion.

The original concept (5) that more severely depressed patients respond better to antidepressants, whereas less severely depressed patients tend to respond to placebo, has held true in recent antidepressant trials (34,35). However, the implementation of this principle has not yielded any better results. Attempts at including patients who have a higher score on rating scales such as the HAM-D prospectively and prior to randomization has simply led to a greater magnitude of placebo response, although the factors behind such a phenomenon remain elusive (36).

Indeed, among the seven antidepressant trials where the severity of depression at baseline using HAM-D-17 was set at a score of 14 or higher, the magnitude of symptom reduction with placebo was 28.2%, while among the ten antidepressant trials where the threshold was set at 20, the magnitude of symptom reduction with placebo was 35.7%. Among the twenty antidepressant trials where the requested severity of depression at baseline using HAM-D-21 was 18 or higher, the magnitude of symptom reduction with placebo was 27.1%, while among the fourteen antidepressant trials where the threshold was set at 20, the magnitude of symptom reduction with placebo was 34.2%. These data raise the concern that forcing a higher pre-randomization severity may simply not work.

In this context, we have observed that using the longer version of HAM-D (21 items) results in a 60% increase in the antidepressant-placebo difference. Such a pattern has persisted over the past twenty-five years. It is possible that this version of HAM-D captures improvement in a larger group of depressive symptoms. Indeed, HAM-D-21 symptoms include diurnal variation in mood, paranoia and sense of hopelessness, reflecting additional dimensions of depression that may be more sensitive to antidepressant effects.

An alternative explanation of the high rate of placebo response when more severe patients are included may be that the investigative site staff rate patients as being more depressed than they actually are for commercial gain (38). However, attempts at having the patients evaluated by clinicians who do not stand to gain commercially by inflating the scores of rating scales, via videos or audiotapes, have not yielded the expected decrease in the magnitude of placebo response. Actually, they have produced an increase in that response and lower antidepressant-placebo differences (39,40).

In summary, given the constraints enforced by the regulators and the inability to control for multiple factors that may influence antidepressant clinical trial outcomes, it is more realistic to set up low expectations. In this context, a recent report by Gibertini et al (41) provides a useful model. These investigators analyzed the data from 81 monoaminergic antidepressant trials conducted in the past three decades, submitted to the FDA for the approval of fifteen antidepressants. They found an effect size of 0.30, which is considered modest.

This finding translates into the design of a prospective antidepressant trial as follows. This trial should compare a known effective dose of the test antidepressant to placebo (50% placebo risk), a two treatment option. Each treatment arm should consist of a minimum of 120 depressed patients and should be implemented at a maximum of twelve investigative sites. A specific successful example of the application of these principles has been the relatively quick and easy approval of vilazodone, based on two out of two positive trials (42).

In this context, it is important to note that clinical trials of medications for other common disorders, such as hypertension (43), asthma (44) and diabetes (45), have produced similar effect sizes, although attracting much less attention and criticism.

## CONCLUSIONS

Patients with a major depressive episode as defined by DSM-III, DSM-IV and DSM-5 are significantly prone to non-specific treatment effects. This applies to both industry and non-industry clinical trials. It is important to note that the high magnitude of response to placebo is not unique to depression, but common to other chronic illness associated with subjective distress. Increasing the blinding of clinicians conducting the antidepressant clinical trial may not result in a decrease in the magnitude of placebo response nor an increase in the magnitude of antidepressant response. In fact, it is likely to do just the opposite.

Drug development regulators such as the FDA and the EMA have a significant, albeit underappreciated, role in how modern antidepressant clinical trials are designed and conducted. Because of their concern about possible false positive results, these regulators require trials that may not have the best design and conduct. Interpretation of data from such trials is difficult and confusing.

Although there are known factors that may influence the outcome of antidepressant trials, taking these factors into account is not easy and is not routinely done. Attempts by researchers to select patients independently from site clinicians by video- or audiotaping have not yielded promising results.

The effect size of current antidepressant trials that include patients with major depressive episode is approximately 0.30 (modest), and this fact needs to be heeded for future antidepressant trials.

## References


1. U.S. Department of Health and Human Services. Depression in primary care: treatment of major depression. Vol. 2, Clinical Practice Guideline No. 5. Washington: U.S. Government Printing Office, 1993.
2. Kuhn R. The treatment of depressive states with G 22355 (imipramine hydrochloride). Am J Psychiatry 1958;115:459-64.
3. Clinical Psychiatry Committee. Clinical trial of the treatment of depressive illness: report to the Medical Research Council. BMJ 1965;1:881-6.
4. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960;23: 56-62.
5. Klerman GL, Cole JO. Clinical pharmacology of imipramine and related antidepressant compounds. Pharmacol Rev 1965;17: 101-41.
6. Brown WA, Khan A. Which depressed patients should receive antidepressants? CNS Drugs 1994;1:341-7.
7. Khan A, Warner H, Brown WA. Symptom reduction and suicide risk in patients treated with placebo in antidepressant clinical trials: an analysis of the FDA database. Arch Gen Psychiatry 2000;57:311-7.
8. Walsh BT, Seidman SN, Sysko R et al. Placebo response in studies of depression: variable, substantial, and growing. JAMA 2002;287:1840-7.
9. Rush AJ, Fava M, Wisniewski SR. Sequenced Treatment Alternatives to Relieve Depression (STAR*D): rationale and design. Control Clin Trials 2004;25: 119-42.
10. Blier P. Do antidepressants really work? J Psychiatry Neurosci 2008;33:89-90.
11. Begley S. Why antidepressants are no better than placebos. Newsweek, January 29, 2010.
12. DeAngelis CD, Fontanarosa PB. Impugning the integrity of medical science: the adverse effects of industry influence. JAMA 2008;299:1833-5.
13. Khan A, Faucett J, Lichtenberg P et al. A systematic review of comparative efficacy of treatments and controls for depression. PLoS One 2012;7:e41778.
14. Sinyor M, Levitt AJ, Cheung AH et al. Does inclusion of a placebo arm influence active antidepressant treatment in randomized controlled trials? Results from pooled and meta-analyses. J Clin Psychiatry 2010;71:270-9.
15. Brown WA. The placebo effect in clinical practice. Oxford: Oxford University Press, 2013.
16. Frank JD, Frank JB. Persuasion and healing: a comparative study of psychotherapy, 3rd ed. Baltimore: Johns Hopkins University Press, 1991.
17. Patel SM, Statson WB, Legedza A et al. The placebo effect in irritable bowel syndrome trials: a meta-analysis. Neurogastroenterol Motil 2005;17:332-40.
18. Sutherland LR, May GR, Shaffer EA. Sulfasalazine revisited: a meta-analysis of 5-aminosalicylic acid in the treatment of ulcerative colitis. Ann Intern Med 1993; 118:540-9.
19. Preston RA, Materson BJ, Reda DJ et al. Placebo-associated blood pressure response and adverse effects in the treatment of hypertension. Arch Int Med 2000;160:1449-54.
20. Joyce DP. The placebo effect in asthma drug therapy trials: a meta-analysis. J Asthma 2000;37:303-18.
21. Mosely JB, O'Malley K, Petersen NJ et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. N Engl J Med 2002;347:81-8.
22. Ondo WG, Sethi KD, Kricorian G. Selegiline orally disintegrating tablets in patients with Parkinson disease and "wearing off" symptoms. Clin Neuropharmacol 2007;30:295-300.
23. Pagnin D, de Quiroz V, Pini S et al. Efficacy of ECT in depression. J ECT 2004; 20:13-20.
24. Rush AJ, Marangell LB, Sackeim HA. Vagus nerve stimulation for treatment resistant depression: a randomized, controlled acute phase trial. Biol Psychiatry 2005;58:347-54.
25. Quitkin FM, Rabkin JG, Ross D et al. Identification of true drug response to antidepressants: use of pattern analysis. Arch Gen Psychiatry 1984;41:782-6.
26. Khan A, Redding N, Brown WA. The persistence of the placebo response in antidepressant clinical trials. J Psychiatr Res 2008;42:791-6.
27. Mayberg HS, Silva AJ, Brannan SK et al. The functional neuroanatomy of the placebo effect. Am J Psychiatry 2002;159:728-37.
28. Benedetti F, Mayberg HS, Wager TD et al. Neurobiological mechanisms of the placebo effect. J Neurosci 2005;25:10390-402.
29. Burt CG, Gordon WF, Holt NF et al. Amitriptyline in depressive states: a controlled trial. Br J Psychiatry 1962;108:711-30.
30. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry 1979;134:382-9.

31. Khan A, Cohen S, Dager S. Onset of response in relation to outcome in depressed outpatients with placebo and imipramine. J Affect Disord 1989;17:33-8.
32. Geddes JR, Carney SM, Davies C. Relapse prevention with antidepressant drug treatment in depressive disorders: a systematic review. Lancet 2003;361:653-61.
33. Rutherford BR, Roose SP. A model of placebo response in antidepressant clinical trials. Am J Psychiatry 2013;170:723-33.
34. Khan A, Leventhal RM, Khan S et al. Severity of depression and response to antidepressants and placebo: an analysis of the FDA database. J Clin Psychopharmacol 2002;22:40-5.
35. Khan A, Brodhead AE, Kolts RL et al. Severity of depressive symptoms and response to antidepressants and placebo in antidepressant clinical trials. J Psychiatr Res 2005;39:145-50.
36. Khan A, Schwartz K, Kolts RL et al. Relationship between depression severity entry criteria and antidepressant clinical trial outcomes. Biol Psychiatry 2007;62:65-71.
37. Khan A, Bhat A, Kolts R et al. Why has the antidepressant-placebo difference in antidepressant clinical trials diminished over the past three decades. CNS Neurosci Ther 2010;16:217-26.
38. Mundt JF, Greist JH, Jefferson JW et al. Is it easier to find what you are looking for if you think you know what it looks like? J Clin Psychopharmacol 2007;27:121-5.
39. Khan A, Faucett J, Brown W. Magnitude of placebo response and response variance in antidepressant clinical trials using structured, taped and appraised rater interviews compared to traditional rating interviews. J Psychiatr Res 2014;51:88-92.
40. Khan A, Faucett J, Brown W. Magnitude of change with antidepressants and placebo in antidepressant clinical trials using structured, taped and appraised rater interviews compared to traditional semi-structured interviews. Psychopharmacology 2014;231:4301-7.
41. Gibertini M, Nations KR, Whitaker JA. Obtained effect size as a function of sample size in approved antidepressants: a real-world illustration in support of better trial design. Int Clin Psychopharmacol 2012;27:100-6.
42. Khan A. Vilazodone, a novel dual-acting serotonergic antidepressant for managing major depression. Expert Opin Investig Drugs 2009;18:1-12.
43. Patel HC, Hayward C, Ozdemir BA et al. Magnitude of blood pressure reduction in the placebo arms of modern hypertension trials. Hypertension 2015;65:401-6.
44. Schilling JP, Khan SRF, Khan A. Outcome measures and serious asthma exacerbation in clinical trials of asthma controller medications. Ann Allergy Asthma Immunol 2012;108:448-53.
45. Shojania KG, Sumant RR, McDonald KM et al. Effects of quality improvement strategies for type II diabetes on glycemic control: a meta-regression analysis. JAMA 2006;296:427-40.

DOI 10.1002/wps.20241

# Clinical trial methodology and drug-placebo differences

**Irving Kirsch**
Harvard Medical School, Beth Israel Deaconess Medical Center, Boston, MA, USA

As Khan and Brown (1) correctly note, the magnitude of the placebo response in antidepressant trials has increased over the years. But it is not only the placebo response that has increased. So too has the response to antidepressants, a fact that has been widely ignored. In the Walsh et al meta-analysis (2), the correlation between the placebo response and year of publication was r = .45; that between the response to selective serotonin reuptake inhibitors (SSRIs) and year of publication was r = .47, and the difference between the two remained relatively constant.

What might account for the finding that the response to both placebo and antidepressants has increased over the years? One thing that it points to is that the drug effect and the placebo effect are probably additive. That is, the response to antidepressants comprises the effect of the drug and the response to placebo, so that when the placebo effect increases, so too does the response to the drug (3). What, then, is responsible for the increase in antidepressant and placebo responses over the years? It cannot be due to decreased baseline severity in more recent trials, because pharmaceutical companies abandoned including mildly and moderately depressed patients in efficacy trials after finding that these patients did not benefit beyond placebo (4). A more likely explanation is that marketing has led to increased public perception that antidepressants are effective, thus enhancing the placebo effect, and because the placebo effect is a component of the drug response, the latter also increased.

As Khan and Brown note, "if you lower the risk of exposure to placebo, then the apparent therapeutic effect with the antidepressants and placebo is greater". Once again, these data suggest additivity. Increasing expectancy of getting a drug rather than a placebo increases the response to the drug and the placebo (5-7).

According to Khan and Brown, the last observation carried forward (LOCF) method acts to minimize antidepressant-placebo differences. My colleagues and I made the same assumption when conducting our first meta-analysis of the trial data submitted to the U.S. Food and Drug Administration (FDA) (8). However, the data proved us wrong. LOCF analyses indicated greater drug-placebo differences than did the observed cases method, in which dropouts were excluded from the analyses.

Caution is needed when drawing conclusions from data indicating that higher baseline scores are associated with greater improvement in both drug and placebo arms. This is exactly what would be expected based on the statistical artifact of regression toward the mean. It is a very substantial effect that is often ignored. Elsewhere I have shown that when difference scores between two random variables are correlated with the score on one of them, the resulting correlation is about r = .70 (9). This is the chance standard against which one might judge an association between baseline severity and improvement scores.

Finally, a note on the "quick and easy approval" of vilazodone by the FDA is needed. Trovis Pharmaceuticals submitted seven clinical trials to the FDA. The first five showed negative results. However, in one of these, the company noted a non-significant trend toward superiority of vilazodone over placebo on the Montgomery-Asberg Depression Rating Scale (MADRS), but not on the Hamilton Rating Scale for Depression (HAM-D), which had been used in previous antidepressant approvals as the primary outcome for assessing efficacy. Trovis Pharmaceuticals was then allowed to designate the MADRS instead of the HAM-D as the primary outcome measure for what were subsequently considered the two pivotal trials. The difference between vilazodone and placebo on the HAM-D improvement was only 1.69 points on the two "pivotal" trials and, considering all seven trials, it was only 1.01. The FDA approved labeling states that "the efficacy of VIIBRYD was established in two 8-week, randomized, double-blind, placebo-controlled trials", but makes no mention of the five negative trials. In an internal memo dated May 4, 1998, P. Leber, writing in his capacity as Director of the FDA Division of Neuropharmacological Drug Products, stated his opinion that "labeling that selectively describes positive studies and excludes mention of negative ones can be viewed as being potentially 'false and misleading'" (10, p. 11).

Despite these minor qualifications, the points made by Khan and Brown are well taken. Drug-placebo differences are small in efficacy trials, and most of the response to antidepressants seems due to expectancy.

## References

1. Khan A, Brown WA. Antidepressants versus placebo in major depression: an overview. World Psychiatry 2015;14:294-300.
2. Walsh BT, Seidman SN, Sysko R et al. Placebo response in studies of major depression: variable, substantial, and growing. JAMA 2002;287:1840-7.
3. Kirsch I. Are drug and placebo effects in depression additive? Biol Psychiatry 2000; 47:733-5.
4. Kirsch I, Deacon BJ, Huedo-Medina TB et al. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. PLoS Med 2008;5:e45.
5. Sinyor M, Levitt AJ, Cheung AH et al. Does inclusion of a placebo arm influence

response to active antidepressant treatment in randomized controlled trials? Results from pooled and meta-analyses. J Clin Psychiatry 2010;71:270-9.
6. Rutherford BR, Sneed JR, Roose SP. Does study design influence outcome? Psychother Psychosom 2009;78:172-81.
7. Sneed JR, Rutherford BR, Rindskopf D et al. Design makes a difference: a meta-analysis of antidepressant response rates in placebo-controlled versus comparator trials in late-life depression. Am J Geriatr Psychiatry 2008;16:65-73.
8. Kirsch I, Moore TJ, Scoboria A et al. The emperor's new drugs: an analysis of antidepressant medication data submitted to the U.S. Food and Drug Administration. Prev Treat 2002;5:23.
9. Kirsch I. The placebo effect revisited: lessons learned to date. Complement Ther Med 2013;21:102-4.
10. Leber P. Approvable action on Forrest Laboratories, Inc. NDA 20-822 Celexa (citalopram HBr) for the management of depression. Memorandum. Washington: Department of Health and Human Services, Public Health Service, Food and Drug Administration, Center for Drug Evaluation and Research, 1998.

DOI 10.1002/wps.20242

# Antidepressants: misnamed and misrepresented

**Joanna Moncrieff**
Department of Mental Health Sciences, University College London, London, UK

Modern evidence reveals that there is little difference between antidepressants and placebo for the treatment of depression. This is the message of Khan and Brown's review (1) of antidepressant research. In fact, older studies came to the same conclusion. In 1969, the authors of a comprehensive review commissioned by the U.S. National Institute of Mental Health concluded that "in well-designed studies, the differences between the effectiveness of antidepressant drugs and placebo are not impressive" (2, p. 19).

Problems with the evidence for antidepressants go even deeper than Khan and Brown suggest, however. Not only does this evidence show that these drugs are little different from placebo, but also that there are no grounds to believe they have specific effects that would justify their classification as "antidepressants".

Like other drugs used for mental health problems, drugs classed as antidepressants are psychoactive substances. Psychoactive substances are drugs that enter the brain and by doing so modify normal thoughts, emotions and behaviours. Recreational drugs have psychoactive properties that some people find pleasant or exciting, but other drugs – including antipsychotics, anticonvulsants and antidepressants – have psychoactive effects that are less appealing. The psychoactive effects of individual antidepressants vary in strength and character, with the effects of some, such as the selective serotonin reuptake inhibitors (SSRIs), being weak and subtle, whereas the effects of others are more profound (e.g., the tricyclics).

The fact that antidepressants are psychoactive substances has major implications for the interpretation of placebo-controlled trials. Firstly, the use of a psychoactive substance will inevitably impact on the experiences and emotions captured by depression rating scales. The sedative effects of antidepressants, such as the tricyclics and newer drugs like mirtazapine, for example, are likely to reduce the degree of agitation, anxiety and insomnia experienced by people with depression. These symptoms feature strongly in measurement scales. Changes in sleep alone can account for up to 6 points on the Hamilton Rating Scale for Depression, for example, whereas typical antidepressant-placebo differences are around 2 to 3 points (3,4). Moreover, psychoactive effects may impact in varied ways on thoughts.

Secondly, the psychoactive effects of antidepressants, along with the physical modifications they produce (both commonly referred to as "side effects", although this is misleading since an independent, therapeutic effect has not been established), will infringe the double blind design. Some of the participants allocated to the active drug will be able to detect that they have received the real drug because of the physical or mental changes the drug produces, especially since they are provided with detailed information on possible side effects. Thus, it has been shown in many placebo-controlled trials of drugs used for mental health problems that participants can guess what they have received better than chance (5). In this situation, people allocated to the active drug are likely to have enhanced expectations of the effectiveness of therapy, and people who suspect they are taking placebo may have unduly negative expectations. Both of these factors may create or exaggerate a difference between antidepressants and placebo.

Unless the psychoactive effects of antidepressants are somehow discounted, differences between antidepressants and placebo cannot be interpreted as providing evidence that those drugs have a specific "antidepressant" effect. Indeed, it transpires that most drugs with psychoactive effects – including many antipsychotics, benzodiazepines, stimulants, buspirone and opiates (6) – produce the same changes as so-called antidepressants in randomized trials in people diagnosed with depression. Moreover, antidepressants themselves come from a wide array of chemical classes, and produce diverse pharmacological effects. Unsurprisingly, it seems that the experience of taking some sort of mind-altering substance produces a slightly different result from taking an inert placebo, when you attempt to measure people's thoughts and feelings (7,8).

If we accept this model, we need to ask whether the psychoactive effects

that antidepressants or other drugs produce might logically be useful in people with depression. There may, for example, be a role for temporary use of drugs with sedative properties to manage insomnia, anxiety and agitation in people who experience these symptoms, balancing proper evidence of benefits against adverse effects, including risks of dependence.

There has been some debate as to whether the SSRIs and related antidepressants produce a state of emotional suppression or disengagement. Antipsychotics are well known for dulling emotions, but the effects of SSRIs are likely to be more subtle. Evidence from converging sources now suggests that SSRIs and other newer antidepressants do have this property, and that it is associated with recognized side effects, such as loss of libido and sexual impairment (9-11). Many people dislike this state of emotional numbing but, in theory, some may find it useful to manage an intense emotional crisis. Again, we need evidence to explore whether this particular effect produces any tangible improvement in people suffering from depression, and whether people find it acceptable or not.

For decades now people have been told that depression is a chemical imbalance and that antidepressants work by correcting that imbalance. This view is not supported by evidence, and is misleading as to the nature and effects of antidepressant drugs. We need to recognize that antidepressants are psychoactive substances, we need more data on the nature of the varied psychoactive effects they produce, and we need to explore whether giving drugs that produce an artificially altered emotional state is a useful and acceptable intervention for people with depression.

## References


1. Khan A, Brown WA. Antidepressants versus placebo in major depression: an overview. World Psychiatry 2015;14:294-300.
2. Smith A, Traganza E, Harrison G. Studies on the effectiveness of antidepressant drugs. Psychopharmacol Bull 1969; 5(Suppl. 1).
3. Kirsch I, Moore TJ, Scoboria A et al. The emperor's new drugs: an analysis of antidepressant medication data submitted to the U.S. Food and Drug Administration. Prev Treat 2002;5:23.
4. Sugarman MA, Loree AM, Baltes BB et al. The efficacy of paroxetine and placebo in treating anxiety and depression: a meta-analysis of change on the Hamilton Rating Scales. PLoS One 2014;9:e106337.
5. Fisher S, Greenberg RP. How sound is the double-blind design for evaluating psychotropic drugs? J Nerv Ment Dis 1993; 181:345-50.
6. Moncrieff J. Are antidepressants overrated? A review of methodological problems in antidepressant trials. J Nerv Ment Dis 2001;189:288-95.
7. Moncrieff J, Cohen D. Rethinking models of psychotropic drug action. Psychother Psychosom 2005;74:145-53.
8. Moncrieff J. The myth of the chemical cure: a critique of psychiatric drug treatment. Basingstoke: Palgrave Macmillan, 2008.
9. Goldsmith L, Moncrieff J. The psychoactive effects of antidepressants and their association with suicidality. Curr Drug Saf 2011;6:115-21.
10. Bolling MY, Kohlenberg RJ. Reasons for quitting serotonin reuptake inhibitor therapy: paradoxical psychological side effects and patient satisfaction. Psychother Psychosom 2004;73:380-5.
11. Price J, Cole V, Goodwin GM. Emotional side-effects of selective serotonin reuptake inhibitors: qualitative study. Br J Psychiatry 2009;195:211-7.


DOI 10.1002/wps.20243

# Antidepressant or antidepressant plus placebo effect?

**Stuart A. Montgomery**
Imperial College School of Medicine, University of London, London, UK

Since the first observations of the therapeutic effect of imipramine, an extensive armamentarium of effective antidepressant medications has been developed, but the level of efficacy achieved appears to be considerably less today than might have been predicted from the discoveries of more than 40 years ago. Public opinion is currently suspicious as to the value of antidepressants in treating depressive illness. Khan and Brown's (1) thoughtful elucidation of possible factors involved helps to clarify if there has been a change in the way studies to establish efficacy are conducted, a change in the patients coming forward for treatment, and whether we should change our expectations on treatment efficacy.

The rise in the response to placebo reported in clinical studies, which certainly makes the demonstration of efficacy more difficult, has been attributed by some to pressure from the pharmaceutical industry to find a rewarding outcome in efficacy studies. The comparison of Khan et al (2) of the symptom reduction data reported in non-pharmaceutical industry studies with the data submitted by pharmaceutical companies to the U.S. Food and Drug Administration (FDA) finds few substantive differences. Differences in assessments between the two datasets appear to be related to size rather than direction, and when the non-pharmaceutical investigators were unaware of the study design the assessments came even closer to those of the FDA data. Industry-sponsored studies submitted for regulatory approval are designed in discussion with the regulatory authorities and have to meet their strict criteria; the similarity of the data from the two sources is reassuring.

Clinicians recognize the powerful contribution of placebo response in depression as in many other conditions. The task of distinguishing the pharmacological from the placebo response has become more difficult over the years, as the proliferation in regulatory requirements to address

efficacy, safety issues, particular patient groups, integrity of the study population, etc. has led to an increase in complexity of study design. The assessments now required to address all these aspects of treatment take considerable amounts of time to complete, and the therapeutic benefit of time spent with the patient is well recognized: an increase in the placebo response and a decrease in the separation of active medication from placebo is to be expected. We should remember that efficacy studies in psychiatry use a combination treatment paradigm: putative antidepressant plus placebo effect vs. placebo.

Khan and Brown (1) rightly point to the need for a reappraisal of the response measures used in determining efficacy. The bar to declaring efficacy differs between the U.S. and the European Union. In the European Union, efficacy has to be not merely established but also shown to be clinically relevant and found in placebo-controlled studies to persist in long-term treatment. Long-term treatment studies tend to be more consistently significant than short-term studies in demonstrating efficacy. Khan and Brown refer to their meta-analysis of response in continuation treatment (3), on the basis of which they consider that response on placebo is persistent. However, the methodology was flawed in that discontinuations from causes such as administrative dropouts or dropouts from side effects were censored and not taken into account. In a more specific study, response to placebo was not found to be persistent (4).

It has to be remembered that the efficacy of antidepressants was originally established in depressed patients, often hospitalized, with clear and relatively severe symptoms. As could be expected, on the basis of this success, the use of those and subsequent antidepressants has been extended to a much wider patient population suffering from a broader range of severity of illness. Antidepressants appear to be less effective or ineffective in patients with mild depression, and the global assessment of the efficacy of antidepressants would be diminished by the inclusion of this patient group. There is also a risk that, instead of identifying those patients who would best benefit from antidepressant treatment, under the pressure of the therapeutic imperative, some members of the "worried well" group receive the diagnosis of depressive illness rather than appropriate reassurance. This group has a high placebo response and should not be included in efficacy studies. The possible inclusion of large numbers of the worried well or of mild depression in U.S. clinical efficacy studies may have contributed to the increased difficulty in demonstrating efficacy of an antidepressant compared to placebo over the years. We would do well to heed the warning of the past chairman of the DSM-IV task force, in relation to the developments in DSM-5, that mild major depression is not major, not depression and not a disorder (5).

In the present climate, where both regulatory authorities under pressure from politicians and institutional review bodies require extra scales and restrictions, it is difficult to design a study which will allow the drug to show efficacy. Candidate antidepressants now need to be more effective in order to separate from placebo in a population with a higher placebo response. Studies carried out in an assay sensitive population cannot be compared with those carried out in a population with a high placebo response rate, and comparison of earlier and current studies is invalid. We should emphasize the success of the effect size achieved in current studies of antidepressants (0.31-0.33), which lies in the same range as many accepted treatments in general medicine.

### References


1. Khan A, Brown WA. Antidepressants versus placebo in major depression: an overview. World Psychiatry 2015;14:294-300.
2. Khan A, Faucett J, Lichtenberg P et al. A systematic review of comparative efficacy of treatments and controls for depression. PLoS One 2012;7:e41778.
3. Khan A, Redding N, Brown WA. The persistence of the placebo response in antidepressant clinical trials. J Psychiatr Res 2008;42:791-6.
4. Montgomery SA, Rasmussen JGC, Tanghoj P. A 24 week study of 20 mg citalopram, 40 mg citalopram and placebo in the prevention of relapse of major depression. Int Clin Psychopharmacol 2008;8:181-8.
5. Frances AJ, Nardo JM. ICD-11 should not repeat the mistakes made by DSM-5. Br J Psychiatry 2013;203:1-2.


DOI 10.1002/wps.20244

# Factors contributing to the increasing placebo response in antidepressant trials

**Siegfried Kasper, Markus Dold**
Department of Psychiatry and Psychotherapy, Medical University of Vienna, Vienna, Austria

In their interesting paper, Khan and Brown (1) sought to summarize the factors contributing to the trend of continuously decreasing drug-placebo differences in antidepressant randomized controlled trials (RCTs) (2). This trend conveys the impression that the newer marketed antidepressants are less efficacious than the older ones, or even that the older and well-established antidepressants have lost efficacy over time. Thus, recognizing the methodological reasons for the decline in antidepressant-placebo separation within RCTs is highly relevant in order to appraise the clinical value of an antidepressant for psychiatric routine

care. This is especially meaningful with regard to newly developed and newly introduced antidepressants.

In our opinion, the constant increase in placebo response over the last decades is the main factor accounting for the mitigation of the drug-placebo differences in antidepressant RCTs. Due to the larger symptom improvement of study participants randomized to placebo, it is much more difficult for an antidepressant to outperform placebo at a statistically significant level. This could be the beginning of an ominous cascade. The diminishing drug-placebo contrast leads to a higher probability of so-called inconclusive or even negative trials. As a consequence, the risk for a newly developed compound to fail the market approval because of negative phase III studies increases. As such late failure is associated with enormous costs for the pharmaceutical industry, this could result in a slowdown of research efforts for new antidepressant agents.

An often cited reason for the continuously increasing placebo response is the fact that placebo administration represents in itself a non-specific treatment (3). Study participants in the placebo groups of RCTs receive enormous clinical attention. Indeed, parallel to the raising placebo response in RCTs, the requirements concerning the accomplishment of RCTs have become more stringent, for instance through the need of an increasingly closer monitoring of participants (4). The intensive contact with the clinical staff can produce positive effects in terms of symptom improvement, particularly in not severely ill participants, who are nowadays increasingly included in RCTs, because the enrollment of more severely ill patients is often not possible due to ethical concerns (a phenomenon called "baseline inflation").

Another element contributing to the magnitude of placebo response in RCTs is the hope of participants in the placebo groups to receive an active, efficacious treatment (a phenomenon called "hope induction" or "expectation bias"). A number of systematic evaluations corroborate this assumption: the more study arms an RCT comprised (i.e., the lower the probability to receive placebo), the higher was the placebo response. On the other hand, the highest antidepressant response was found in direct drug comparisons (head-to-head trials), where participants were certain to receive active drug treatment (5).

Furthermore, in many antidepressant RCTs, the enrolled subjects are not representative of clinical practice, in which a number of depressive patients suffer from severe comorbidities or suicidal ideation. Exactly those severely ill patients who are excluded from RCTs might particularly benefit from antidepressant pharmacotherapy. However, it must be noted that, in some analyses, a higher symptom severity at baseline was associated with a higher placebo response (6), a phenomenon which is not fully understood as yet.

There is a large body of evidence suggesting that clinical studies carried out in the U.S. are characterized by a larger placebo response compared to non-U.S. trials. This phenomenon could be observed, to provide a recent example, in RCTs of vortioxetine (7). The inclusion of so-called professional research participants in U.S. studies may account for this finding. These subjects are mainly recruited by advertisements, and their motivation is often the prospect for free medication or other financial compensation. Therefore, they often aim to please the investigators in order to be invited again for participation in a clinical study.

It appears meaningful to emphasize that the drug-placebo contrast in long-term, relapse-prevention studies is usually higher compared to acute-phase trials (8). Interestingly, in a double-blind, long-term citalopram RCT, responders to acute-phase treatment were randomized to either placebo or continuing citalopram, while placebo responders of the acute phase continued the placebo administration under double-blind conditions. Both placebo responders and citalopram responders receiving placebo in the long-term study exhibited a higher relapse rate compared to the participants in the citalopram continuation arm (9), suggesting that the underlying biological abnormality is not sensitive to placebo.

In summary, the reasons contributing to the raising placebo response over time and the subsequent decreasing drug-placebo separation should be critically considered in the interpretation of antidepressant clinical trial results. Expectation bias and increased clinical attention are often called "unspecific effects" of placebo administration, and these effects are not present in routine clinical care (10). Therefore, it can be assumed that the effectiveness of an antidepressant in the routine care is higher than the antidepressant-placebo difference of RCTs indicates. The increase over time of placebo response should also be considered in meta-analyses (11), in which data are pooled from trials carried out in different periods.

## References


1. Khan A, Brown WA. Antidepressants versus placebo in major depression: an overview. World Psychiatry 2015;14:294-300.
2. Walsh BT, Seidman SN, Sysko R et al. Placebo response in studies of major depression: variable, substantial, and growing. JAMA 2002;287:1840-7.
3. Adam D, Kasper S, Moller HJ et al. Placebo-controlled trials in major depression are necessary and ethically justifiable: how to improve the communication between researchers and ethical committees. Eur Arch Psychiatry Clin Neurosci 2005;255:258-60.
4. Rutherford BR, Cooper TM, Persaud A et al. Less is more in antidepressant clinical trials: a meta-analysis of the effect of visit frequency on treatment response and dropout. J Clin Psychiatry 2013;74:703-15.
5. Rutherford BR, Roose SP. A model of placebo response in antidepressant clinical trials. Am J Psychiatry 2013;170:723-33.
6. Dunlop BW, Thase ME, Wun CC et al. A meta-analysis of factors impacting detection of antidepressant efficacy in clinical trials: the importance of academic sites. Neuropsychopharmacology 2012;37:2830-6.
7. Pae CU, Wang SM, Han C et al. Vortioxetine: a meta-analysis of 12 short-term, randomized, placebo-controlled clinical trials for the treatment of major depressive disorder. J Psychiatry Neurosci 2015;40:174-86.
8. Leucht S, Hierl S, Kissling W et al. Putting the efficacy of psychiatric and general

medicine medication into perspective: review of meta-analyses. Br J Psychiatry 2012;200:97-106.
9. Montgomery SA, Rasmussen JGC, Tanghoj P. A 24-week study of 20 mg citalopram, 40 mg citalopram, and placebo in the prevention of relapse of major depression. Int Clin Psychopharmacol 1993;8:181-8.
10. Hegerl U, Mergl R. The clinical significance of antidepressant treatment effects cannot be derived from placebo-verum response differences. J Psychopharmacol 2010;24:445-8.
11. Huf W, Kalcher K, Pail G et al. Meta-analysis: fact or fiction? How to interpret meta-analyses. World J Biol Psychiatry 2011;12:188-200.

DOI 10.1002/wps.20245

# Time to abandon placebo control in pivotal phase III trials?

**John R. Geddes, Andrea Cipriani**
Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford OX3 7JX, UK

Randomized controlled trials are the best way of testing therapeutic products to determine if they perform as expected and actually make a difference in treating a specific disease. Once preliminary evidence from phase II studies reveals that a treatment is probably effective, phase III trials are carried out to fully examine the risk/benefit profile of an experimental drug in a broader population over a longer period of time.

Hundreds of potential therapies are generated in laboratories, but only very few survive the early development stage and reach the point of human testing (1). There are many obstacles to the development of new treatments and the production of reliable evidence. These include the length of time and high financial cost involved in conducting clinical trials, the regulatory requirements for studies involving human subjects, and the difficulties in recruiting the appropriate patient population. Antidepressant trials are a good example of how difficult is to innovate in psychopharmacology, and Khan and Brown (2) discuss this issue focusing on the place of placebo-controlled studies in major depression.

A truly novel antidepressant has not been introduced for 30 years. Among multiple potential explanations, the positive impact that selective serotonin reuptake inhibitors (SSRIs) have had on clinicians and researchers may have played a key role. Notwithstanding the success of the SSRIs, however, pharmacological treatment for depression remains far from being optimal. Key challenges in depression research include the lack of objective markers for diagnosing depression (which is still largely based on subjective evaluation even in DSM-5) and the fact that trials of antidepressants are not primarily focused on answering the most important clinical issues (i.e., comparative effectiveness between interventions, long-term outcomes).

In terms of study design, the use of placebo is probably the most compelling issue (3). The majority of clinical trials in depression are placebo-controlled, because regulators require them for licensing approval (4). The rate of placebo responders in these trials, however, has added a layer of complexity and difficulty to the process of designing trials and interpreting results. Placebo responders in antidepressant trials have been increasing over time since 1981 (5). How can this be explained? Either participants are becoming more suggestible and/or placebo more effective, or the increase in placebo response must be artefactual.

In fact, the increasing placebo response rate can at least partially be explained by the phenomenon called "inflation of baseline severity" (6). Entry criteria for clinical trial participants are based not only on the categorical diagnosis, but also on the severity of the illness. Usually, the minimum criterion for enrolling an individual in an antidepressant trial is a diagnosis of major depressive disorder and a total score greater than a pre-specified threshold according to a standardized rating scale. These measures of severity of depressive symptoms, even if rated by trained assessors, are subjective and can be easily unconsciously manipulated. Researchers always struggle to find participants who meet the entry criteria, are eligible for randomization and are willing to accept randomization in a trial in which they know they may receive placebo. To recruit to time and target, investigators may tend to overemphasize some symptoms and give a higher score to some of the items to reach the minimum overall score on the rating scale and get the patient into the study. As a result, bias is introduced in the selection of participants (skewed distribution) and too mildly ill patients are enrolled who are more likely to remit "spontaneously", which means to respond to placebo. After their high initial ratings, in fact, physicians begin to rate the condition of patients more accurately. The main consequence of inflation of baseline severity is a large drop in the severity scores between randomization and endpoint, also in the mildly ill patients who remit "spontaneously" without receiving any active treatment, thus making the placebo appear more effective.

Innovation in psychopharmacology is urgently needed not only in terms of drug discovery and development, but also in terms of the design of phase III clinical trials. The need for comparative effectiveness has been reported and highlighted many times recently (7). Therefore, the key question to ask is whether we still routinely need

to have placebo-controlled phase III studies. It has been suggested that they are needed in the field of depression because findings of equivalence between a new drug and standard treatment are not evidence of efficacy unless the new drug is also significantly more effective than placebo (8). This assay sensitivity may be required in phase II studies, but in phase III one could argue that the essential question is whether a new treatment is superior to existing therapies.

We now know that some antidepressants are better than others and that individual drugs can be ranked according to their efficacy and acceptability profiles (9). Comparative effectiveness research is a key element of current efforts in health care reform in Europe and also in the U.S. (10). Prioritizing this kind of research in the field of antidepressants presents a methodological challenge. To improve treatments and patient outcomes, we need phase III trials with a superiority design against an active comparator, chosen among the most effective and better tolerated treatments already available on the market via a reliable and transparent meta-analytical process. This will set a more ambitious and clinically meaningful target, and will foster much-needed innovation in psychopharmacology to developing more effective interventions. It will also be in line with the ethical requirements of not exposing patients to placebo when an effective treatment is available.

Despite the importance of creating new therapies, investment in neuroscience is no longer a top priority for pharmaceutical industry. Drug discovery in psychiatry is just too difficult for current commercial models. It is time for academics to join forces with industry, creating new models of drug discovery (11) and driving innovation in the methodology of drug development. Abandoning the standard requirement for placebo control in pivotal phase III trials, and setting more ambitious targets for treatment advance, should be a bold first step.

## References


1. Institute of Medicine. Transforming clinical research in the United States: challenges and opportunities: Workshop summary. Washington: National Academies Press, 2010.
2. Khan A, Brown WA. Antidepressants versus placebo in major depression: an overview. World Psychiatry 2015;14:294-300.
3. Papakostas GI, Østergaard SD, Iovieno N. The nature of placebo response in clinical studies of major depressive disorder. J Clin Psychiatry 2015;76:456-66.
4. Laughren TP. The scientific and ethical basis for placebo-controlled trials in depression and schizophrenia: an FDA perspective. Eur Psychiatry 2001;16:418-23.
5. Walsh BT, Seidman SN, Sysko R et al. Placebo response in studies of major depression: variable, substantial, and growing. JAMA 2002;287:1840-7.
6. Landin R, DeBrota DJ, DeVries TA et al. The impact of restrictive entry criterion during the placebo lead-in period. Biometrics 2000;56:271-8.
7. Li T, Vedula SS, Scherer R et al. What comparative effectiveness research is needed? A framework for using guidelines and systematic reviews to identify evidence gaps and research priorities. Ann Intern Med 2012;156:367-77.
8. Kupfer DJ, Frank E. Placebo in clinical trials for depression: complexity and necessity. JAMA 2002;287:1853-4.
9. Cipriani A, Furukawa TA, Salanti G et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. Lancet 2009;373:746-58.
10. American Recovery and Reinvestment Act. Public Law 111-5, 111th Congress, 2009.
11. Norman TC, Bountra C, Edwards AM et al. Leveraging crowdsourcing to facilitate the discovery of new medicines. Sci Transl Med 2011;3:88mr1.


DOI 10.1002/wps.20246

# The role of regulators, investigators, and patient participants in the rise of the placebo response in major depressive disorder

**Maurizio Fava**

Clinical Trials Network and Institute (CTNI), Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

Khan and Brown's comprehensive overview (1) provides numerous opportunities for reflection on the role of regulators, investigators, and patient participants in the progressive rise of the placebo response in major depressive disorder (MDD). They rightfully argue that the DSM-III and subsequent iterations have led to an expansion of the population meeting criteria for that disorder, and that some of the patients entering MDD trials may have questionable forms of this condition. This is certainly consistent with our experience of independently interviewing patients considered to meet criteria for MDD by site investigators (2).

Khan and Brown also review the robust evidence for investigators' bias influencing the magnitude of symptom reduction across all treatments, including placebo. Their arguments are consistent with a review from our group (3), which has shown smaller effect sizes in antidepressant trials using active placebos compared to those observed in the presumably less blinded trials with inert placebo, and has pointed out that a trial of quetiapine in bipolar depression found no difference in efficacy between active treatment and placebo in the groups of patients reporting sedation as an adverse event.

While I fully agree with Khan and Brown's views on the role of patient participants and investigators in the rise of placebo response in antidepressant

trials, I differ on their perspectives on the regulatory contributions to this problem. In particular, they argue that U.S. Food and Drug Administration (FDA)'s decisions have, at times, negatively influenced the design and execution of antidepressant trials in the past three decades by using conservative approaches to both design and analysis of trials. I would argue the opposite: the FDA, in my opinion, has shown a great deal of openness to innovation and to novel designs and methods, but we, as a field, have justified our conservative stances by using the anticipation of a negative FDA response as a reason for holding on to obsolete standard designs. Let me offer a few examples.

Khan and Brown mention the regulatory burden of the required use of an active control to establish assay sensitivity, an approach shown to lead to higher placebo response rates due to expectations of increased odds of receiving a form of active treatment (4). As far as I can tell, this has never been a regulatory requirement in adult MDD trials (though often suggested), but it became a popular approach among sponsors with the goal of de-risking investments in novel therapies of uncertain efficacy. In fact, the FDA has approved vilazodone as an antidepressant in the absence of any data involving an active control (5).

Similarly, Khan and Brown state that dose-response data are a regulatory requirement that has led to inflated placebo response rates because of the expectations of increased odds of receiving a form of active treatment (4). Once again, although the FDA likes dose-response data in a submission, there is no official requirement in phase III and, in fact, vilazodone was approved based on two studies that evaluated only one dose (5).

Khan and Brown also state that the FDA staff has never accepted the concept of therapeutic response as a valid primary outcome and that the FDA considers valid only the use of the change in total scores of scales such as the Hamilton Depression Rating Scale (HAM-D) or the Montgomery-Asberg Depression Rating Scale (MADRS) as primary outcome measure. Once again, sponsors have favored this approach because of the greater sensitivity to detect changes of a continuous measure compared to a dichotomous measure, but, in fact, the FDA has accepted proposed registration studies using time to response as the primary outcome.

Khan and Brown add that the FDA does not accept, contrary to the European Medicines Agency (EMA), randomized withdrawal study designs. Actually, S. Borges and other colleagues from the Division of Psychiatry Products of the FDA usually include a post-marketing requirement for such type of maintenance studies in the initial approval, and have in fact published a paper (6) strongly endorsing this design as having good sensitivity to detect signals with antidepressants. Indeed, lamotrigine was approved for maintenance treatment of bipolar disorder by the FDA only based on two relapse prevention trials (7), i.e., in the absence of acute treatment data.

Another example of using the anticipation of a negative FDA response as a reason for holding on to obsolete standard designs is the fact that a number of sponsors have often used longer duration of trials "because the FDA requires it". On the contrary, N.A. Khin and other colleagues from the Division of Psychiatry Products of the FDA (8) have shown in their pooled analyses of 81 trials of antidepressants that 6-week trials had a higher success rate than 8-week trials (55% vs. 42%), and trials as short as four weeks are considered acceptable by the FDA.

Finally, adaptive designs have been mentioned as critical innovations by the FDA Director of Psychiatry Products at that time (9), and novel study designs aimed at reducing the placebo response, such as the sequential parallel comparison design, have been used in Phase II and Phase III antidepressant trials (10), with FDA statisticians having published on new methods to analyze them (11).

While it is important to identify factors that may have contributed to rising placebo response in depression trials, there is no evidence that the regulatory agencies *per se* played any role in this. On the contrary, we, as a field, have justified our conservative stances by using the anticipation of a negative FDA response as a reason for holding on to obsolete standard designs and methodologies.

## References


1. Khan A, Brown WA. Antidepressants versus placebo in major depression: an overview. World Psychiatry 2015;14:294-300.
2. Desseilles M, Witte J, Chang TE et al. Massachusetts General Hospital SAFER criteria for clinical trials and research. Harv Rev Psychiatry 2013;21:269-74.
3. Perlis RH, Ostacher M, Fava M et al. Assuring that double-blind is blind. Am J Psychiatry 2010;167:250-2.
4. Tedeschini E, Fava M, Goodness TM et al. Relationship between probability of receiving placebo and probability of prematurely discontinuing treatment in double-blind, randomized clinical trials for MDD: a meta-analysis. Eur Neuropsychopharmacol 2010;20:562-7.
5. Laughren TP, Gobburu J, Temple RJ et al. Vilazodone: clinical basis for the US Food and Drug Administration's approval of a new antidepressant. J Clin Psychiatry 2011;72:1166-73.
6. Borges S, Chen YF, Laughren TP et al. Review of maintenance trials for major depressive disorder: a 25-year perspective from the US Food and Drug Administration. J Clin Psychiatry 2014; 75:205-14.
7. McIntyre RS, Cha DS, Kim RD et al. A review of FDA-approved treatment options in bipolar depression. CNS Spectr 2013;18(Suppl. 1):4-20.
8. Khin NA, Chen YF, Yang Y et al. Exploratory analyses of efficacy data from major depressive disorder trials submitted to the US Food and Drug Administration in support of new drug applications. J Clin Psychiatry 2011;72:464-72.
9. Laughren TP. What's next after 50 years of psychiatric drug development: an FDA perspective. J Clin Psychiatry 2010;71:1196-204.
10. Heger M. Trial designs advance to overcome bitter pill of placebo effect. Nat Med 2013;19:1353.
11. Chen YF, Yang Y, Hung HM et al. Evaluation of performance of some enrichment designs dealing with high placebo response in psychiatric clinical trials. Contemp Clin Trials 2011;32:592-604.


DOI 10.1002/wps.20247

# The responsiveness of the different versions of the Hamilton Depression Scale

**Per Bech**

Psychiatric Research Unit, Psychiatric Centre North Zealand, Copenhagen University Hospital, Hillerød, Denmark

In clinical pharmacology, the evidence proving that a drug has a therapeutic effect in a specific medical condition is based on two major elements: superiority of the drug over placebo in randomized clinical trials within the medical condition under examination, and a systematic relationship between the dose of the drug and the magnitude of the response it elicits.

In their overview of antidepressants versus placebo, Khan and Brown (1) conclude that "no clear dose-response relationship has been established to date for most of the new antidepressants", while the superiority of the antidepressants over placebo in terms of effect size statistics is approximately 0.30, a level they find "less than impressive".

In their review of double-blind, placebo-controlled trials of antidepressants conducted from 1981 to 2008, Khan et al (2) observed that the Hamilton Depression Scale (HAM-D) had been used as an outcome scale in most of the trials. However, the HAM-D was actually used in two different versions, the 21-item version (HAM-D-21) and the 17-item version (HAM-D-17). The HAM-D-21 was used in one third of the trials and the HAM-D-17 in two thirds. Unfortunately, authors who use the HAM-D-21 rarely provide information about the results on the HAM-D-17. Khan and Brown (1) highlight now that the antidepressant-placebo difference seems to be higher in HAM-D-21 trials compared to the trials in which the HAM-D-17 has been used as outcome measure. This is a tautological finding in so far as the standard deviation of this difference is not available, which is the case for most of the trials reviewed by Khan et al (2).

Among the trials collected by Khan et al (2) it is possible, however, to identify ten publications in which the six-item HAM-D (HAM-D-6) is compared to both HAM-D-17 and HAM-D-21, or to HAM-D-28. The HAM-D-6 covers the core symptoms of depression: depressed mood, work and interests, guilt feelings, psychomotor retardation, psychic anxiety, and general somatic (fatigability). These six items have clinical and psychometric validity (3). In two of these ten trials, a dose-response relationship was investigated. Fabre et al (4) showed that sertraline was significantly superior to placebo at all three doses (50, 100, 200 mg daily) when using HAM-D-6, but only at 50 mg daily when using HAM-D-17. Liebowitz et al (5) showed that desvenlafaxine was superior to placebo at both 50 and 100 mg daily when using HAM-D-6, but only at 50 mg daily when using HAM-D-17.

An analysis of all placebo-controlled trials of desvenlafaxine showed that at doses of 200 or 400 mg daily the effect size was negative during the first week of treatment (superiority of placebo) when using HAM-D-17 but not when using the HAM-D-6, implying that the HAM-D-17 includes symptoms which might be side effects of the drug (6). In placebo-controlled trials of fluoxetine, over a dose range from 20 to 60 mg daily, the effect size using HAM-D-17 was approximately 0.30, but when using HAM-D-6 it was approximately 0.40 (3). For escitalopram, a dose of 10 mg daily obtained an effect size of 0.38 mg using HAM-D-6 and a dose of 20 mg daily gave an effect size of 0.61 (3).

Over the past decade, the goal when evaluating the effect of an antidepressant has been the event of remission rather than response (7). Remission in major depression is defined as a minimal level of the core symptoms of depression (7). The syndrome reflected by the HAM-D-6 is a unidimensional measure for specific drug targets, and a cut-off score below 5 indicates that the individual symptoms of the scale are only present to a very doubtful degree (remission).

Khan and Brown (1) refer to the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) as an example of a poor response to citalopram treatment. Indeed, when using the conventional definition of remission (HAM-D-17 $<8$), only approximately 36% of the patients achieved remission. However, when using a HAM-D-6 score of $<5$ as the definition of remission, approximately 45% of the patients in that study achieved remission ($p<0.001$) (8).

From a statistical point of view, failed trials are merely a consequence of insufficient power, as the inability to reject the null hypothesis is inherently associated with low statistical power. This has recently been illustrated in a re-analysis of a failed study which had used the HAM-D-17 to evaluate the effect of erythropoietin as augmentation in patients with treatment-resistant depression (9). By focusing on the HAM-D-6, fewer patients are needed to reject the null hypothesis.

Psychometrically, Khan and Brown (1) correctly focus on the use of the HAM-D-17 as the major factor for the "less than impressive" effect size of 0.30 and the lack of a dose-response relationship. However, their solution to go for a larger HAM-D version (HAM-D-21) is not justified. The solution is to go for the brief, clinically and psychometrically valid subscale (HAM-D-6).

The use of the HAM-D-6 as outcome measure in placebo-controlled clinical trials of antidepressants increases the effect size to 0.40, which is indicative of clinical significance. Using the HAM-D-6 as outcome measure, a dose-response relationship has been established for newer antidepressants

such as escitalopram and desvenlafaxine. Moreover, fewer patients are then needed to identify antidepressant effect in controlled trials, which has important ethical implications (fewer patients need to receive placebo).

In my opinion, we need to aim at establishing "dose-remission" rather than dose-response relationship in future trials of antidepressants. The HAM-D-6 contains the core symptoms of depression by which to define the event of remission.

### References


1. Khan A, Brown WA. Antidepressants versus placebo in major depression: an overview. World Psychiatry 2015;14:294-300.
2. Khan A, Bhat A, Kolts R et al. Why has the antidepressant-placebo difference in antidepressant clinical trials diminished over the past three decades? CNS Neurosci Ther 2010;16:217-26.
3. Bech P. Is the antidepressive effect of second-generation antidepressants a myth? Psychol Med 2010;40:181-6.
4. Fabre LF, Abuzzahab FS, Amin M et al. Sertraline safety and efficacy in major depression: a double-blind fixed-dose comparison with placebo. Biol Psychiatry 1995; 38:592-602.
5. Liebowitz MR, Manley AL, Padmanabhan SK et al. Efficacy, safety, and tolerability of desvenlafaxine 50 mg/day and 100 mg/day in outpatients with major depressive disorder. Curr Med Res Opin 2008;24:1877-90.
6. 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.
7. Rush AJ, Kraemer HC, Sackeim HA et al. Report by the ACNP Task Force on response and remission in major depressive disorder. Neuropsychopharmacology 2006;31:1841-53.
8. Ostergaard SD, Bech P, Trivedi MH et al. Brief, unidimensional melancholia rating scales are highly sensitive to the effect of citalopram and may have biological validity: implications for the Research Domain Criteria (RDoC). J Affect Disord 2014;163:18-24.
9. Ostergaard SD, Bech P, Miskowiak KW. Fewer study participants needed to demonstrate superior antidepressant efficacy when using the Hamilton melancholia subscale (HAM-D6) as outcome measure. J Affect Disord (in press).


DOI 10.1002/wps.20248

# What if a placebo effect explained all the activity of depression treatments?

**Pim Cuijpers, Ioana A. Cristea**

[1]Department of Clinical Psychology, Faculty of Psychology and Education, University of Amsterdam, Amsterdam, The Netherlands, [2]Department of Clinical Psychology and Psychotherapy, Babes Bolyai University, Cluj, Romania

Many randomized trials have shown that when depressed patients receive no active treatment, e.g. they are administered pill placebo, a large part of them improve anyway. This improvement can be partly explained by natural remission or by the patients' expectations that a treatment will have an effect on their problems (even when they receive pill placebo). The corollary is that many patients remit even when undergoing exotic therapies, such as Argentian tango, swimming with dolphins or horticulture (1-3).

This phenomenon makes it difficult to examine the additional effects of specific treatments. This is not only true for pharmacotherapy, but also for psychotherapies for depression. In a recent meta-analysis, we found that 62% of patients meeting criteria for major depression at baseline did no longer meet these criteria after treatment (4). But among the patients receiving only care-as-usual, 48% also no longer met criteria for major depressive disorder. So, therapists may think that more than 60% of patients get better because of the psychotherapy, while in fact the additional benefit of psychotherapy over usual care is only 14%. Khan and Brown (5) indicate that comparable outcomes take place for pharmacotherapy, with symptom reduction of about 40% with antidepressants and 30% with placebo. That is in line with Kline's conclusion in 1964 that "in the treatment of depression, one always has an ally in the fact that most depressions terminate in spontaneous remissions" (6).

Given this large proportion of patients who remit spontaneously, patients as well as therapists can easily be led into the idea that their treatment is highly successful, while in fact the effects of this treatment may be only moderate. This may also explain why the exotic treatments mentioned earlier are believed by some to be effective, while most clinicians would consider the specific effects of such treatments as not very credible. "But we see that patients get better" is a phrase that supporters of such therapies often use.

Due to the discrepancy between the relatively high rate of spontaneous remission and the low additional value of specific (pharmacological and psychological) treatments, several important issues arise. One question is whether these treatments do in fact have any effects. Of course, randomized trials show that pharmacotherapy and psychotherapy are effective for treating depression, with small effect sizes of 0.30 for antidepressants (5) and 0.25 for psychotherapies (7). But we also know that these effects are much higher when risk of bias is not taken into account. In fact, only the highest quality studies show such small effects, and only after publication bias has been adjusted for.

But suppose there is still a bias lingering in these trials. For example, since patients getting a placebo know that they are not receiving active medication because they experience no side effects, this breaks the blinding and serves to lower their expectations.

A meta-analysis of trials with active placebos pointed exactly in this direction (8). Or investigators' choices may influence the trial outcomes in ways that are just not known, for example by selecting those patients who are expected to respond well to treatment but not to placebo (9). The effects of the active treatments are so small that only slight tweaking because of some bias may further them, and make them clinically irrelevant. The same is true for psychotherapies. Their effects compared with pill placebo are very small and, because patients cannot be blinded at all, expectations may have a considerable effect on the outcomes. Only a small adjustment because of an unknown risk of bias could move these effects into clinical irrelevance as well.

The other implication is that research should focus much more on how spontaneous remission takes place. Now most of the research is focused on the brain changes and the psychological mechanisms involved in the action of biological and psychological therapies. However, the process through which spontaneous remission occurs is at least as important as the mechanisms through which these specific treatments work, particularly since their additional effectiveness is not as high as has been thought for a long time.

Hence, a clinical issue in need of much more investigation is by what mechanisms spontaneous remission can be optimized. For example, it can be assumed that when expectations of outcome are higher, spontaneous remission is more likely to occur. If we understood this process better, we could also find ways to optimize expectations and thus increase remission rates. That would eventually reduce the relative contribution of current treatments towards remission, though they still may lead to better outcomes for patients.

Khan and Brown conclude that the effects of antidepressants are modest, and other research shows that the same holds for psychological treatments for depression. We argue that the high rate of spontaneous remission introduces considerable confusion about the effectiveness of treatments. In order to improve outcomes for patients we have to face facts, and focus much more on the process of natural recovery instead of on the limited contributions of specific treatments.

## References


1. Pinniger R, Brown RF, Thorsteinsson EB et al. Argentine tango dance compared to mindfulness meditation and a waiting-list control: a randomised trial for treating depression. Complement Ther Med 2012; 20:377-84.
2. Kamioka H, Tsutani K, Yamada M et al. Effectiveness of horticultural therapy: a systematic review of randomized controlled trials. Complement Ther Med 2014;22:930-43.
3. Kamioka H, Okada S, Tsutani K et al. Effectiveness of animal-assisted therapy: a systematic review of randomized controlled trials. Complement Ther Med 2014;22:371-90.
4. Cuijpers P, Karyotaki E, Weitz E et al. The effects of psychotherapies for major depression in adults on remission, recovery and improvement: a meta-analysis. J Affect Disord 2014;159:118-26.
5. Khan A, Brown WA. Antidepressants versus placebo in major depression: an overview. World Psychiatry 2015;14:294-300.
6. Kline N. The practical management of depression. JAMA 1964;190:122-30.
7. Cuijpers P, Turner EH, Mohr DC et al. Comparison of psychotherapies for adult depression to pill placebo control groups: a meta-analysis. Psychol Med 2014;44: 685-95.
8. Moncrieff J, Wessely S, Hardy R. Active placebos versus antidepressants for depression. Cochrane Database Syst Rev 2004;1:CD003012.
9. Iovieno N, Papakostas GI. Correlation between different levels of placebo response rate and clinical trial outcome in major depressive disorder: a meta-analysis. J Clin Psychiatry 2012;73:1300-6.


DOI 10.1002/wps.20249

# Pragmatic randomized controlled trial of long-term psychoanalytic psychotherapy for treatment-resistant depression: the Tavistock Adult Depression Study (TADS)

Peter Fonagy[1], Felicitas Rost[2], Jo-anne Carlyle[2], Susan McPherson[3], Rachel Thomas[2], R.M. Pasco Fearon[1], David Goldberg[4], David Taylor[2]

[1]Research Department of Clinical, Educational and Health Psychology, University College London, London, UK; [2]Adult Department, Tavistock & Portman NHS Foundation Trust, London, UK; [3]School of Health and Human Sciences, University of Essex, Colchester, UK; [4]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

*This pragmatic randomized controlled trial tested the effectiveness of long-term psychoanalytic psychotherapy (LTPP) as an adjunct to treatment-as-usual according to UK national guidelines (TAU), compared to TAU alone, in patients with long-standing major depression who had failed at least two different treatments and were considered to have treatment-resistant depression. Patients (N=129) were recruited from primary care and randomly allocated to the two treatment conditions. They were assessed at 6-monthly intervals during the 18 months of treatment and at 24, 30 and 42 months during follow-up. The primary outcome measure was the 17-item version of the Hamilton Depression Rating Scale (HDRS-17), with complete remission defined as a HDRS-17 score ≤8, and partial remission defined as a HDRS-17 score ≤12. Secondary outcome measures included self-reported depression as assessed by the Beck Depression Inventory - II, social functioning as evaluated by the Global Assessment of Functioning, subjective wellbeing as rated by the Clinical Outcomes in Routine Evaluation - Outcome Measure, and satisfaction with general activities as assessed by the Quality of Life Enjoyment and Satisfaction Questionnaire. Complete remission was infrequent in both groups at the end of treatment (9.4% in the LTPP group vs. 6.5% in the control group) as well as at 42-month follow-up (14.9% vs. 4.4%). Partial remission was not significantly more likely in the LTPP than in the control group at the end of treatment (32.1% vs. 23.9%, p=0.37), but significant differences emerged during follow-up (24 months: 38.8% vs. 19.2%, p=0.03; 30 months: 34.7% vs. 12.2%, p=0.008; 42 months: 30.0% vs. 4.4%, p=0.001). Both observer-based and self-reported depression scores showed steeper declines in the LTPP group, alongside greater improvements on measures of social adjustment. These data suggest that LTPP can be useful in improving the long-term outcome of treatment-resistant depression. End-of-treatment evaluations or short follow-ups may miss the emergence of delayed therapeutic benefit.*

**Key words:** Treatment-resistant depression, psychoanalytic psychotherapy, long-term treatment, delayed therapeutic effect

***(World Psychiatry 2015;14:312–321)***

The potential to follow a chronic, relapsing course is a substantial part of what makes depression one of the largest contributors to the burden of human disease worldwide (1,2). Treatments for major depressive disorder generally have medium effect sizes (3), but observational studies and trials consistently report high rates of non-response (4,5), with 12 to 20% of depressed patients not benefitting even from multiple courses of treatment (6). This is often termed treatment-resistant depression.

Recent systematic reviews of treatment research for this patient group, whether considered separately (7) or combined with chronic major depressive disorder (8), revealed that existing studies are mostly of poor quality and design (9). Trials of novel neuromodulation therapies – such as repetitive transcranial magnetic stimulation (10), deep brain stimulation (11) and vagus nerve stimulation (12) – with these patients have shown serious limitations. There is some evidence supporting the augmentation of initial antidepressant medication with other classes of drugs (e.g., atypical antipsychotics) (7), or the adjunct of cognitive-behavioral therapy (CBT) to that medication (8,13), at least for patients with severe but non-chronic (episode ≤2 years) major depression (14).

Evidence is accumulating that, in order to be effective, interventions for treatment-resistant depression may need to be longer and more complex than first-line treatments of depression (15), and that follow-ups should be longer (16).

Some empirical evidence for short-term psychodynamic psychotherapies in the treatment of depression is available (e.g., 17). However, given the likelihood that a longer intervention will be needed, these therapies may have little relevance to populations of patients with treatment-resistant depression (18).

Evidence-gathering regarding the effectiveness of longer-term, more intensive psychoanalytic treatments is in its early stages (19). One recent meta-analysis identified 27 studies, most being either observational or quasi-randomized, with groups matched retrospectively (20). One quasi-randomized but otherwise methodologically strong study found long-term psychodynamic psychotherapy to be less effective over the short term than brief focused therapies for a sample of mood-disordered patients. However, after a 3-year follow-up, long-term psychodynamic psychotherapy was found to be superior (21).

Notwithstanding their various methodological shortcomings, the findings of studies with a multi-year follow-up period do suggest that there may be benefits from long-term psychodynamic psychotherapy (≥50 sessions) for patients with depression (20,22), particularly in the longer term (18,23).

Given the limitations of the evidence base concerning management of treatment-resistant depression, the present pragmatic randomized controlled trial assessed whether this condition is more likely to improve when long-term psychoanalytic psychotherapy (LTPP) is provided in addition to treatment-as-usual according to UK national guidelines (TAU), but excluding the short-term forms of psychological therapy recommended by those guidelines. We hypothesized, on the basis of accumulating evidence from non-randomized controlled studies (20-22), that the effect of LTPP would increase over the course of a longer than usual follow-up period.

## METHODS

### Study design and participants

Patients were recruited from primary care from February 2002 to May 2009 and assessed at the Adult Service of the Tavistock & Portman National Health Service (NHS) Foundation Trust in London. They were not paid and consented only after receiving a complete written description and thorough discussion of the study.

After baseline assessment, randomization to an 18-month course of LTPP plus TAU or TAU alone was carried out offsite by an independent statistician using a stochastic minimization program (MINIM) balancing for gender, depression severity (scores of 21-39 or 40+ on the Beck Depression Inventory - II, BDI-II (24)), and medication (on/off). Patients were then followed up for 24 to 42 months post-randomization according to an intention-to-treat design.

The trial methodology was published in advance of trial completion and data analysis (25). The study protocol was registered with the International Randomized Controlled Trial Number Register (ISRCTN40586372), and approved by the Institutional Review Board of NHS West Midlands Research Ethics Committee (MREC02/07/035).

In total, 308 patients were screened for eligibility. Of these, 235 attended for interview. Inclusion criteria were: age 18-65 years; current DSM-IV diagnosis of major depressive disorder as ascertained by the Structured Clinical Interview for DSM-IV (SCID-I, 26); minimum duration of two years of the current depressive episode; minimum score of 14 on the 17-item version of the Hamilton Depression Rating Scale (HDRS-17, 27) and of 21 on the BDI-II; and at least two failed treatment attempts (elicited at interview and verified from medical records), one of which must have included treatment with an antidepressant medication, and the other with either an antidepressant medication or a psychological intervention. Exclusion criteria were: receiving psychodynamic psychotherapy in the past two years; currently, or in the past five years, meeting DSM-IV criteria for psychotic disorder or bipolar I disorder; receiving psychiatric input for substance dependence in the past two years; moderate or severe learning disability, and evidence of organic brain disorder. No assessment for presumed suitability or unsuitability for psychoanalytic forms of therapy was performed.

### Treatments

LTPP consisted of 60 (50 min) sessions of once-weekly individual psychoanalytic psychotherapy over 18 months. The treatment manual (28) describes the intervention and methods, which are based on the view that depression is an outgrowth of current life difficulties arising out of painful and continuing ambivalence first felt in relation to those of the greatest emotional significance to the patient early in the course of his/her development.

The theory employed in LTPP assumes that, in patients with treatment-resistant depression, problems with psychosocial functioning impair help-seeking and illness-combating behaviors, and may also have an emotional impact upon health care/service providers in a way that affects the care they offer (29,30). LTPP enables these patients to gradually internalize a psychological capacity to relate to pathogenic personal experiences, memories, feelings, beliefs and relationships in a reflective, yet also more active, manner (31).

All the therapists (N=22; average years of experience: 17.45) had a mental health qualification and a training approved by the British Psychoanalytic Council. All therapy sessions were audio-recorded. Fidelity to treatment was assessed with the 100-item Psychotherapy Process Q-Sort (32). Three randomly selected sessions from the early, middle and end phases of each treatment were rated (183 sessions in total). Inter-rater reliability, assessed in a subsample of 90 sessions, was excellent: intraclass correlation coefficients (ICCs) after Spearman-Brown correction ranged from 0.68 to 0.98 (mean 0.87). As expected, analysis revealed that in 82.2% of cases the highest correlation obtained was with the psychodynamic prototype (mean $r=0.45$, $p<0.001$), with the remainder (17.8%) best resembling the CBT prototype (mean $r=0.28$, $p<0.05$).

TAU consisted of interventions as directed by the referring practitioner. This could include referral for other specialist provisions. In the UK's NHS, the range of these interventions is defined, and to an extent specified, in the treatment guidelines of the National Institute for Clinical Excellence (33). Referral to psychoanalytic psychotherapy is not within the guidance. In the LTPP group, the short-term forms of psychological therapy included in the guidelines were not allowed. Treatments received were recorded using the Client Service Receipt Inventory (34) and health care records.

### Assessments

Assessments were based on data collected at entry; at 6, 12 and 18 months over the course of treatment; and at 24, 30 and 42 months during follow-up.

The primary outcome measure was the HDRS-17, modified to include increases in sleep, appetite and weight (35). Trained interviewers blinded to treatment condition conducted the evaluations. All evaluations were recorded, and all interviews were double-rated by an independent blinded coder to establish inter-rater reliability. An ICC of 0.89 was obtained for the total HDRS-17 score with the following severity bands: 0-7 not depressed, 8-13 mild depression, 14-18 moderate depression, 19-22 severe depression, ≥23 very severe depression. Full remission was defined as an HDRS-17 score of 8 or less (36). Following Hollon et al (14), HDRS-17 scores ≤12 were considered to meet criteria for partial remission.

Secondary outcomes included self-reported depression as assessed by BDI-II; social functioning as evaluated by the Global Assessment of Functioning (GAF, 37); subjective wellbeing as rated by the Clinical Outcomes in Routine Evaluation - Outcome Measure (CORE-OM, 38); and satisfaction with general activities as assessed by the Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q, 39).

### Statistical analysis

Data analysis was by intention to treat. All analyses were carried out using Stata Statistical Software Release 14 (40). Power calculations were based on statistical analysis of data from another trial of long-term psychodynamic therapy with a similarly heterogeneous population (41). We conservatively assumed an intra-cluster correlation coefficient for therapists of 0.05: with a minimum of 10 therapists delivering each therapy, each seeing on average five patients, the study with N=129 has 80% power to reject the null hypothesis of equivalence, with a non-infidelity margin equal to an effect size of 0.5 using a 95% one-sided confidence interval, on the basis of a 80% rate of follow-up to 42 months. Adequacy of randomization was assessed by between-group comparisons of baseline characteristics on all measures, using $\chi^2$ tests for dichotomous variables and Kruskal-Wallis statistics and t-tests for count and interval data.

Treatment differences and changes over time were analyzed using the STATA ME package, which fits mixed-effects models (also known as multilevel models and hierarchical models) for a variety of distributions of the response conditional on normally distributed random effects (42). Mixed-effects models use all available data. The MIXED procedure was used for the continuous variables, including HDRS-17, BDI-II and Q-LES-Q scores. MELOGIT was used for categorical outcomes. With outcome measures that proved highly positively skewed, multilevel mixed-effects ordered probit regression (MEOPROBIT) models were applied. All model parameters for continuous outcome measures are presented here as partial standardized effects. Those for the categorical outcome measures are presented as conditional odds ratios (ORs).

The six time points of assessment were coded as −7 (baseline), −6 (6 months), −5 (12 months), −4 (18 months) of the review period, and −3 (24 months), −2 (30 months) and 0 (42 months) of the follow-up, in all models where 6-monthly data were available, thereby implying that regression coefficients involving time measured the linear rate of change from baseline to 42-month follow-up, and that regression intercepts referenced group differences at the last follow-up point. Models with random intercepts were initially fitted. Random slopes were added when likelihood ratio tests indicated a significant improvement of fit. In preliminary models, there was evidence of strong non-linear change effects in both patient groups. A quadratic time variable was therefore included in all models, but was removed if the likelihood ratio test indicated a non-significant improvement in fit.

Categorical outcome measures were best fitted by a logistic proportional odds random intercepts and slopes model. Continuous outcomes were best represented by a linear random intercepts and slopes model. Where data were seriously positively skewed, we fitted multilevel mixed-effects ordered probit regression models where the actual values taken on by the dependent variable were irrelevant, except that larger values were assumed to correspond to "higher" outcomes. As the LTPP group proved to be significantly better educated, despite random assignment to treatment groups, effects for all outcome measures were adjusted by additionally incorporating covariates for higher education into all fitted models. Adjusting for education also controlled for correlated observed asymmetries in employment and being in receipt of state (welfare) benefits.

Only those primary model parameters directly relevant to the study's objectives are presented here. These are: the overall significance of the model (Wald $\chi^2$ statistic); modelled (intention-to-treat) group differences at 42 months (indicating whether LTPP plus TAU was better or worse than TAU alone at the last follow-up time point); the linear rate of change from baseline to 42 months for both groups combined (indicating the extent to which participants improved or deteriorated over the 3.5 years of the study); and the differential rate of change for the LTPP group (indicating whether the rate of improvement or deterioration in this group was substantially greater than in the control group).

## RESULTS

### Baseline characteristics

The 42 patients who, after interview, declined to participate did not differ significantly from those who accepted on any clinical variable.

Table 1 summarizes pre-treatment demographic and clinical characteristics of the 129 patients who were randomized to the two treatment conditions. The majority of these patients scored within the severe range on both HDRS-17 and BDI-II. The reported average of almost four previously

**Table 1** Pre-treatment demographic and clinical characteristics of the LTPP and control groups

| | LTPP group (N=67) | Control group (N=62) |
|---|---|---|
| Age (years, mean±SD) | 42.7 ± 10.4 | 46.1 ± 9.9 |
| Gender (female, %) | 66.7 | 66.1 |
| Currently married or cohabiting (%) | 17.9 | 17.7 |
| Living alone (%) | 82.1 | 82.3 |
| Tertiary education (%)** | 59.7 | 35.5 |
| Current employment (%)* | 52.2 | 29.0 |
| Receiving state benefits (%)** | 41.8 | 64.5 |
| Duration of depressive illness (years, mean±SD) | 24.4 ± 11.6 | 19.6 ± 10.8 |
| Duration of current episode (years, mean±SD) | 3.7 ± 3.4 | 3.8 ± 2.6 |
| Previously failed treatment attempts (N, mean±SD) | 3.5 ± 1.4 | 3.9 ± 1.8 |
| Previous suicide attempts (N, mean±SD) | 0.9 ± 1.3 | 0.9 ± 1.3 |
| HDRS-17 score (mean±SD) | 19.8 ± 5.1 | 20.4 ± 4.9 |
| HDRS-17 severe or very severe depression (%) | 53.7 | 59.6 |
| HDRS-17 moderate depression (%) | 34.3 | 33.9 |
| HDRS-17 mild depression (%) | 11.9 | 6.5 |
| BDI-II score (mean±SD) | 36.5 ± 10.1 | 36.7 ± 9.5 |
| BDI-II severe depression (score >29) (%) | 74.6 | 77.4 |
| Any comorbid anxiety disorder (%) | 73.1 | 77.4 |
| Any comorbid substance use disorder (%) | 19.4 | 17.7 |
| Any comorbid eating disorder (%) | 16.4 | 9.7 |
| Current Axis I diagnoses (N, mean±SD) | 3.5 ± 1.4 | 3.2 ± 1.4 |
| GAF score (mean±SD) | 49.1 ± 7.0 | 48.8± 6.1 |
| GAF <50 (%) | 53.7 | 56.5 |
| CORE global distress score (mean±SD) | 22.8 ± 6.0 | 22.5± 6.1 |
| CORE severe distress (score >26) (%) | 44.5 | 40.0 |

LTPP – long-term psychoanalytic psychotherapy, HDRS-17 – 17-item Hamilton Depression Rating Scale, BDI-II – Beck Depression Inventory - II, GAF – Global Assessment of Functioning, CORE – Clinical Outcomes in Routine Evaluation
*p<0.02, **p<0.01

failed treatment attempts and the average GAF score <50 also highlight the considerable clinical challenge presented by this severely and chronically depressed patient group.

Patient flow is displayed in Figure 1. Attrition over four years was relatively low at 25%. Missing values were not a major problem: across all points, observations were available for 82% of primary and 75% of secondary outcome variables. There was no difference in the distribution of completer categories between the treatment groups ($\chi^2$=1.87, df=2, p=0.18).

The two groups did not differ significantly on any pre-treatment characteristics, except that patients randomized to the LTPP group had more tertiary education (p<0.01), were more often employed (p<0.02), and received fewer state benefits (p<0.02) (see Table 1). All subsequent analyses statistically controlled for this asymmetry.

## Outcomes

Complete remission (HDRS ≤8) was infrequent in both groups at the end of treatment (9.4% vs. 6.5%; $\chi^2$=0.3; p=0.59; relative risk, RR=1.4; 95% CI: 0.3-5.8; number needed to treat, NNT=34) and at 42-month follow-up (14.9% vs. 4.4%; $\chi^2$=2.9; p=0.09; RR=3.4; 95% CI: 0.7-15.6; NNT=9.6).

As shown in Table 2, partial remission (HDRS ≤12) was not significantly more likely in the LTPP than in the control group at the end of treatment (32.1% vs. 23.9%; $\chi^2$=0.8; p=0.37; RR=1.3; 95% CI: 0.6-2.5; NNT=12.3), but significant differences emerged during follow-up (at 24 months: 38.8% vs. 19.2%, $\chi^2$=4.5, p=0.03, RR=2.0, 95% CI: 1.1-4.1, NNT=5.1; at 30 months: 34.7% vs. 12.2%, $\chi^2$=6.9, p=0.008, RR=2.8, 95% CI: 1.2-6.6, NNT=4.5; at 42 months: 30.0% vs. 4.4%, $\chi^2$=10.3, p=0.001, RR=6.7, 95% CI: 1.6-28.3, NNT=3.9).

The odds of partial remission increased for both groups during the review period, but was 40% higher per 6-month period for the LTPP group. The difference between the estimated odds was significant at 24 months (Δ=1.1, 95% CI: 0.08-2.1, p=0.034); 30 months (Δ=1.5, 95% CI: 0.32-2.5, p=0.012); 36 months (Δ=1.8, 95% CI: 0.50-3.1, p=0.007) and 42 months (Δ=2.1, 95% CI: 0.64-3.6, p=0.005).

Mean HDRS-17 scores for all time points are displayed in Table 3. The difference between the group means became significant only at 24 months. The linear decrease in depression scores was significantly greater for the LTPP group (p<0.05). The model yielded a significant difference between groups at 42 months (p<0.01).

Using a cut-off point of 24 on the BDI-II for partial remission from moderate or severe depression, significantly more of the LTPP than the control group were in remission at 42 months (52.4% vs. 20.0%; $\chi^2$=9.3; p=0.002; RR=2.6; 95% CI: 1.3-5.2; NNT=3.2). The mixed-effects model analysis, which predicted self-reported remission based on all observations (intention to treat) and included adjustments for covariates, confirmed the significance of the group differences at 42 months, and the decrease in the OR was significantly steeper for the LTPP group (Table 2). Modelling individual BDI-II scores showed the linear rate of decrease to be somewhat greater for the LTPP group (p<0.05). Again,

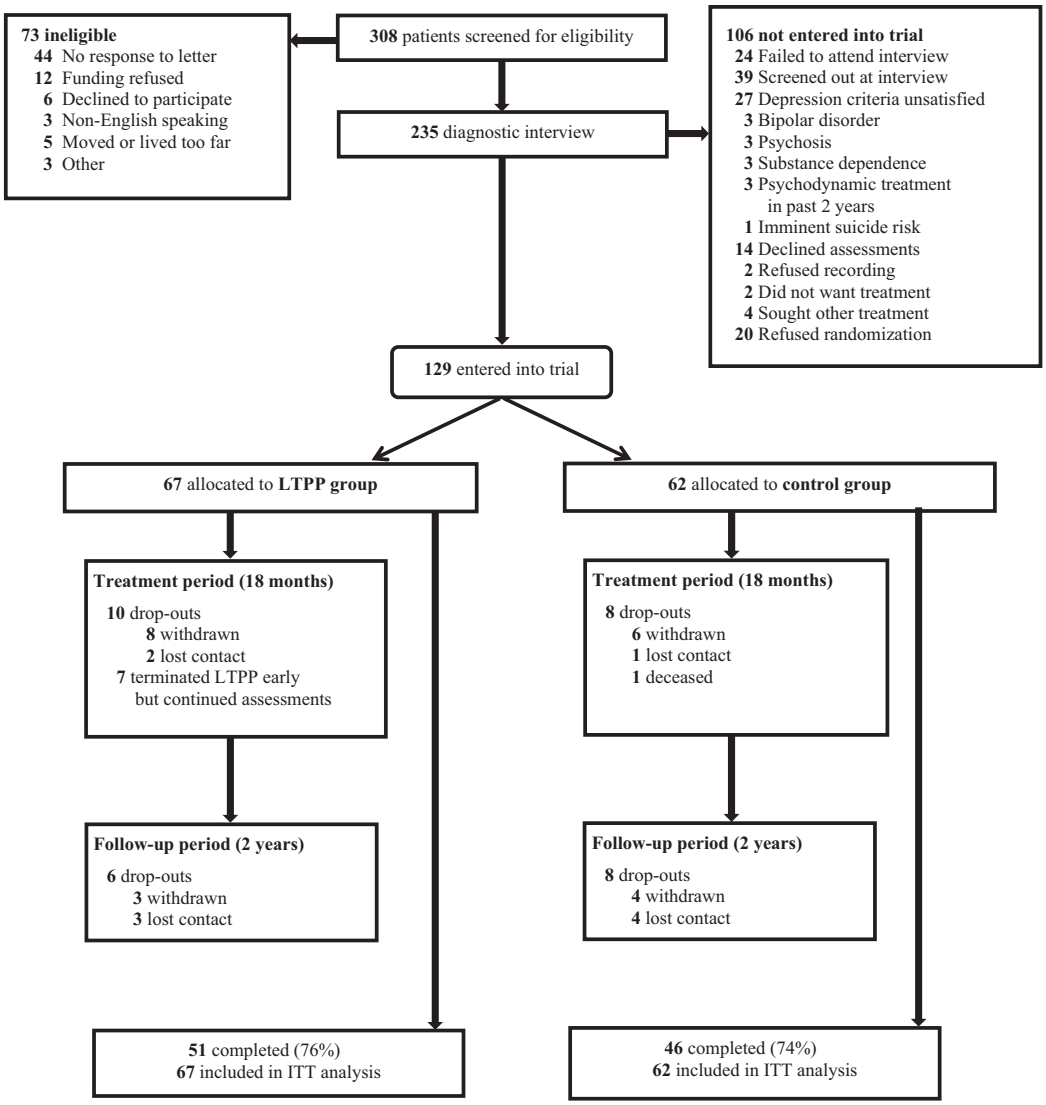


**Figure 1** CONSORT diagram of patient flow through the study. LTPP – long-term psychoanalytic psychotherapy, ITT – intention to treat

the model yielded a significant difference between groups at 42 months ($p<0.05$) (Table 3). Graphical representations of these data are available from the authors upon request.

The number of participants no longer meeting DSM-IV criteria for major depressive disorder is also shown in Table 2. Mixed-effects logistic regression indicated a significant differential change in proportional ORs across the measurement points. By 42 months, 44% of the LTPP group but only 10% of the control group were in remission ($\chi^2=14.7$; $p=0.0002$; RR=4.4; 95% CI: 1.7-10.8; NNT=2.9).

Table 4 includes the mean ratings on the GAF scale. These improved for both groups over the 18-month treatment and the two years of follow-up. Improvement in the LTPP group was greater, with a highly significant observed difference at 42 months ($t=3.3$; $p=0.001$; $d=0.69$; 95% CI: 0.26-1.11). Table 4 also shows observed and modelled improvement for both groups on self-rated subjective well-being (CORE-OM) and satisfaction with general activities (Q-LES-Q), but with substantially greater benefits accruing to the LTPP group.

## Treatments received

There were no significant between-group differences in the total number of prescribed medications, which increased from an average of just over two to over five in the course of the treatment; there were no significant reductions in these figures during the follow-up period (Table 5). As per protocol, the LTPP group received more psychoanalytic psychotherapy (average 41 hours, $p<0.0001$), while the control group received larger amounts of other types of psychosocial treat-

**Table 2** Group differences on indicators of depression (categorical measures)

| | Partial remission (HDRS-17) | | | Partial remission from moderate/severe depression (BDI-II) | | | Remission of major depression diagnosis (SCID) | | |
|---|---|---|---|---|---|---|---|---|---|
| | LTPP group | Control group | $\chi^2$ | LTPP group | Control group | $\chi^2$ | LTPP group | Control group | $\chi^2$ |
| 6 months | 12/61 (19.7%) | 6/56 (10.7%) | 1.8 | 14/48 (29.2%) | 11/39 (28.2%) | 0.0 | Not collected | | |
| 12 months | 13/56 (23.2%) | 11/52 (21.1%) | 0.1 | 21/46 (45.7%) | 7/40 (17.5%) | 7.7** | Not collected | | |
| 18 months | 17/53 (32.1%) | 11/46 (23.9%) | 0.8 | 21/45 (46.7%) | 11/39 (28.2%) | 3.0 | 20/55 (36.4%) | 6/52 (11.5%) | 9.0** |
| 24 months | 19/49 (38.8%) | 9/47 (19.2%) | 4.5* | 20/41 (48.8%) | 10/38 (26.3%) | 4.2* | 24/53 (45.3%) | 8/53 (15.1%) | 11.5*** |
| 30 months | 17/49 (34.7%) | 6/49 (12.2%) | 6.9** | 21/43 (48.8%) | 14/41 (34.1%) | 1.9 | 18/51 (35.3%) | 7/54 (13.0%) | 7.2** |
| 42 months | 14/47 (30.0%) | 2/45 (4.4%) | 10.3*** | 22/42 (52.4%) | 8/40 (20.0%) | 9.3** | 22/50 (44.0%) | 5/50 (10.0%) | 14.7*** |
| | ***Modelled odds ratios (95% CI)*** | | | ***Modelled odds ratios (95% CI)*** | | | ***Modelled odds ratios (95% CI)*** | | |
| Model: Wald $\chi^2(df=5)$ | 60.2*** | | | 49.7*** | | | 39.2*** | | |
| Linear change (both groups) | 4.67*** (2.84, 7.70) | | | 2.71*** (1.80, 4.11) | | | 4.20** (1.51, 11.40) | | |
| Quadratic change (both groups) | 0.81*** (0.76, 0.86) | | | 0.88*** (0.83, 0.93) | | | 0.79*** (0.70, 0.88) | | |
| Differential linear change (LTPP) | 1.41* (1.05, 1.89) | | | 1.33* (1.05, 1.68) | | | 2.37* (1.18, 4.84) | | |
| Group differences at 42 months | 0.09* (0.01, 0.16) | | | 0.13* (0.01, 0.24) | | | 0.22*** (0.09, 0.36) | | |

LTPP – long-term psychoanalytic psychotherapy, HDRS-17 – 17-item Hamilton Depression Rating Scale, BDI-II – Beck Depression Inventory - II, SCID – Structured Clinical Interview for DSM-IV
*p<0.05, **p<0.01, ***p<0.001

ments (average 11 hours, p<0.002), particularly counseling (27%) and CBT (19%). Patients of the control group were also significantly more likely to receive psychiatric/medical attention (37% vs. 21%).

Over follow-up, the two groups were not significantly different in terms of treatment received, although LTPP participants were slightly more likely to have received further psychodynamic psychotherapy outside the trial.

**Table 3** Group differences on indicators of depression (continuous measures)

| | HDRS-17 scores (mean±SD) | | | BDI-II scores (mean±SD) | | |
|---|---|---|---|---|---|---|
| | LTPP group | Control group | t | LTPP group | Control group | t |
| Baseline (N=129) | 19.8±5.1 | 20.2±4.8 | 0.6 | 36.5±10.1 | 36.7±9.5 | 0.2 |
| 6 months (N=117) | 16.8±6.0 | 18.3±5.8 | 1.4 | 29.9±12.4 | 32.6±15.3 | 0.9 |
| 12 months (N=108) | 17.1±6.1 | 17.9±6.3 | 0.6 | 27.4±14.5 | 34.7±13.4 | 2.4** |
| 18 months (N=99) | 16.4±6.2 | 17.9±6.5 | 1.1 | 28.0±12.8 | 34.3±16.6 | 2.1* |
| 24 months (N=96) | 15.4±6.6 | 17.6±6.1 | 1.7* | 25.9±16.4 | 34.1±16.1 | 2.3** |
| 30 months (N=98) | 16.7±7.4 | 19.4±6.5 | 1.9* | 27.0±16.0 | 31.0±15.8 | 1.3 |
| 42 months (N=92) | 15.9±6.8 | 20.1±5.4 | 3.2*** | 24.0±14.4 | 34.5±14.2 | 3.3*** |
| | ***Adjusted model coefficients (95% CI)*** | | | ***Adjusted model coefficients (95% CI)*** | | |
| Model: Wald $\chi^2(df=5)$ | 53.3*** | | | 44.4*** | | |
| Linear change (both groups) | −1.20*** (−1.64, −0.74) | | | −2.22*** (−3.26, −1.17) | | |
| Quadratic change (both groups) | 0.17*** (0.11, 0.22) | | | 0.28*** (0.14, 0.42) | | |
| Differential linear change (LTPP) | −0.36** (−0.64, −0.07) | | | −0.84* (−1.57, −0.12) | | |
| Group differences at 42 months | −2.71** (−5.16, −0.29) | | | −6.94* (−12.87, −1.00) | | |

LTPP – long-term psychoanalytic psychotherapy, HDRS-17 – 17-item Hamilton Depression Rating Scale, BDI-II – Beck Depression Inventory-II
*p<0.05, **p<0.01, ***p<0.001

**Table 4** Group differences in measures of social functioning, subjective wellbeing and satisfaction with general activities

| | GAF scores (mean±SD) | | | Subjective wellbeing, CORE-OM (mean±SD) | | | Satisfaction with general activities, Q-LES-SQ (mean±SD) | | |
|---|---|---|---|---|---|---|---|---|---|
| | LTPP group | Control group | t | LTPP group | Control group | t | LTPP group | Control group | t |
| Baseline (N=129) | 49.1±7.1 | 48.8±6.1 | 0.2 | 2.4±0.6 | 2.3±0.6 | 0.8 | 28.9±14.7 | 29.2±15.1 | 0.1 |
| 6 months (N=115) | Not collected | | | 2.2±0.7 | 2.2±0.8 | 0.2 | 36.3±15.8 | 35.3±17.5 | 0.3 |
| 12 months (N=106) | Not collected | | | 2.2±0.7 | 2.3±0.7 | 0.7 | 37.1±15.2 | 35.2±16.8 | 0.6 |
| 18 months (N=96) | 57.3±9.8 | 52.5±9.2 | 2.4** | 2.0±0.7 | 2.3±0.8 | 1.8* | 38.8±18.0 | 32.6±19.9 | 1.5 |
| 24 months (N=94) | 60.1±9.7 | 54.3±9.2 | 3.0** | 1.9±0.8 | 2.2±0.8 | 1.6* | 43.1±21.2 | 30.9±21.0 | 2.5** |
| 30 months (N=95) | 58.6±12.5 | 52.6±11.9 | 2.4** | 1.9±0.8 | 2.1±0.9 | 0.7 | 41.7±20.1 | 35.3±22.0 | 1.4 |
| 42 months (N=91) | 60.0±12.9 | 52.4±8.1 | 3.3*** | 1.8±0.8 | 2.3±0.7 | 2.9** | 45.6±19.9 | 32.0±19.0 | 3.1*** |
| | *Adjusted model coefficients (95% CI)* | | | *Adjusted model coefficients (95% CI)* | | | *Adjusted model coefficients (95% CI)* | | |
| Model: Wald $\chi^2(df=5)$ | 98.0*** | | | 29.3*** | | | 40.1*** | | |
| Linear change (both groups) | 2.29*** (1.53, 3.05) | | | −0.08** (−0.14, −0.02) | | | 2.12** (0.62, 3.62) | | |
| Quadratic change (both groups) | −0.25*** (−0.34,−0.15) | | | 0.01** (0.00, 0.02) | | | −0.29** (−0.48,−0.10) | | |
| Differential linear change (LTPP) | 0.81** (0.24, 1.38) | | | −0.06** (−0.10,−0.01) | | | 1.75*** (0.67, 2.82) | | |
| Group differences at 42 months | 6.01** (1.80, 10.22) | | | −0.32* (−0.64, −0.00) | | | 10.33** (2.46, 18.21) | | |

LTPP – long-term psychoanalytic psychotherapy, GAF – Global Assessment of Functioning, CORE-OM – Clinical Outcomes in Routine Evaluation - Outcome Measure, Q-LES-SQ – Quality of Life Enjoyment and Satisfaction Questionnaire
*p<0.05, **p<0.01, ***p<0.001

## DISCUSSION

This is the first fully randomized controlled trial of a manualized LTPP for treatment-resistant depression. Improvements in depression were modest but comparable between the LTPP and the control group until termination of treatment, while differences emerged from 24 months post-randomization, with the LTPP group mostly maintaining the gains achieved while the control group appeared to be at greater risk of relapse. At 2-year follow-up, almost one-third of the participants receiving LTPP were still in partial remission, compared with only 4% of those in the control group. At that time, 44% of the LTPP group no longer met diagnostic criteria for major depressive disorder, compared with 10% of those receiving TAU alone.

The effect sizes observed are in the medium range. The long-term outcomes of LTPP compare favorably with effect sizes reported in comprehensive reviews (3), including those used by the UK treatment recommendations (33). Studies that show stronger effects tend to observe patients in whom treatment resistance is less evident and lack information about long-term outcomes (43). Further comparisons, including longer manualized treatments based upon other (non-psychoanalytic) psychological therapy modalities such as CBT, are needed to establish the specificity of the therapeutic gain reported here.

As predicted, differences between the LTPP and the control group increased during follow-up on most measures. A Finnish longitudinal study of LTPP has reported a similar pattern with a less chronically depressed patient group (44), suggesting that LTPP may require some time post-treatment for its full effects to become evident (45). End-of-treatment evaluations or follow-ups that are too short may miss the emergence of this delayed therapeutic benefit.

While this study has ecological validity in that it employed a relatively unselected sample and incorporated a comparatively long follow-up, it has several limitations. First, the design of the study did not allow masking of patients to the treatment allocation, which may have generated an expectation bias. Second, although mixed-effects models are thought to be robust even to selective loss of data (46), we still failed to collect primary outcome data for over 25% of patients at 42 months, despite an unusually good level of retention for patients with depression of this severity. Third, the differences between the effects associated with the two treatments could have arisen as a result of the disparities between their respective numbers of contact hours, intensity, and quality of case management (47). Fourth, in spite of robust procedures, randomization yielded a difference between groups in education level, with associated asymmetries in employment and state benefits, which we were forced to adjust for statistically. Reanalysis in which the samples were balanced by selectively excluding patients did not alter the basic pattern of findings. Fifth, while we were concerned to measure outcome over an extended period, we omitted to include an interval depression measure such as the Longitudinal Interval Follow-up Evaluation (48). Sixth, since the study was planned and conducted by the developers of the inter-

**Table 5** Treatments delivered to patients of LTPP and control groups in periods before randomization (6 months), during treatment (18 months) and during follow-up (24 months)

| | Period before randomization | | | Treatment period | | | Follow-up period | | |
|---|---|---|---|---|---|---|---|---|---|
| | LTPP group | Control group | t or $\chi^2$ | LTPP group | Control group | t or $\chi^2$ | LTPP group | Control group | t or $\chi^2$ |
| **Medications** | | | | | | | | | |
| Antidepressants (%) | 82.0 | 80.7 | <1 | 85.0 | 79.0 | <1 | 79.0 | 74.2 | <1 |
| Anxiolytics/hypnotics (%) | 41.8 | 45.2 | <1 | 40.3 | 41.9 | <1 | 34.3 | 35.5 | <1 |
| Antipsychotics/ mood stabilizers (%) | 9.0 | 3.2 | <1 | 11.9 | 11.3 | <1 | 13.4 | 16.1 | <1 |
| Analgesics (%) | 37.3 | 40.3 | <1 | 35.8 | 41.9 | <1 | 29.9 | 41.9 | $\chi^2$=2.05 |
| Other medications (%) | 23.9 | 30.6 | <1 | 23.9 | 33.9 | $\chi^2$=1.57 | 28.4 | 37.1 | $\chi^2$=1.12 |
| No medication (%) | 9.0 | 6.5 | <1 | 7.5 | 6.4 | <1 | 15.0 | 9.7 | <1 |
| Number of medications (mean±SD) | 2.1 ± 1.4 | 2.0 ± 1.2 | <1 | 5.0 ± 4.2 | 5.3 ± 3.9 | <1 | 4.6 ± 4.4 | 5.2 ± 4.1 | <1 |
| **Psychosocial treatments** | | | | | | | | | |
| Psychodynamic psychotherapy (hours, mean±SD) | 0.8 ± 6.3 | 0 | <1 | 41.4 ± 21.4 | 0.4 ± 3.0 | $t$=15.0*** | 3.6 ± 11.0 | 0.8 ± 6.6 | $t$=1.7 |
| Other therapies (hours, mean±SD) | 6.2 ± 11.3 | 7.7 ± 14.7 | <1 | 3.2 ± 11.6 | 11.2 ± 18.4 | $t$=2.98*** | 6.2 ± 11.5 | 8.1 ± 16.2 | <1 |
| CBT (%) | 9.0 | 8.1 | <1 | 1.5 | 19.4 | $\chi^2$=11.4*** | 10.5 | 8.1 | <1 |
| Counseling (%) | 37.3 | 42.0 | <1 | 1.5 | 27.4 | $\chi^2$=18.1*** | 16.4 | 17.7 | <1 |
| Clinical psychologist (%) | 22.4 | 17.7 | <1 | 11.9 | 14.5 | <1 | 13.4 | 11.3 | <1 |
| Psychotherapist (%) | 12.0 | 13.0 | <1 | 7.5 | 11.2 | $\chi^2$=2.4 | 16.4 | 23.6 | $\chi^2$=2.5 |
| **Other interventions** | | | | | | | | | |
| Psychiatric/medical (hours, mean±SD) | 2.3 ± 5.2 | 0.5 ± 1.9 | $t$=2.56** | 1.3 ± 3.7 | 1.5 ± 3.1 | <1 | 1.2 ± 3.5 | 1.8 ± 6.1 | <1 |
| Psychiatric/medical (%) | 31.3 | 16.1 | $\chi^2$=4.1* | 20.9 | 37.1 | $\chi^2$=4.1* | 22.4 | 27.4 | <1 |
| Social worker/OT/nurse (%) | 9.0 | 9.7 | <1 | 9.0 | 9.7 | <1 | 7.5 | 6.5 | <1 |
| Self-help groups (%) | 4.5 | 4.8 | <1 | 4.5 | 4.8 | <1 | 4.5 | 4.8 | <1 |
| Day centre (%) | 0 | 0 | <1 | 1.5 | 1.6 | <1 | 1.5 | 0 | <1 |
| Hospital admissions (%) | 4.0 | 0 | $\chi^2$=3.82 | 3.0 | 1.6 | <1 | 4.5 | 4.9 | <1 |

LTPP – long-term psychoanalytic psychotherapy, CBT – cognitive-behavioral therapy, OT – occupational therapy
*p<0.05, **p<0.01, ***p<0.001

vention, there is a risk of allegiance bias (49). We tried to minimize this risk by having the primary outcome measure assessments made by interviewers who were blinded to the treatment condition. Seventh, these results were delivered by a single provider organization. This may limit generalizability. However, a multi-center German trial (the LAC Study) (50), testing LTPP using the same manual with a similar patient group, will shortly report.

In conclusion, while the benefit of both interventions for this severely affected group of patients with major depressive disorder was limited, a moderate difference emerged over long-term follow-up in favor of the LTPP condition. Further studies are needed to replicate this finding, ascertain its clinical utility, understand the mechanisms involved, and identify factors associated with response or non-response to treatment.

## Acknowledgements

The study was supported by the Tavistock Clinic Charitable Foundation and the Tavistock & Portman NHS Foundation Trust, plus a small grant from the International Psychoanalytic Association. P. Fonagy was in receipt of a National Institute of Health Research Senior Investigator Award (NF-SI-0510-10228), which provided support for the study. The authors are indebted to the patients who participated and to all the research assistants, interns, students, administrators

and clinicians who worked on the trial. They are grateful to the distinguished scientists of the Trial's Steering Committee, A. Faulkner, S. Blake, M. Buszewicz, J. Cape, P. McCrone, M. Knapp and I. Nazareth. Finally, they wish to recognize the contribution of P. Richardson, the trial's original principal investigator, who sadly died in 2007.

## References


1. Moussavi S, Chatterji S, Verdes E et al. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. Lancet 2007;370:851-8.
2. Greden JF. The burden of disease for treatment-resistant depression. J Clin Psychiatry 2001;62(Suppl. 16):26-31.
3. Huhn M, Tardy M, Spineli LM et al. Efficacy of pharmacotherapy and psychotherapy for adult psychiatric disorders: a systematic overview of meta-analyses. JAMA Psychiatry 2014;71:706-15.
4. Stimpson N, Agrawal N, Lewis G. Randomised controlled trials investigating pharmacological and psychological interventions for treatment-refractory depression. Systematic review. Br J Psychiatry 2002;181:284-94.
5. Thomas L, Kessler D, Campbell J et al. Prevalence of treatment-resistant depression in primary care: cross-sectional data. Br J Gen Pract 2013;63:e852-8.
6. Kubitz N, Mehra M, Potluri RC et al. Characterization of treatment resistant depression episodes in a cohort of patients from a US commercial claims database. PLoS One 2013;8:e76882.
7. McIntyre RS, Filteau MJ, Martin L et al. Treatment-resistant depression: definitions, review of the evidence, and algorithmic approach. J Affect Disord 2014;156:1-7.
8. Spijker J, van Straten A, Bockting CL et al. Psychotherapy, antidepressants, and their combination for chronic major depressive disorder: a systematic review. Can J Psychiatry 2013;58: 386-92.
9. Carvalho AF, Berk M, Hyphantis TN et al. The integrative management of treatment-resistant depression: a comprehensive review and perspectives. Psychother Psychosom 2014;83:70-88.
10. Lam RW, Chan P, Wilkins-Ho M et al. Repetitive transcranial magnetic stimulation for treatment-resistant depression: a systematic review and metaanalysis. Can J Psychiatry 2008;53:621-31.
11. Daban C, Martinez-Aran A, Cruz N et al. Safety and efficacy of Vagus Nerve Stimulation in treatment-resistant depression. A systematic review. J Affect Disord 2008;110:1-15.
12. Morishita T, Fayad SM, Higuchi MA et al. Deep brain stimulation for treatment-resistant depression: systematic review of clinical outcomes. Neurotherapeutics 2014;11:475-84.
13. Trivedi RB, Nieuwsma JA, Williams JW Jr. Examination of the utility of psychotherapy for patients with treatment resistant depression: a systematic review. J Gen Intern Med 2011;26:643-50.
14. Hollon SD, DeRubeis RJ, Fawcett J et al. Effect of cognitive therapy with antidepressant medications vs antidepressants alone on the rate of recovery in major depressive disorder: a randomized clinical trial. JAMA Psychiatry 2014;71:1157-64.
15. Hollon SD, Ponniah K. A review of empirically supported psychological therapies for mood disorders in adults. Depress Anxiety 2010;27:891-932.
16. Rawlins M. De Testimonio: on the evidence for decisions about the use of therapeutic interventions. Clin Med 2008;8:579-88.
17. Driessen E, Van HL, Don FJ et al. The efficacy of cognitive-behavioral therapy and psychodynamic therapy in the outpatient treatment of major depression: a randomized clinical trial. Am J Psychiatry 2013;170:1041-50.
18. Fonagy P. The effectiveness of psychodynamic psychotherapies: an update. World Psychiatry 2015;14:137-50.
19. Leichsenring F, Klein S. Evidence for psychodynamic psychotherapy in specific mental disorders: a systematic review. Psychoanal Psychother 2014;28:4-32.
20. de Maat S, de Jonghe F, de Kraker R et al. The current state of the empirical evidence for psychoanalysis: a meta-analytic approach. Harv Rev Psychiatry 2013;21:107-37.
21. Knekt P, Lindfors O, Laaksonen MA et al. Quasi-experimental study on the effectiveness of psychoanalysis, long-term and short-term psychotherapy on psychiatric symptoms, work ability and functional capacity during a 5-year follow-up. Helsinki Psychotherapy Study Group. J Affect Disord 2011;132:37-47.
22. Leichsenring F, Rabung S. Long-term psychodynamic psychotherapy in complex mental disorders: update of a meta-analysis. Br J Psychiatry 2011;199:15-22.
23. Shedler J. The efficacy of psychodynamic psychotherapy. Am Psychol 2010;65:98-109.
24. Beck AT, Steer R, Brown G. Manual for Beck Depression Inventory - II (BDI-II). San Antonio: Psychological Corporation, 1996.
25. Taylor D, Carlyle JA, McPherson S et al. Tavistock Adult Depression Study (TADS): a randomised controlled trial of psychoanalytic psychotherapy for treatment-resistant/treatment-refractory forms of depression. BMC Psychiatry 2012;12:60.
26. First MB, Spitzer RL, Gibbon M et al. User's Guide for the Structured Clinical Interview for DSM-IV Axis I Disorders - Research Version. New York: Biometrics Research, New York State Psychiatric Institute, 2001.
27. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960;23:56-62.
28. Taylor D. Treatment manuals and the advancement of psychoanalytic knowledge: the treatment manual of the Tavistock Adult Depression Study. Int J Psychoanalysis 2015;96:845-75.
29. McPherson S, Armstrong D. Negotiating 'depression' in primary care: a qualitative study. Soc Sci Med 2009;69:1137-43.
30. Andrews G. Reducing the burden of depression. Can J Psychiatry 2008;53:420-7.
31. Milton J. Psychoanalysis and cognitive behaviour therapy – rival paradigms or common ground? Int J Psychoanal 2001;82:431-47.
32. Jones EE. Therapeutic action: a guide to psychoanalytic therapy. Northvale: Jason Aronson, 2000.
33. National Institute for Health and Clinical Excellence. Depression in adults: the treatment and management of depression in adults. Clinical Guideline 90. London: National Institute for Health and Clinical Excellence, 2009.
34. Beecham JK, Knapp MRJ. Costing psychiatric interventions. In: Thornicroft G, Brewin C, Wing JK (eds). Measuring mental health needs. London: Gaskell/Royal College of Psychiatrists, 1992:163-83.
35. Reimherr FW, Amsterdam JD, Quitkin FM et al. Optimal length of continuation therapy in depression: a prospective assessment during long-term fluoxetine treatment. Am J Psychiatry 1998;155: 1247-53.
36. Frank E, Prien RF, Jarrett RB et al. Conceptualization and rationale for consensus definitions of terms in major depressive disorder. Remission, recovery, relapse, and recurrence. Arch Gen Psychiatry 1991;48:851-5.
37. Hilsenroth MJ, Ackerman SJ, Blagys MD et al. Reliability and validity of DSM-IV axis V. Am J Psychiatry 2000;157:1858-63.
38. Evans C, Mellor-Clark J, Margison F et al. CORE: Clinical Outcomes in Routine Evaluation. J Ment Health 2000;9:247-55.
39. Endicott J, Nee J, Harrison W et al. Quality of Life Enjoyment and Satisfaction Questionnaire: a new measure. Psychopharmacol Bull 1993;29:321-6.
40. Statacorp. Stata statistical software: Release 14. College Station: StataCorp LP, 2015.
41. Bateman A, Fonagy P. Randomized controlled trial of outpatient mentalization-based treatment versus structured clinical management for borderline personality disorder. Am J Psychiatry 2009; 166:1355-64.

42. Rabe-Hesketh S, Skrondal A. Multilevel and longitudinal modeling using STATA, 3rd ed. College Station: Stata Press, 2012.
43. Keller MB, McCullough JP, Klein DN et al. A comparison of nefazodone, the cognitive behavioral-analysis system of psychotherapy, and their combination for the treatment of chronic depression. N Engl J Med 2000;342:1462-70.
44. Knekt P, Lindfors O, Harkanen T et al. Randomized trial on the effectiveness of long-and short-term psychodynamic psychotherapy and solution-focused therapy on psychiatric symptoms during a 3-year follow-up. Helsinki Psychotherapy Study Group. Psychol Med 2008;38:689-703.
45. Muratori F, Picchi L, Bruni G et al. A two-year follow-up of psychodynamic psychotherapy for internalizing disorders in children. J Am Acad Child Adolesc Psychiatry 2003;42:331-9.
46. Verbeke G, Molenberghs G. Linear mixed models for longitudinal data. New York: Springer, 2000.
47. Rossouw TI, Fonagy P. Mentalization-based treatment for self-harm in adolescents: a randomized controlled trial. J Am Acad Child Adolesc Psychiatry 2012;51:1304-13.
48. Keller MB, Lavori PW, Friedman B et al. The Longitudinal Interval Follow-up Evaluation. A comprehensive method for assessing outcome in prospective longitudinal studies. Arch Gen Psychiatry 1987;44:540-8.
49. Luborsky L, Diguer L, Seligman DA et al. The researcher's own therapy allegiances: a "wild card" in comparisons of treatment efficacy. Clin Psychol Sci Pract 1999;6:95-106.
50. Beutel ME, Leuzinger-Bohleber M, Ruger B et al. Psychoanalytic and cognitive-behavior therapy of chronic depression: study protocol for a randomized controlled trial. Trials 2012; 13:117.

DOI 10.1002/wps.20267

# At risk or not at risk? A meta-analysis of the prognostic accuracy of psychometric interviews for psychosis prediction

Paolo Fusar-Poli[1,2], Marco Cappucciati[1], Grazia Rutigliano[1], Frauke Schultze-Lutter[3], Ilaria Bonoldi[1], Stefan Borgwardt[4], Anita Riecher-Rössler[4], Jean Addington[5], Diana Perkins[6], Scott W. Woods[7], Thomas H. McGlashan[7], Jimmy Lee[8], Joachim Klosterkötter[9], Alison R. Yung[10], Philip McGuire[1,2]

[1]King's College London, Institute of Psychiatry, London, UK; [2]OASIS Service, South London and Maudsley NHS Foundation Trust, London, UK; [3]University Hospital of Child and Adolescent Psychiatry and Psychotherapy, University of Bern, Bern, Switzerland; [4]University of Basel Psychiatric Clinics, Basel, Switzerland; [5]Hotchkiss Brain Institute, Department of Psychiatry, University of Calgary, Calgary, Alberta, Canada; [6]Department of Psychiatry, University of North Carolina, Chapel Hill, NC, USA; [7]Department of Psychiatry, Yale University, New Haven, CT, USA; [8]Department of General Psychiatry, Institute of Mental Health, Singapore, Singapore; [9]Department of Psychiatry and Psychotherapy, University of Cologne, Cologne, Germany; [10]Institute of Brain, Behaviour and Mental Health, University of Manchester, Manchester, UK

*An accurate detection of individuals at clinical high risk (CHR) for psychosis is a prerequisite for effective preventive interventions. Several psychometric interviews are available, but their prognostic accuracy is unknown. We conducted a prognostic accuracy meta-analysis of psychometric interviews used to examine referrals to high risk services. The index test was an established CHR psychometric instrument used to identify subjects with and without CHR (CHR+ and CHR−). The reference index was psychosis onset over time in both CHR+ and CHR− subjects. Data were analyzed with MIDAS (STATA13). Area under the curve (AUC), summary receiver operating characteristic curves, quality assessment, likelihood ratios, Fagan's nomogram and probability modified plots were computed. Eleven independent studies were included, with a total of 2,519 help-seeking, predominately adult subjects (CHR+: N=1,359; CHR−: N=1,160) referred to high risk services. The mean follow-up duration was 38 months. The AUC was excellent (0.90; 95% CI: 0.87-0.93), and comparable to other tests in preventive medicine, suggesting clinical utility in subjects referred to high risk services. Meta-regression analyses revealed an effect for exposure to antipsychotics and no effects for type of instrument, age, gender, follow-up time, sample size, quality assessment, proportion of CHR+ subjects in the total sample. Fagan's nomogram indicated a low positive predictive value (5.74%) in the general non-help-seeking population. Albeit the clear need to further improve prediction of psychosis, these findings support the use of psychometric prognostic interviews for CHR as clinical tools for an indicated prevention in subjects seeking help at high risk services worldwide.*

**Key words:** Psychosis, prevention, psychometric interviews, high risk services, prognostic accuracy

***(World Psychiatry 2015;14:322–332)***

Treatments for psychosis have been in wide use for nearly half a century, yet there is little evidence that they have substantially improved outcomes (1). Therefore, indicated preventive intervention is the main paradigm yielding new hope for impacting the course of psychosis (2). However, this intervention requires an accurate identification of individuals at clinical high risk (CHR), that relies on the use of accurate prognostic tools to detect psychosis as early as possible, so that its progress can be arrested and, if possible, reversed.

Prognostic testing is commonly used in preventive medicine (3). While a screening test should identify all individuals who may develop the disease (4), a prognostic test is used to predict the development or not of the future disease when a patient shows some heralding signs or symptoms. Examples of predictive testing in somatic medicine include fasting glucose and oral glucose tolerance test and glycated haemoglobin to detect subjects at high risk for diabetes (pre-diabetes or intermediate hyperglycaemia) (5). Pre-diabetes closely resembles the CHR state, in that only about 5-10% of people per year will progress to diabetes, with the same proportion converting back to normoglycaemia (5).

No biological tests such as those used to detect pre-diabetes are available in clinical psychiatry (6). Therefore, for an indicated prevention of psychosis, prognostic testing is usually accomplished by administration of specific psychometric interviews, which assess validated CHR criteria (7). These instruments include: the Comprehensive Assessment of At Risk Mental State (CAARMS, 8,9), the Structured Interview for Psychosis-Risk Syndrome (SIPS, 10) and the Basel Screening Instrument for Psychosis (BSIP, 11) for the assessment of "ultra-high risk" (UHR) criteria (12); and the Bonn Scale for the Assessment of Basic Symptoms (BSABS, 13) and the Schizophrenia Proneness Instruments (Adult version, SPI-A, 14, and Child & Youth version, SPI-CY, 15) for the assessment of basic symptom (BS) criteria (16).

The UHR criteria include attenuated psychotic symptoms (APS), brief limited intermittent psychotic symptoms (BLIPS) and trait vulnerability plus a marked decline in psychosocial functioning (genetic risk and functional deterioration syndrome: GRFD). The two partially overlapping BS criteria rely on subjectively experienced disturbances of perception, thinking, language and attention (17).

These CHR instruments show excellent reliability when used by trained raters: the overall inter-rater agreement is 0.95 for the SIPS (18), 0.85 for the CAARMS (12) and 0.91 for the SPI-A (19). Yet, their prognostic accuracy is still uncertain. For an ideal instrument, all subjects actually about to develop psychosis should be classified as "at risk" (CHR+) while those suffering from other complaints not leading to frank psychosis should be classified as "not at risk" (CHR−).

The prognostic accuracy of a test can be quantified by different measures – sensitivity (Se), specificity (Sp), summary receiver operating characteristic (SROC) curves, area under the curve (AUC) – whose evaluation requires follow-up not only of CHR+ but also of CHR− subjects. So far, no robust meta-analysis has addressed the consistency and magnitude of the prognostic accuracy of psychometric CHR testing, and the few available studies reported inconsistent prognostic accuracy findings (18,20). Because of this, the overall clinical utility (i.e., predictive value) of psychometric interviews in help-seeking and non-help-seeking subjects is still unknown.

Predictive values are not fixed indicators of a test performance, but are affected by the prevalence of the condition (4). Within help-seeking CHR+ samples, the ability of the above psychometric instruments to identify true positives is accumulating to 29% at 2-year follow-up (21,22) – a finding comparable to other preventive approaches in medicine (23). On the contrary, the predictive value and potential clinical utility of these instruments in samples with a lower prevalence of the condition, such as the general population, still await results from follow-ups (24-26). Similarly, the predictive value in other samples with a variable psychosis risk, such as unselected adolescents with psychiatric problems (27), subjects accessing public treatment services, psychiatric patients in forensic units (28), primary care patients, genetic high risk samples, prisoners, post-partum women, people with 22q11.2 deletion syndrome, users of high potency cannabis, military, black ethnic minorities, refugees, people with borderline personality disorders or epilepsy, is still largely unknown.

To overcome this lack of knowledge, we conducted the first robust meta-analysis to examine the consistency and magnitude of the prognostic accuracy of instruments used for psychosis prediction, while at the same time investigating their potential clinical utility in help-seeking samples of high risk services, in the general population and across other groups.

## METHODS

### Search strategy

Two investigators (MC, GR) conducted a two-step literature search. At a first step, the Web of Knowledge database was searched, incorporating both the Web of Science and MEDLINE. The search was extended until March 2015, including only abstracts in English. The electronic research adopted several combinations of the following keywords: "at risk mental state", "psychosis risk", "prodrome", "prodromal psychosis", "ultra-high risk", "high risk", "help-seeking", "diagnostic accuracy", "sensitivity", "specificity", "psychosis prediction", "psychosis onset", and name of the CHR assessment instruments. The second step involved the use of Scopus to investigate citations of previous systematic reviews on transition outcomes in CHR subjects and a manual search of the reference lists of the retrieved articles.

Articles identified through these two steps were then screened for the selection criteria on the basis of abstract reading. The articles surviving this selection were assessed for eligibility on the basis of full text reading. To achieve a high standard of reporting, we adopted the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) checklist (29).

### Selection criteria

Studies were eligible for inclusion if: a) they were reported in original articles, written in English or in German; b) they had used in the same pool of referrals an established CHR psychometric instrument (index test); c) they had followed up both CHR+ and CHR− subjects for psychosis onset (reference index) using established international diagnostic manuals (ICD or DSM); d) they had reported sufficient prognostic accuracy data. With respect to this last point, when data were not directly presented, they were indirectly extracted from associated data. Additionally, we contacted all corresponding authors to request additional data when needed.

We excluded: a) abstracts, pilot datasets, reviews, articles in a language other than English or German; b) studies in which interviews were not conducted in the same pool of referrals or that used an external CHR− group of healthy controls; c) studies with overlapping datasets. In case of multiple publications deriving from the same study population, we selected the article reporting the largest and most recent data set. The literature search was summarized according to PRISMA guidelines (30).

### Recorded variables

Data extraction was independently performed by two investigators (MC, GR). Data included author, year of publication, characteristics of subject samples (baseline sample sizes, mean age and age range, proportion of females), the CHR psychometric instrument used, exposure to antipsychotics, diagnostic criteria used at follow-ups to assess the psychotic outcome, follow-up time, prognostic accuracy data (number of true and false positives, true and false negatives or associated data) and quality assessment conducted

with the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) checklist (31).

### Statistical analysis

The statistical analysis followed the Cochrane Guidelines for Systematic Reviews of Diagnostic Test Accuracy, Version 1.0 (32) and the Methods Guide for Authors of Systematic Reviews of Medical Tests by the Agency for Healthcare Research and Quality (chapter 8) (33). Evaluating test accuracy requires knowledge of two quantities: the test's Se and Sp. Meta-analysis methods for diagnostic test accuracy thus have to deal with two summary statistics simultaneously rather than one (32). Methods for undertaking analyses which account for both Se and Sp, the relationship between them, and the heterogeneity in test accuracy, require fitting advanced hierarchical random effects models (32).

For each study we constructed a two-by-two table, which included true positive, false positive, true negative, and false negative values. When studies reported different data at different follow-up times, we used data from the longest follow-up. The baseline sample size was conservatively used as the base reference to avoid a bias towards overly high transition risks at longer follow-ups and related higher drop-out rates of transition negatives.

Data were then analyzed with MIDAS (Meta-analytical Integration of Diagnostic Accuracy Studies) (34), a comprehensive program of statistical and graphical routines for undertaking meta-analysis of diagnostic/prognostic test performance in STATA 13 software. The index tests of CHR status (CHR+ or CHR−) and reference tests of transition to psychosis according to international diagnostic manuals (ICD or DSM as gold standard) were dichotomous.

Primary data synthesis was performed within the bivariate mixed-effects regression framework for the logit transforms of Se and Sp (34). In addition to accounting for study size, the bivariate model estimates and incorporates the intrinsic negative correlation that may arise between Se and Sp within studies (threshold effect) (35), as a result of differences in the test threshold between studies (36). The bivariate model allows for heterogeneity beyond chance as a result of clinical and methodological differences between studies (36).

We estimated the summary Se and Sp and the estimated hierarchical SROC curves (32). A SROC graph across each predictor, with the y-axis representing the predictor's Se and the x-axis representing 1-specificity, was used to plot around the summary estimates a 95% confidence region and a 95% prediction region to illustrate the precision with which the summary values were estimated (confidence ellipse of a mean), and to show the amount of between-study variation (prediction ellipse; the likely range of values for a new study). We also estimated the AUC. Finally, for sensitivity analyses of the impact of follow-up times, supplementary analyses were conducted by grouping the data at each specific time point of 6, 12, 24 and $\geq$30 months.

Heterogeneity across studies was assessed using the $I^2$, with values of 25%, 50% and 75% representing mild, moderate and severe inconsistency, respectively (37). Within MIDAS, forest plots and heterogeneity statistics can be created for each test performance parameter individually or may be displayed as paired plots. Subgroups analyses and meta-regressions were used to examine the influence of CHR instruments used, mean age, gender (% females), follow-up time, sample size, exposure to antipsychotics, and quality assessment (QUADAS) on meta-analytical estimates. To control for biases associated with imbalanced datasets (38), we further tested the impact of the proportion of CHR+ subjects in the overall samples. The meta-regressions were used if there was substantial heterogeneity ($I^2 > 50\%$) (39).

Model diagnostic analyses included quantile plot of residual based goodness-of-fit; chi-squared probability plot of squared Mahalanobis distances for assessment of the bivariate normality assumption; spike plot for checking for particularly influential observations using Cook's distance; a scatter plot for checking for outliers using standardized predicted random effects (standardized level-2 residuals) (34). Sensitivity analyses (i.e., exclusion of outliers and rerunning of the model) were conducted to further explore heterogeneity. We did not test publication bias (40), because no proven statistical method exists for this type of meta-analysis (41).

In a second step, we employed the probability-modifying plot and the Fagan's nomogram to estimate the clinical or patient-relevant utility of the CHR interview in subjects seeking help at early detection services, in the general population, as well as in other samples (i.e., genetic high risk samples, prisoners, post-partum women, people with 22q11.2 deletion syndrome, users of high potency cannabis, military, black ethnic minorities, people with borderline personality disorders, and unselected psychiatric samples).

The clinical utility was evaluated using the positive and negative likelihood ratios (LR+ and LR−) to calculate post-test probability (PostTP) based on Bayes' theorem (with pre-test probability, PrePT, being the prevalence of the condition in the target population), as follows: $\text{PostTP} = \text{LR} \times \text{PreTP}/[(1-\text{PreTP}) + (\text{PreTP} \times \text{LR})]$ (35). Specifically, the probability-modifying plot (34) is a graphical sensitivity analysis of the test's predictive values across a baseline psychosis risk continuum in people seeking help at early detection services. It depicts separate curves for positive and negative tests and uses general summary statistics (i.e., unconditional positive and negative predictive values, NPV and PPV, which permit underlying psychosis risk heterogeneity) to evaluate the effect of the CHR assessment on predictive values (42). The PreTP probability of psychosis risk in subjects seeking help at early detection services was computed in the current dataset as the proportion of subjects developing psychosis on the total baseline sample (CHR+ plus CHR−) (34).

Fagan's nomogram, a two-dimensional graphical tool for estimating how much the result of a test changes the pre-test probability that a patient will develop psychosis, was used

**Table 1** Studies included in the meta-analysis

| Study | QUADAS score (max. 14); exposure to antipsychotics at baseline | Predictor (index test) | Psychosis diagnosis (reference standard) | Age (years, mean±SD, range) | Gender (% females) | Follow-up (months) | CHR+ subjects (baseline) | CHR− subjects (baseline) |
|---|---|---|---|---|---|---|---|---|
| Klosterkötter et al (52) | 14; No | BSABS (BS) | DSM-IV | 29.3±10.0 (15-53) | 47.5 | 0, ≥30 | 110 | 50 |
| Yung et al (45) | 12; Yes (% NA) | CAARMS (UHR) | CAARMS | 18.1 (15-24) | 51.0 | 0, 6, 24 | 119 | 173 |
| Riecher-Rössler et al (11) | 13.5; No | BSIP (UHR plus 4th criterion) | BPRS | 26.8±8.9 (18-60) | 41.4 | 0, 6, 12, 24, ≥30 | 58 | 32 |
| Woods et al (20) | 13.5; Yes (11.6%) | SIPS (UHR) | DSM-IV or medical records | 17.8±4.4 (12-36) | 39.5 | 0, 6, 12, 24 | 259 | 111 |
| Addington et al (48) | 13.5; Yes (1.8%) | SIPS (UHR) | DSM-IV | 19.8±4.5 (12-31) | 47.8 | 0, 6, 12, 24 | 172 | 100 |
| Liu et al (49) | 2.5; Yes (79.7%) | SIPS (UHR) | DSM-IV | 21.4±4.0 (16-24) | 47.7 | 0, 24 | 59 | 48 |
| Simon et al (50) | 6; No | SIPS/SPI-A (BS/UHR) | DSM-IV | 21.0 (14-40) | 32.4 | 0, 12, 24 | 99 | 49 |
| Lee et al (44) | 13; No | CAARMS (UHR) | DSM-IV | 21.6±3.5 (14-29) | 39.9 | 0, 6, 12, 24, ≥30 | 173 | 494 |
| Schultze-Lutter et al (46) | 13; Yes (13.8%) | SPI-A/SIPS (BS/UHR) | DSM-IV | 24.9±6.0 (15-39) | 37.0 | 0, 6, 12, 24, ≥30 | 194 | 52 |
| Kotlicka-Antczak et al (47) | 11.5; Yes (10.2%) | CAARMS (UHR) | ICD-10 | 19.0±3.6 (15-29) | 51.1 | ≥30 | 94 | 33 |
| Spada et al (51) | 11; No | CAARMS (UHR) | DSM-IV | 15.8±1.7 (12-17) | 47.5 | 0, 6 | 22 | 18 |

QUADAS – Quality Assessment of Diagnostic Accuracy Studies checklist, CHR – clinical high risk, UHR – ultra-high risk, BS – basic symptoms, BSABS – Bonn Scale for the Assessment of Basic Symptoms, BPRS – Brief Psychiatric Rating Scale, BSIP – Basel Screening Instrument for Psychosis, CAARMS – Comprehensive Assessment of At Risk Mental State, SIPS – Structured Interview for Prodromal Syndromes, SPI-A – Schizophrenia Proneness Instrument, NA – not available

to estimate the clinical value of psychometric CHR interview in the general population and in the other samples. Again, the clinical value is calculated on the LR+ and LR− obtained from the current meta-analysis (43) and using the pre-test psychosis risk in the different samples as estimated from the available literature.

Statistical tests were two-sided and statistical significance was defined as p values $<0.05$.

## RESULTS

### Database

The literature review (PRISMA flow chart available from the authors upon request) produced eleven independent studies that met the inclusion criteria, for a total of 2,519 subjects (CHR+: N=1,359; CHR−: N=1,160) referred to high risk services (Table 1). The proportion of CHR+ subjects in the total sample was 0.54%, revealing an overall balanced dataset.

Four studies employed the CAARMS, three the SIPS, one the BSIP, one the BSABS, and two both the SIPS and the SPI-A. The mean follow-up time was 37.72 months (SD 27.81, median=33). QUADAS ratings ranged from 2.5 to 14 (the latter is the highest possible score). The main reasons for a non-optimal rating were (partial) exposure to antipsychotics and unsatisfactory reporting of results.

### Prognostic accuracy of CHR interview

Across the eleven studies interviewing help-seeking subjects for CHR symptoms, the summary meta-analytical estimate of Se and the AUC were outstanding, while the estimate of Sp was poor (Figure 1). There was moderate to substantial heterogeneity for Se ($I^2=51$, $p=0.02$) and severe heterogeneity for Sp ($I^2=95$, $p<0.001$), 17% of which was due to threshold effects.

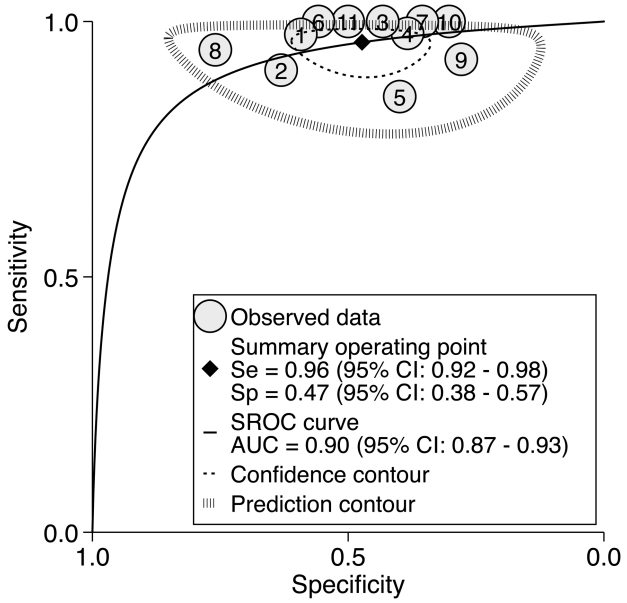


**Figure 1** Meta-analytical summary receiver operating characteristic (SROC) curve of clinical high risk (CHR) psychometric interviews. Se – sensitivity, Sp – specificity, AUC – area under the curve, 1 – Klosterkötter et al (52), 2 – Yung et al (45), 3 – Riecher-Rössler et al (11), 4 – Woods et al (20), 5 – Addington et al (48), 6 – Liu et al (49); 7 – Simon et al (50), 8 – Lee et al (44), 9 – Schultze-Lutter et al (46), 10 – Kotlicka-Antczak et al (47), 11 – Spada et al (51)

Sensitivity analyses revealed that the two studies with the highest proportion of CHR− subjects in the total sample had the highest Sp (44,45), while the two studies with the lowest proportion of CHR− subjects had the lowest Sp (46,47). However, meta-regression analyses showed that the proportion of CHR+ subjects in the total sample had no impact on the overall AUC (38).

Across SIPS samples (20,46,48-50), Se was 0.96 (95% CI: 0.88-0.99) and Sp was 0.39 (95% CI: 0.32-0.46). Across CAARMS samples (44,45,47,51), Se was 0.96 (95% CI: 0.82-0.99) and Sp was 0.56 (95% CI: 0.38-0.73). There were not enough data to perform subgroups meta-analyses in BSIP samples (11), BSABS/SPI-A samples (46,50,52) and samples combining the SIPS and SPI-A (46).

Meta-regression analyses revealed no significant effects for mean age, gender, follow-up time, sample size and quality assessment (QUADAS), but there was a significant effect for exposure to antipsychotics at baseline (p=0.04). This effect was driven by a significant decrease of Se (0.94) in the five studies where subjects were exposed to antipsychotics as compared to the six studies where subjects were not exposed (Se=0.98).

Model diagnostics revealed a good fit of the model and indicated that one study was close to the outlier threshold (44). Sensitivity analyses confirmed a very good AUC (0.84) after this study was removed from the dataset.

Supplementary analyses were conducted grouping the available samples at specific time points of 6, 12, 24 and ≥30 months. The AUCs were outstanding at each time point: at 6 months (seven samples, AUC=0.97, 95% CI: 0.95-0.98), at 12 months (six samples, AUC=0.94, 95% CI: 0.92-0.96), at 24 months (eight samples, AUC=0.94, 95% CI: 0.92-0.96), and at ≥30 months (seven samples, AUC=0.91, 95% CI: 0.88-0.93).

## Clinical utility of psychometric CHR interviews in subjects seeking help at high risk services

The 38-month psychosis risk in the 2,519 help-seeking subjects was 15% (95% CI: 0.9%-24%). On the basis of this prior distribution, the continuous relationship between PreTP and PostTP probability is summarized in Figure 2. Being CHR+ was associated with a 26% (95% CI: 23%-30%) risk of developing psychosis within 38 months, yet a small LR+ of just 1.82 (95% CI: 1.52-2.18), while being CHR− was associated with a 1.56% (95% CI: 0.7%-2.42%) risk of developing psychosis and a large LR− of 0.09 (CI 95%: 0.04-0.18) (Figure 3).

## Estimated clinical utility of psychometric CHR interviews in the general population and in other samples

Based on a lifetime prevalence of all non-organic psychotic disorders of 3.27% (53) and the above LRs, Fagan's

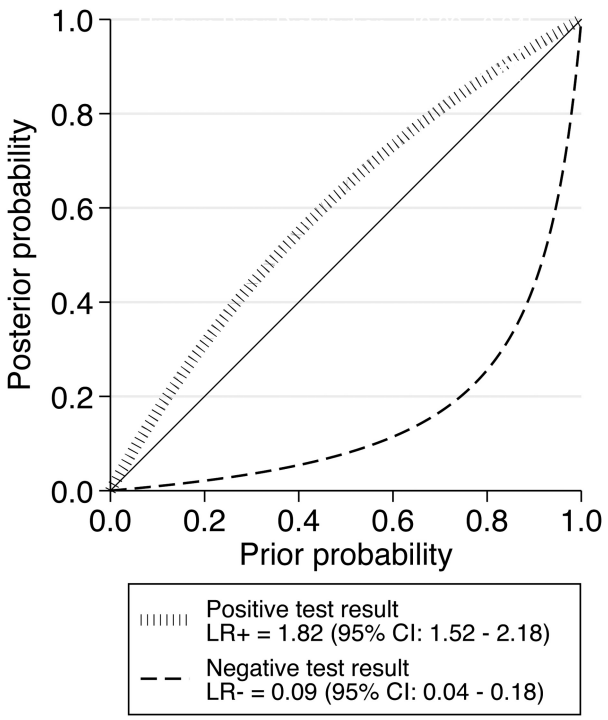


**Figure 2** Meta-analytical probability modifying plot, illustrating the relationship between pre-test probability (PreTP) (9 to 24% psychosis risk at 38 months in subjects seeking help at early detection services) and post-test probability (PostTP) (psychosis risk at 38 months in help-seeking subjects based on clinical high risk psychometric interviews), computed as the likelihood of a positive (above diagonal line; LR+) or negative (below diagonal line, LR−) test result over the 0-1 range of PreTP

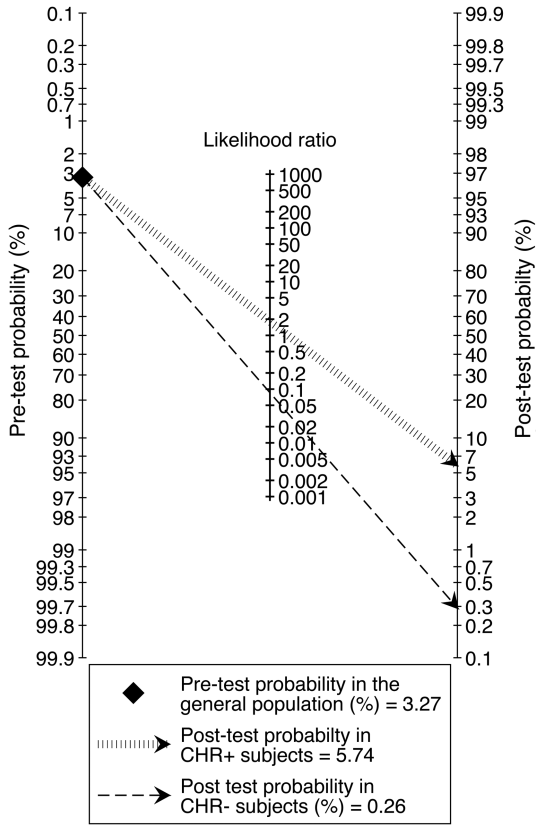


**Figure 3** Fagan's nomogram illustrating the meta-analytical clinical value (post-test probability) of clinical high risk (CHR) psychometric interviews in the general population in order to predict risk of psychosis at 38 months, given an assumed psychosis risk (pre-test probability) of 3.27%, as reported in a nationally representative sample of general population subjects aged 30-44 years (see 53)

nomogram revealed only limited clinical utility for CHR instruments in the general population. Testing positive for CHR was associated with a 5.74% lifetime risk of developing psychosis, while testing negative was associated with hardly any such risk (0.26%). Corresponding figures for other clinical and non-clinical samples are displayed in Table 2.

## DISCUSSION

This is the first study to present a robust and elaborated meta-analytical estimate of the prognostic accuracy of CHR psychometric interviews for psychosis prediction. Assessing help-seekers referred to a high risk service with a CHR interview generally revealed an excellent overall prognostic performance in terms of the AUC at 38-month follow-up (values of 0.9-1.0 are considered outstanding, of 0.8-0.9 excellent and of 0.7-0.8 acceptable) (66), which is comparable to other preventive approaches in medicine. However, excellent AUC values were mainly mediated by an outstanding ability of the instruments to rule out psychosis (i.e., very satisfyingly low LR− and high Se), at an expense of their ability to rule in psychosis (i.e., unsatisfactory low LR+ and only moderate overall Sp), which indicates some need to further improve prediction. On the contrary, the clinical utility of current CHR instruments in non-help-seeking subjects in the general population was estimated to be low.

Our first aim was to investigate at meta-analytical level the overall prognostic accuracy of CHR instruments in determining the risk of developing psychosis at 38 months in young help-seeking subjects referred to high risk services. We first estimated the AUC, which serves as a global measure of test performance and indicates the overall goodness

**Table 2** Estimated clinical utility of clinical high risk psychometric instruments for psychosis prediction in various populations

| Sample | Psychosis risk | Positive test result | Negative test result |
|---|---|---|---|
| Unselected psychiatric adolescents (27) | 3.13% (12 mo.) | 3.13% | 0.29% |
| Subjects in contact with public treatment services (54) | 0.35% (lifetime) | 0.63% | <0.001% |
| Psychiatric patients in forensic units (55) | 74% (lifetime) | 83.38% | 20.39% |
| Primary care patients (56) | 0.045% (per year) | <0.001% | <0.001% |
| Prisoners (57) | 3.90% (lifetime) | 6.87% | 0.36% |
| Post-partum women (58) | 4% (12 mo.) | 7.04% | 0.37% |
| 22q11.2 deletion syndrome (59) | 16% (48 mo.) | 25.74% | 1.68% |
| Young adults at familial risk for psychosis (60) | 12% (30 mo.) | 19.88% | 1.21% |
| Users of high potency cannabis (61) | 24% (lifetime) | 36.49% | 2.76% |
| Military (62) | 0.014% (per year) | <0.001% | <0.001% |
| Black ethnic minority (63) | 1.45% (lifetime) | 2.60% | 0.13% |
| Refugees (64) | 3.3% (lifetime) | 5.84% | 0.31% |
| Epilepsy (65) | 5.6% (lifetime) | 9.74% | 0.53% |

**Table 3** Prognostic accuracy of indicated prevention tests in clinical medicine

| At-risk population | Outcome | Diagnostic test | Sensitivity (follow-up) | Specificity (follow-up) | AUC (follow-up) |
|---|---|---|---|---|---|
| Patients presenting for CHR evaluation | Psychosis | CHR interview | 0.96 (2 yrs.) | 0.47 (2 yrs.) | 0.89 (2 yrs.) |
| Men at risk for prostate cancer | Prostate cancer | PSA (72,73) | 0.69 (5 yrs.) | 0.89 (5 yrs.) | 0.88 (5 yrs.) |
| Men at risk for colorectal cancer | Colorectal cancer | Risk prediction model (74) | NA (5 yrs.) | NA (5 yrs.) | 0.80 (5 yrs.) |
| Women at risk for colorectal cancer | Colorectal cancer | Risk prediction model (74) | NA (5 yrs.) | NA (5 yrs.) | 0.73 (5 yrs.) |
| Patients with transient ischemic attack | Stroke | ABCD2 score (75,76) | 0.57 (30 days) | 0.32 (30 days) | 0.72 (7 days) |
| Patients with stable coronary disease | Coronary event | Framingham risk score + number of diseased vessels (77) | NA (8.5 yrs.) | NA (8.5 yrs.) | 0.67 (77) (8.5 yrs.) |
| Pre-diabetes | Diabetes | 30-min plasma glucose (78) | 0.91 (9 yrs.) | 0.39 (9 yr.) | 0.67 (9 yrs.) |
| Mild cognitive impairment | Alzheimer's disease | ADAS-cog subscale (79) | 0.62 (1 yr.) | 0.73 (1 yr.) | 0.67 (1 yr.) |
| Women at risk for breast cancer | ER-positive invasive breast cancer | Gail model (80) | 0.50 (5 yrs.) | 0.65 (5 yrs.) | 0.60 (5 yrs.) |

CHR – clinical high risk, AUC – area under the curve, PSA – prostate specific antigen, ER – estrogen receptor, NA – not available, ADAS-cog – Alzheimer Disease Assessment Scale-cognitive part

of a diagnostic tests. Thereby, we adopted a robust methodological approach following international guidelines for diagnostic/prognostic accuracy meta-analysis, to avoid the serious flaws observed in a previous meta-analytical attempt, such as overlapping samples, missing studies and lack of control for several moderators (67,68). Our finding of consistent prognostic accuracy across CHR instruments is particularly important, given the significant differences in their criteria (69). This evidence of a negligible role of the CHR assessment instrument (i.e., CAARMS vs. SIPS) is in line with our previous meta-analysis, which found no differences in pooled annual transition risks between these instruments (21). This finding was also confirmed by a second independent meta-analysis (22).

We further revealed that, despite an excellent overall prognostic accuracy, there is a need to specifically improve the ability to rule in subsequent psychosis, i.e., to improve LR+ and Sp, while preserving the outstanding ability to rule it out. This is particularly relevant given that interviewing subjects seeking help at high risk services is particularly difficult: these individuals are assumed to lay on an upper mid-range of a symptomatic continuum by showing mild and often infrequent symptoms of yet some clinical significance already (24).

However, differentiating between such gradual symptoms with specific tests or interviews is not a problem specific to psychosis prevention or other preventive approaches in psychiatry. For example, in case of the at-risk state of diabetes, the World Health Organization (WHO) proposed the use of the term "intermediate hyperglycaemia" (i.e., pre-diabetes) to accurately reflect the observation that glycaemia is a continuous variable and that their defined categories are based on somewhat arbitrary decisions on where to draw a line between normality and abnormality (70). Similarly to the different cut-offs and criteria used to identify CHR subjects, the definition of pre-diabetes is based on cut-off points for glycaemia (5) for which there are different operationalizations (e.g., by WHO and by the American Diabetes Association) (5). Furthermore, as for the CHR state (7), progression to diabetes is not inevitable in pre-diabetes; some individuals, in the absence of any intervention, may remain in that state or even revert to normoglycaemia (5). Because of this, various risk assessment tools based on socio-demographic or questionnaire data are available to identify subjects with pre-diabetes, and their overall prognostic accuracy is comparable to our meta-analytical estimates, such as the AUC=0.76 reported for the Cambridge risk score (71). More broadly, the overall prognostic accuracy of the CHR instruments was comparable if not superior to various other medical tests used for an indicated prevention (Table 3).

However, it is important to highlight that the high AUC of CHR instruments is secondary to an accurate training of raters and ongoing close supervision provided by expert clinicians (7). Thus, a recent guidance on the early detection

of psychosis explicitly recommends CHR assessment to be conducted in specialized centres by well-trained raters and/or clinical supervision by such raters (22).

The imbalance between an excellent Se (0.96) and an only modest Sp (0.47) may have some relevant clinical implications, when considering that we have selectively included only studies discriminating CHR+ from CHR− within the same pool of help-seeking subjects. Since these patients were seeking help at or were subsequently referred to early detection services and frequently presented also with psychosocial and functional impairment (81) and other non-psychotic symptoms (82) and disorders (83), the use of CHR assessments should not be thought of as identifying and treating an unselected and asymptomatic group at risk of a poor outcome (universal prevention) (84). Rather, the use of CHR assessment follows the approach of an indicated prevention, which is concerned with detecting a disease in its earliest stages, before frank symptoms appear, and with intervening to slow or stop its progression into the full-blown medical picture. Therefore, the above-mentioned recent guidance explicitly restricts CHR assessment to the clients of mental health services (22).

With regard to the potential CHR+ misdiagnosis of persons who do not in fact develop psychosis, or the potential CHR− misdiagnosis of persons who will develop psychosis, the low Sp suggests a stepped and multi-component strategy. In a first sensitivity-preserving step, CHR instruments could be used to rule out true negatives, i.e. subjects who are unlikely to develop psychosis. In a second step, additional clinical, neurocognitive, biological or combined models of risk stratification could be applied to the CHR+ group, with the aim of increasing Sp and prognostic reliability. This would enable risk stratification and personalized treatments accordingly (85,86).

We further estimated the clinical utility of CHR assessments in other clinical and non-clinical populations, as clinical utility is affected by the underlying psychosis risk in a population. We found that testing positive for CHR was associated with a 26% risk of developing psychosis within 38 months, a proportion comparable with our previous meta-analysis (95% CI: 23-35) (21) of transition risks in CHR+ subjects. This was due to a small LR+ of 1.82. We could also show here for the first time that being CHR− was associated with only a 1.56% risk of developing the illness, corresponding to a large LR− of 0.09. It is important to note that the PostTP, as estimated from the likelihood ratio and PreTP, is generally more accurate than if estimated from the PPV of the test. In fact, with the help of these two measures (LR+ and LR−), it was possible to estimate the PostTP in different settings characterized by a variable PrePT of psychosis risk, which however will still require empirical studies.

We clearly estimated for the first time a limited clinical utility of CHR interviews in the general population, revealing only a small and inadequate PPV of 5.74%. This estimate is in line with meta-analytical results indicating that self-reported psychotic-like experiences in the young non-help-seeking general population are associated with a negligible risk of transitioning to psychotic disorders over time (87). Yet, as self-reported psychotic experiences are only a poor estimate of clinician-assessed CHR symptoms, these findings might not reflect the true predictive power of CHR criteria in the community. Similarly, it appears there is no scope to use psychometric CHR interviews in unselected psychiatric adolescent samples, patients accessing public treatment or primary care services, patients admitted to forensic units, post-partum women, ethnic minorities, military, refugees, patients with epilepsy and prisoners. The latter finding is in line with a recent study indicating that the CHR state does not predict psychosis in adolescent delinquent samples (28). On the other hand, our estimates provide some support for the clinical utility of CHR assessments in subjects with two psychotic relatives, in patients with 22q11.2 deletion syndrome and in subjects using high potency cannabis, as well as for preventive trials already proposed in some of these clinical samples (88).

The additional novel finding is that our probability-modifying plot allows future power calculation studies in samples characterized by an underlying variable psychosis risk that is ranging from 0 to 1. For example, with our plot available, researchers may draw a vertical line from the selected pre-test probability of the sample to the appropriate likelihood ratio line and then read the post-test probability off the vertical scale.

Some limitations of this meta-analysis should be acknowledged. First, because of the limited statistical power, we were unable to directly compare the prognostic accuracy of different psychometric instruments. However, subgroups analyses revealed comparable SIPS vs. CAARMS AUCs. Furthermore, two independent meta-analyses (21,22) did not reveal any significant impact of the type of psychometric instrument employed on risk estimates. Also, we were unable to explain all the observed heterogeneity across individual studies. However, some of this was accounted for by threshold effects and the effect of antipsychotics exposure on Se. An effect of age, with lower transition risks in younger CHR+ subjects, was observed in our first meta-analysis (21) and recently confirmed in another re-analysis (22). Such an age effect might have been missed in our analyses, as only the by far smallest of the included studies, with an only 6-month follow-up (69), was on minors only.

Furthermore, the individual studies included here varied with respect to follow-up time, although meta-regression did not reveal any significant effect of this variable. We additionally conducted supplementary analyses at each specific time point, and these analyses confirmed excellent AUCs. Furthermore, there is new meta-analytical evidence that, in UHR samples, transition to psychosis is most likely to occur within the first 2 years after presentation to clinical services, with a stable plateau after 36 months (89). Since our mean follow-up time (38 months) falls in this plateau period, follow-up had no significant impact on the meta-analytical estimates across samples mainly at risk for UHR criteria.

## CONCLUSIONS

The present prognostic accuracy meta-analysis indicated that currently used interviews for psychosis prediction have an excellent overall prognostic performance. This supports their use as clinical tools for an indicated prevention in subjects seeking help at mental health services worldwide, provided raters have undergone adequate training, while discouraging their use for prevention in non-help-seeking subjects in the general population.


### Acknowledgements

P. Fusar-Poli was supported in part by a 2014 NARSAD Young Investigator Award. The last two authors contributed equally to this work.


### References


1. Insel TR. Rethinking schizophrenia. Nature 2010;468:187-93.
2. McGorry PD. Early clinical phenotypes, clinical staging, and strategic biomarker research: building blocks for personalized psychiatry. Biol Psychiatry 2013;74:394-5.
3. Arbyn M, Verdoodt F, Snijders P et al. Accuracy of human papillomavirus testing on self-collected versus clinician-collected samples: a meta-analysis. Lancet Oncol 2014;15:172-83.
4. Michel C, Schultze-Lutter F, Schimmelmann BG. Screening instruments in child and adolescent psychiatry: general and methodological considerations. Eur Child Adolesc Psychiatry 2014;23:725-7.
5. Tabak AG, Herder C, Rathmann W et al. Prediabetes: a high-risk state for diabetes development. Lancet 2012;379:2279-90.
6. Kapur S, Phillips AG, Insel TR. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? Mol Psychiatry 2012;17:1174-9.
7. Fusar-Poli P, Borgwardt S, Bechdolf A et al. The psychosis high-risk state: a comprehensive state-of-the-art review. JAMA Psychiatry 2013;70:107-20.
8. Yung AR, Phillips LJ, Yuen HP et al. Comprehensive Assessment of at Risk Mental State. Parkville: PACE Clinic, ORYGEN Research Centre, University of Melbourne, 2006.
9. Yung AR, Yuen HP, McGorry PD et al. Mapping the onset of psychosis: the Comprehensive Assessment of At-Risk Mental States. Aust N Zeal J Psychiatry 2005;39:964-71.
10. McGlashan T, Walsh B, Woods S. The psychosis-risk syndrome: handbook for diagnosis and follow-up. Oxford: Oxford University Press, 2010.
11. Riecher-Rössler A, Aston J, Ventura J et al. The Basel screening instrument for psychosis (BSIP): development, structure, reliability and validity. Fortschr Neurol Psychiatr 2008;76:207-16.
12. Yung AR, Stanford C, Cosgrave E et al. Testing the ultra high risk (prodromal) criteria for the prediction of psychosis in a clinical sample of young people. Schizophr Res 2006;84:57-66.
13. Klosterkotter J, Gross G, Huber G et al. Evaluation of the Bonn Scale for the Assessment of Basic Symptoms - BSABS as an instrument for the assessment of schizophrenia proneness: a review of recent findings. Neurol Psychiatry Brain Res 1997;5: 137-50.
14. Schultze-Lutter F, Addington J, Ruhrmann S et al. Schizophrenia Proneness Instrument, Adult version (SPI-A). Rome: Fioriti, 2007.
15. Schultze-Lutter F, Koch E. Schizophrenia Proneness Instrument, Child and Youth version (SPI-CY). Rome: Fioriti, 2010.
16. Schultze-Lutter F, Ruhrmann S, Fusar-Poli P et al. Basic symptoms and the prediction of first-episode psychosis. Curr Pharm Des 2012;18:351-7.
17. Schultze-Lutter F. Subjective symptoms of schizophrenia in research and the clinic: the basic symptom concept. Schizophr Bull 2009;35:5-8.
18. Miller TJ, McGlashan TH, Rosen JL et al. Prodromal assessment with the Structured Interview for Prodromal Syndromes and the Scale of Prodromal Symptoms: predictive validity, interrater reliability, and training to reliability. Schizophr Bull 2003;29:703-15.
19. Schultze-Lutter F, Klosterkötter J, Picker H et al. Predicting first-onset of psychosis by basic symptom criteria. Clin Neuropsychiatry 2007;4:11-22.
20. Woods SW, Addington J, Cadenhead KS et al. Validity of the prodromal risk syndrome for psychosis: findings from North American Prodrome Longitudinal Study. Schizophr Bull 2009;35:894-908.
21. Fusar-Poli P, Bonoldi I, Yung AR et al. Predicting psychosis: a meta-analysis of transition outcomes in individuals at high clinical risk. Arch Gen Psychiatry 2012;69:1-10.
22. Schultze-Lutter F, Michel C, Schmidt SJ et al. EPA guidance on the early detection of clinical high risk states of psychoses. Eur Psychiatry 2015;30:405-16.
23. Fusar Poli P, Carpenter W, Woods SW et al. Attenuated psychosis syndrome: ready for DSM-5.1? Annu Rev Clin Psychol 2014;10: 155-92.
24. Schultze-Lutter F, Michel C, Ruhrmann S et al. Prevalence and clinical significance of DSM-5-attenuated psychosis syndrome in adolescents and young adults in the general population: the Bern Epidemiological At-Risk (BEAR) Study. Schizophr Bull 2014;40: 1499-08.
25. Kelleher I, Murtagh A, Molloy C et al. Identification and characterization of prodromal risk syndromes in young adolescents in the community: a population-based clinical interview study. Schizophr Bull 2012;38:239-46.
26. Schimmelmann B, Michel C, Martz-Irngartinger A et al. Age matters in the prevalence and clinical significance of ultra-high-risk for psychosis symptoms and criteria in the general population: findings from the BEAR and BEARS-Kid studies. World Psychiatry 2015;14:189-97.
27. Lindgren M, Manninen M, Kalska H et al. Predicting psychosis in a general adolescent psychiatric sample. Schizophr Res 2014;158:1-6.
28. Manninen M, Lindgren M, Therman S et al. Clinical high-risk state does not predict later psychosis in a delinquent adolescent population. Early Interv Psychiatry 2014;8:87-90.
29. Stroup DF, Berlin JA, Morton SC et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000;283:2008-12.
30. Moher D, Liberati A, Tetzlaff J et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009;339:b2535.
31. Whiting P, Rutjes AW, Reitsma JB et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003;3:25.
32. Macaskill P, Gatsonis C, Deeks J et al. Cochrane handbook for systematic reviews of diagnostic test accuracy Version 1.0. In: Deeks JJ, Bossuyt PM, Gatsonis C (eds) The Cochrane Collaboration, 2010.
33. Smetana GW, Umscheid CA, Chang S et al. Methods guide for authors of systematic reviews of medical tests: a collaboration between the Agency for Healthcare Research and Quality (AHRQ) and the Journal of General Internal Medicine. In: Chang SM, Matchar DB, Smetana GW et al (eds). Methods guide for medical test reviews. Rockville: Agency for Healthcare Research and Quality, 2012.

34. Dwamena BA. MIDAS: computational and graphical routines for meta-analytical integration of diagnostic accuracy studies in Stata. Ann Arbor: Division of Nuclear Medicine, Department of Radiology, University of Michigan Medical School, 2007.
35. Harbord R, Whiting P. Meta-analysis of diagnostic accuracy using hierarchical logistic regression. Stata J 2009;9:211-29.
36. Janda S, Shahidi N, Gin K et al. Diagnostic accuracy of echocardiography for pulmonary hypertension: a systematic review and meta-analysis. Heart 2011;97:612-22.
37. Lipsey M, Wilson D. Practical meta-analysis. Thousand Oaks: Sage Publications, 2000.
38. Bekkar M, Djemaa H, Alitouche T. Evaluation measures for models assessment over imbalanced data sets. J Inf Eng Appl 2013;3:27-38.
39. Higgins JP, Thompson SG, Deeks JJ et al. Measuring inconsistency in meta-analyses. BMJ 2003;327:557-60.
40. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol 2005;58:882-93.
41. Wanders LK, East JE, Uitentuis SE et al. Diagnostic performance of narrowed spectrum endoscopy, autofluorescence imaging, and confocal laser endomicroscopy for optical diagnosis of colonic polyps: a meta-analysis. Lancet Oncol 2013;14:1337-47.
42. Li J, Fine J, Safdar N. Prevalence-dependent diagnostic accuracy measures. Stat Med 2007;26:3258-73.
43. Fagan TJ. Letter: Nomogram for Bayes theorem. N Engl J Med 1975;293:257.
44. Lee J, Rekhi G, Mitter N et al. The Longitudinal Youth at Risk Study (LYRIKS) – an Asian UHR perspective. Schizophr Res 2013;151:279-83.
45. Yung AR, Nelson B, Stanford C et al. Validation of "prodromal" criteria to detect individuals at ultra high risk of psychosis: 2 year follow-up. Schizophr Res 2008;105:10-17.
46. Schultze-Lutter F, Klosterkotter J, Ruhrmann S. Improving the clinical prediction of psychosis by combining ultra-high risk criteria and cognitive basic symptoms. Schizophr Res 2014;154: 100-6.
47. Kotlicka-Antczak M, Pawelczyk T, Rabe-Jablonska J et al. PORT (Programme of Recognition and Therapy): the first Polish recognition and treatment programme for patients with an at-risk mental state. Early Interv Psychiatry (in press).
48. Addington J, Piskulic D, Perkins DO et al. Affect recognition in people at clinical high risk of psychosis. Schizophr Res 2012;140: 87-92.
49. Liu CC, Lai MC, Liu CM et al. Follow-up of subjects with suspected pre-psychotic state in Taiwan. Schizophr Res 2011;126: 65–70.
50. Simon AE, Gradel M, Cattapan-Ludewig K et al. Cognitive functioning in at-risk mental states for psychosis and 2-year clinical outcome. Schizophr Res 2012;142:108-15.
51. Spada G, Molteni S, Pistone C et al. Identifying children and adolescents at ultra high risk for psychosis in Italian neuropsychiatry services: a feasibility study. Eur Child Adolesc Psychiatry (in press).
52. Klosterkötter J, Hellmich M, Steinmeyer EM et al. Diagnosing schizophrenia in the initial prodromal phase. Arch Gen Psychiatry 2001;58:158-64.
53. Perala J, Suvisaari J, Saarni SI et al. Lifetime prevalence of psychotic and bipolar I disorders in a general population. Arch Gen Psychiatry 2007;64:19-28.
54. Morgan VA, McGrath JJ, Jablensky A et al. Psychosis prevalence and physical, metabolic and cognitive co-morbidity: data from the second Australian national survey of psychosis. Psychol Med 2013;23:1-14.
55. Andreasson H, Nyman M, Krona H et al. Predictors of length of stay in forensic psychiatry: the influence of perceived risk of violence. Int J Law Psychiatry 2014;37:635-42.
56. Hardoon S, Hayes JF, Blackburn R et al. Recording of severe mental illness in United Kingdom primary care, 2000-2010. PLoS One 2013;8:e82365.
57. Fazel S, Seewald K. Severe mental illness in 33,588 prisoners worldwide: systematic review and meta-regression analysis. Br J Psychiatry 2012;200:364-73.
58. Vesga-Lopez O, Blanco C, Keyes K et al. Psychiatric disorders in pregnant and postpartum women in the United States. Arch Gen Psychiatry 2008;65:805-15.
59. Gothelf D, Schneider M, Green T et al. Risk factors and the evolution of psychosis in 22q11.2 deletion syndrome: a longitudinal 2-site study. J Am Acad Child Adolesc Psychiatry 2013;52:1192-203.
60. Johnstone EC, Ebmeier KP, Miller P et al. Predicting schizophrenia: findings from the Edinburgh High-Risk Study. Br J Psychiatry 2005;186:18-25.
61. Di Forti M, Marconi A, Carra E et al. Proportion of patients in south London with first-episode psychosis attributable to use of high potency cannabis: a case-control study. Lancet Psychiatry (in press).
62. Cowan DN, Weber NS, Fisher JA et al. Incidence of adult onset schizophrenic disorders in the US military: patterns by sex, race and age. Schizophr Res 2011;127:235-40.
63. Qassem T, Bebbington P, Spiers N et al. Prevalence of psychosis in black ethnic minorities in Britain: analysis based on three national surveys. Soc Psychiatry Psychiatr Epidemiol 2015;50:1057-64.
64. Llosa AE, Ghantous Z, Souza R et al. Mental disorders, disability and treatment gap in a protracted refugee setting. Br J Psychiatry 2014;204:208-13.
65. Clancy MJ, Clarke MC, Connor DJ et al. The prevalence of psychosis in epilepsy; a systematic review and meta-analysis. BMC Psychiatry 2014;14:75.
66. Hosmer W, Lemeshow S. Applied survival analysis: regression modeling of time to event data. New York: Wiley, 1999.
67. Mitchell AJ. Predicting the development of schizophrenia. Br J Psychiatry 2012;200:254.
68. Fusar-Poli P. Predicting the development of schizophrenia. Br J Psychiatry 2012;200:254-5.
69. Schultze-Lutter F, Schimmelmann BG, Ruhrmann S et al. 'A rose is a rose is a rose', but at-risk criteria differ. Psychopathology 2013; 46:75-87.
70. World Health Organization. Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia: Report of a WHO/IDF Consultation. Geneva: World Health Organization, 2006.
71. Thomas C, Hypponen E, Power C. Type 2 diabetes mellitus in midlife estimated from the Cambridge Risk Score and body mass index. Arch Intern Med 2006;166:682-88.
72. Mettlin C, Murphy GP, Babaian RJ et al. The results of a five-year early prostate cancer detection intervention. Investigators of the American Cancer Society National Prostate Cancer Detection Project. Cancer 1996;77:150-9.
73. Littrup PJ, Kane RA, Mettlin CJ et al. Cost-effective prostate cancer detection. Reduction of low-yield biopsies. Investigators of the American Cancer Society National Prostate Cancer Detection Project. Cancer 1994;74:3146-58.
74. Shin A, Joo J, Yang HR et al. Risk prediction model for colorectal cancer: National Health Insurance Corporation study, Korea. PLoS One 2014;9:e88079.
75. Ghia D, Thomas P, Cordato D et al. Low positive predictive value of the ABCD2 score in emergency department transient ischaemic attack diagnoses: the South Western Sydney Transient Ischaemic Attack Study. Int Med J 2012;42:913-8.
76. Giles MF, Rothwell PM. Systematic review and pooled analysis of published and unpublished validations of the ABCD and ABCD2 transient ischemic attack risk scores. Stroke 2010;41:667-73.
77. Sugamata W, Nakamura T, Uematsu M et al. The combined assessment of flow-mediated dilation of the brachial artery and brachial-ankle pulse wave velocity improves the prediction of

future coronary events in patients with chronic coronary artery disease. J Cardiol 2014;64:179-84.
78. Kim YA, Ku EJ, Khang AR et al. Role of various indices derived from an oral glucose tolerance test in the prediction of conversion from prediabetes to type 2 diabetes. Diabetes Res Clin Pract 2014; 106:351-9.
79. Rozzini L, Vicini Chilovi B, Bertoletti E et al. The importance of Alzheimer disease assessment scale-cognitive part in predicting progress for amnestic mild cognitive impairment to Alzheimer disease. J Geriatr Psychiatry Neurol 2008;21:261-7.
80. Chlebowski RT, Anderson GL, Lane DS et al. Predicting risk of breast cancer in postmenopausal women by hormone receptor status. J Natl Cancer Inst 2007;99:1695-705.
81. Fusar-Poli P, Rocchetti M, Sardella A et al. Disorder, not just a state of risk: meta-analysis of functioning and quality of life in subjects at high clinical risk for psychosis. Br J Psychiatry (in press).
82. Fusar-Poli P, Nelson B, Valmaggia L et al. Comorbid depressive and anxiety disorders in 509 individuals with an at risk mental state: impact on psychopathology and transition to psychosis Schizophr Bull 2014;40:120-31.
83. Fusar-Poli P, Byrne M, Badger S et al. Outreach and support in South London (OASIS), 2001-2011: ten years of early diagnosis and treatment for young individuals at clinical high risk for psychosis. Eur Psychiatry 2012;28:315-26.
84. Yung AR, Woods SW, Ruhrmann S et al. Whither the attenuated psychosis syndrome? Schizophr Bull 2012;38:1130-4.
85. Ruhrmann S, Schultze-Lutter F, Salokangas RK et al. Prediction of psychosis in adolescents and young adults at high risk: results from the prospective European prediction of psychosis study. Arch Gen Psychiatry 2010;67:241-51.
86. Koutsouleris N, Riecher-Rössler A, Meisenzahl EM et al. Detecting the psychosis prodrome across high-risk populations using neuroanatomical biomarkers. Schizophr Bull 2015;41:471-82.
87. Fusar-Poli P, Yung AR, McGorry P et al. Lessons learned from the psychosis high-risk state: towards a general staging model of prodromal intervention. Psychol Med 2014;44:17-24.
88. Armando M, De Crescenzo F, Vicari S et al. Indicated prevention with long-chain polyunsaturated omega-3 fatty acids in patients with 22q11DS genetically at high risk for psychosis. Protocol of a randomized, double-blind, placebo-controlled treatment trial. Early Interv Psychiatry (in press).
89. Kempton M, Bonoldi I, Valmaggia L et al. Speed of psychosis progression in people at ultra high clinical risk: a complementary meta-analysis. JAMA Psychiatry 2015;72:622-3.

DOI 10.1002/wps.20250

# Etiological overlap between obsessive-compulsive disorder and anorexia nervosa: a longitudinal cohort, multigenerational family and twin study

Martin Cederlöf[1], Laura M. Thornton[2], Jessica Baker[2], Paul Lichtenstein[1], Henrik Larsson[1], Christian Rück[3], Cynthia M. Bulik[1,2,4], David Mataix-Cols[3]

[1]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; [2]Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; [3]Department of Clinical Neuroscience, Centre for Psychiatric Research and Education, Karolinska Institutet and Stockholm County Council, Stockholm, Sweden; [4]Department of Nutrition, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

*Obsessive-compulsive disorder (OCD) often co-occurs with anorexia nervosa (AN), a comorbid profile that complicates the clinical management of both conditions. This population-based study aimed to examine patterns of comorbidity, longitudinal risks, shared familial risks and shared genetic factors between OCD and AN at the population level. Participants were individuals with a diagnosis of OCD (N=19,814) or AN (N=8,462) in the Swedish National Patient Register between January 1992 and December 2009; their first-, second- and third-degree relatives; and population-matched (1:10 ratio) unaffected comparison individuals and their relatives. Female twins from the population-based Swedish Twin Register (N=8,550) were also included. Females with OCD had a 16-fold increased risk of having a comorbid diagnosis of AN, whereas males with OCD had a 37-fold increased risk. Longitudinal analyses showed that individuals first diagnosed with OCD had an increased risk for a later diagnosis of AN (risk ratio, RR=3.6), whereas individuals first diagnosed with AN had an even greater risk for a later diagnosis of OCD (RR=9.6). These longitudinal risks were about twice as high for males than for females. First- and second-degree relatives of probands with OCD had an increased risk for AN, and the magnitude of this risk tended to increase with the degree of genetic relatedness. Bivariate twin models revealed a moderate but significant degree of genetic overlap between self-reported OCD and AN diagnoses ($r_a$=0.52, 95% CI: 0.26-0.81), but most of the genetic variance was disorder-specific. The moderately high genetic correlation supports the idea that this frequently observed comorbid pattern is at least in part due to shared genetic factors, though disorder-specific factors are more important. These results have implications for current gene-searching efforts and for clinical practice.*

**Key words:** Obsessive-compulsive disorder, anorexia nervosa, eating disorders, genetic epidemiology, comorbidity, shared genetic factors

***(World Psychiatry 2015;14:333–338)***

An association between obsessive-compulsive disorder (OCD) and eating disorders, particularly anorexia nervosa (AN), has long been observed. In clinical academic settings, these conditions co-occur far more frequently than would be expected by chance, with lifetime prevalence estimates of OCD ranging from 9.5 to 62% in patients with eating disorders (1,2). Similarly, the estimated lifetime prevalence of eating disorders in OCD samples ranges between 11 and 42% (3-5). OCD may, in fact, precede the onset of eating disorders in as many as a quarter of cases, though the relevant studies, with few exceptions (5), have been retrospective (1,2,6,7).

Clinically, the comorbidity between OCD and eating disorders poses special challenges. For example, due to the cognitive effects of starvation, very low-weight patients with AN have difficulty engaging in, and benefiting from, cognitive behavioral therapy for OCD.

Given these clinical considerations, there is a need to better understand the nature of the association between OCD and AN. One possibility is that OCD shares familial risk factors with AN. Controlled family studies have indicated elevated rates of OCD in relatives of patients with eating disorders, particularly restricting-type AN (8,9). On the other hand, family studies have not observed elevated rates of eating disorders in relatives of OCD patients (10,11), although this could be related to the low population prevalence of eating disorders. Thus, it is currently unclear whether there is shared familial transmission between OCD and eating disorders (12).

In this study, we linked longitudinal national Swedish registers, including multigenerational families and twins, to shed new light on the nature of the relationships between OCD and AN. We first examined the comorbidity patterns between OCD and AN at the population level. We next employed longitudinal analyses to examine the sequential risk of AN in individuals first diagnosed with OCD, and the sequential risk of OCD in patients first diagnosed with AN. Next, we investigated the risk of AN in relatives of individuals with OCD who did not have a lifetime diagnosis of AN, compared with the risk in relatives of individuals without a diagnosis of OCD or AN, stratified by degree of genetic relatedness to the probands. Finally, we conducted a bivariate twin analysis of self-reported OCD and AN diagnoses in a large population-based female twin sample. Our multi-method approach controls for many of the disease-related confounding factors that can create spurious associations between disorders.

## METHODS

### National registers

Following approval from the Regional Ethics Board in Stockholm, we linked three Swedish national registers,

using the individual personal identification numbers assigned at birth or, for resident immigrants, upon arrival to the country. The Total Population Register contains demographic data on all individuals registered as Swedish inhabitants since 1968, and is extended by the Multi-Generation Register, which contains information about the identity of biological parents of all individuals born in Sweden since 1932 and individuals living in Sweden since 1961. The Swedish National Patient Register (13) covers psychiatric inpatient care since 1969 and psychiatric outpatient care since 2001.

### Definition and validity of ICD codes for OCD and AN

OCD probands were defined as individuals identified in the National Patient Register with at least one ICD-10 diagnosis of OCD (F42). The ICD-10 codes for OCD were validated by obtaining a random sample of patient records (N=68) from three Swedish counties. Each file was reviewed and blindly rated by two independent psychiatrists. The ICD-10 codes had excellent validity, with a positive predictive value of 91% (rater 1) and 96% (rater 2). The inter-rater agreement was outstanding (kappa=0.98, $p<0.001$) (14).

AN probands were defined as individuals identified in the National Patient Register with at least one ICD-10 diagnosis of AN (F50.0 or F50.1). The ICD-10 codes for AN were validated by comparing the eating disorder diagnoses in the National Patient Register to the diagnoses in two specialized quality registers: the Riksät-National Quality Register for Specialized Treatment for Eating Disorders and the Stepwise-regional Quality Assurance System for Eating Disorders (15). This yielded a positive predictive value of 83% and a negative predictive value of 73%.

### Twin data

Twins were recruited from the population-based STAGE (Screening Twin Adults: Genes and Environment) study, based on all twins from the Swedish Twin Registry born from 1959 to 1985 (16). The STAGE target population included approximately 43,000 eligible twins. In 2005-2006, twins were invited by mail to participate in the study; nearly 25,000 individuals responded to the questionnaire, which covered common complex diseases. Twins could also opt to complete a phone interview with a trained interviewer using a computer-based data collection method.

Self-reported OCD was assessed using a single item: "Do you have/have you ever had OCD?" Response options were "yes", "no", and "don't know/refuse". AN was assessed using an expanded, on-line Structured Clinical Interview for DSM-IV (SCID)-based instrument. Study criteria for AN were lifetime lowest illness-related body mass index < 18.55, at least slightly afraid of gaining weight or becoming fat while at low weight, and feeling at least slightly fat while at low weight. Participants were coded "1" if all criteria were present, "0" if fewer than all criteria were present, and "missing" if a diagnosis could not be made. Because there were too few men (N=10) with a diagnosis of AN, only women from monozygotic and same sex dizygotic pairs were included in the twin analyses (17).

STAGE was approved by the Regional Ethics Board, and participants provided informed consent by responding to the questionnaire or verbally over the telephone before participation. This study was also approved by the Biomedical Institutional Review Board at the University of North Carolina.

### Statistical analyses

In the population analyses, we first examined the risk for AN in individuals with OCD, compared with individuals without OCD at the time of the first diagnosis of the probands. For each individual with OCD, 10 comparison individuals matched by birth year, sex, and county of residence were randomly selected from the general population. Comparison individuals had to be alive, living in Sweden, and not diagnosed with OCD at the date of the first OCD diagnosis of the proband.

In longitudinal analyses, we estimated the risk that individuals with OCD would receive a later diagnosis of AN during the follow-up period, compared with individuals without an OCD diagnosis. Conversely, we examined the risk that individuals first diagnosed with AN would later receive a diagnosis of OCD during the follow-up period, compared with individuals without an AN diagnosis. We also calculated the median number of years (plus interquartile range) between the first diagnosis (e.g., OCD) and the subsequent diagnosis (e.g., AN).

We used the multigenerational family design to examine the possible etiological overlap between OCD and AN. Specifically, the risk of AN in relatives of individuals with OCD who did not have a lifetime diagnosis of AN was compared with the risk in relatives of individuals without a diagnosis of OCD or AN. For each proband-relative pair, 10 randomly selected unexposed-relative pairs were matched by birth year and sex, and these individuals had to be alive, living in Sweden, and without a diagnosis of OCD at the time of the first diagnosis of the proband. This method reduces the potential bias introduced by individuals in the population registers entering the study at different times (left truncation). OCD-affected relatives of individuals with OCD were excluded, in order to be sure that we studied independent transmission of the conditions. Shared familial (genetic and environmental) risk factors are assumed when individuals with the index disorder (i.e., OCD) have relatives with the other disorder (i.e., AN) but not the index disorder (18). First-, second- and third-degree relatives were analyzed separately to examine the extent to which the familial associations were influenced by genetic and shared environmental factors.

To estimate the concurrent and sequential risks of AN in individuals with OCD (and vice versa), we calculated risk

**Table 1** Risk of anorexia nervosa in individuals with OCD, compared with matched comparison individuals without OCD from the general population

| | OCD probands (N=19,512)[a] | Matched comparison individuals (N=195,120) | RR (95% CI) |
|---|---|---|---|
| Females + males | 572 (2.9%) | 368 (0.2%) | 16.9 (14.8-19.4) |
| Females | 524 (4.8%) | 355 (0.3%) | 16.1 (14.0-18.5) |
| Males | 48 (0.6%) | 13 (0.02%) | 36.9 (20.0-68.1) |

OCD – obsessive-compulsive disorder, RR – risk ratio, CI – confidence interval
[a]302 patients with OCD could not be assigned comparison individuals per the matching criteria

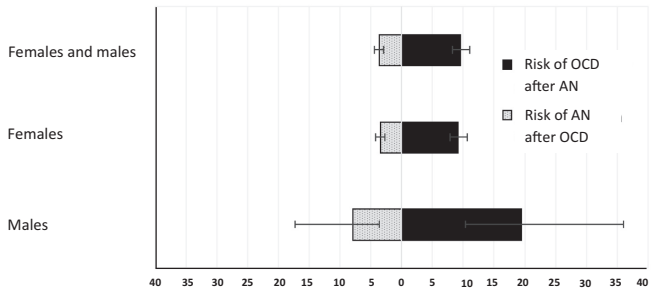


**Figure 1** Sequential risks of receiving a diagnosis of obsessive-compulsive disorder (OCD) after having received an initial diagnosis of anorexia nervosa (AN) (right side, in black), and vice versa (left side, in grey), by proband gender. Values represent risk ratios and confidence intervals

ratios (RR) and 95% confidence intervals (CI) using conditional logistic regression. When assessing risks within families, CI were obtained with a robust sandwich estimator function to adjust for the correlated data structure. All analyses were conducted in SAS version 9.4 (SAS Institute, Cary, NC, USA).

Twin analyses were done using the Mx program (http://www.vcu.edu/mx/). The classic twin study evaluates the proportion of phenotypic variation attributable to genetic variation among individuals (heritability) and what proportions are due to common and unique environmental factors. Specifically, it estimates the proportion of variance due to: a) additive genetic effects (representing the cumulative impact of several genes, i.e., heritability, $a^2$); b) common environmental effects (environmental influences to which both members of a twin pair are exposed regardless of zygosity, $c^2$); and c) unique environmental effects (environmental effects impacting one twin but not the other) and measurement error ($e^2$). Thus, the sum of $a^2 + c^2 + e^2 = 1$ (total variance).

An extension of this twin model, a bivariate structural equation model using Cholesky's decomposition, was fitted to the data. We applied a reduced model including estimates for two sources of variation (additive genetic effects and unique environmental effects, AE model) for OCD and for AN, and correlations indicating the proportion of variance that the two traits share due to genetic ($r_a$) and unique environmental ($r_e$) factors. Model selection was based on the best fitting models for both OCD and AN published elsewhere (19,20) (there was no loss of fit of the AE model compared with the full model). We applied the raw ordinal data option in Mx, which allows data from both complete and incomplete twin pairs to be analyzed. We report parameter estimates with their 95% CI.

## RESULTS

### Comorbidity

We identified 19,814 individuals ever diagnosed with OCD (43.5% males) and 8,462 individuals ever diagnosed with AN (6.4% males). Individuals with OCD had a 17 times higher risk of having a comorbid diagnosis of AN. Although males with OCD had a lower absolute risk of AN (0.6%) compared to females with OCD (4.8%), the relative risk was significantly higher for male than female OCD probands (Table 1).

### Longitudinal analyses

Individuals first diagnosed with OCD had a 4-fold higher risk of receiving a later diagnosis of AN during the follow-up period, compared with individuals without OCD (Figure 1). The median time between the first diagnosis of OCD and the subsequent first diagnosis of AN was 2.2 years (interquartile range, IR=2.8). These risks were approximately double for male than female OCD patients (Table 2).

Conversely, individuals first diagnosed with AN had a 10-time higher risk of receiving a later diagnosis of OCD during the follow-up period, compared with individuals without AN (Figure 1). The median time between diagnoses was 2.4 years (IR=3.0). These risks were significant for both female and male AN patients, though the magnitude of the risk was more than doubled in males (Table 3).

**Table 2** Longitudinal risk of receiving a later diagnosis of anorexia nervosa during the follow-up period in probands with an initial diagnosis of OCD compared with individuals without OCD

| | Initial diagnosis of OCD | | |
|---|---|---|---|
| | Present (N=19,069)[a] | Absent (N=190,690) | RR (95% CI) |
| Females + males | 129 (0.7%) | 366 (0.2%) | 3.6 (2.9-4.4) |
| Females | 118 (1.1%) | 352 (0.3%) | 3.4 (2.7-4.2) |
| Males | 11 (0.1%) | 94 (0.02%) | 7.9 (3.6-17.3) |

OCD – obsessive-compulsive disorder, RR – risk ratio, CI – confidence interval
[a]OCD patients with patients with a prior diagnosis of anorexia nervosa where excluded from the analyses

**Table 3** Longitudinal risk of receiving a later diagnosis of OCD during the follow-up period in probands with an initial diagnosis of anorexia nervosa compared with individuals without anorexia nervosa

| | Initial diagnosis of anorexia nervosa | | |
|---|---|---|---|
| | Present (N=8,192)[a] | Absent (N=81,920) | RR (95% CI) |
| Females + males | 369 (4.5%) | 403 (0.5%) | 9.6 (8.3-11.1) |
| Females | 339 (4.4%) | 386 (0.5%) | 9.2 (7.9-10.7) |
| Males | 30 (6.0%) | 17 (0.3%) | 19.4 (10.4-36.1) |

OCD – obsessive-compulsive disorder, RR – risk ratio, CI – confidence interval
[a]Anorexia nervosa patients with a prior diagnosis of OCD were excluded from the analyses

## Family analyses

When the proband had OCD (but not AN), his/her OCD-unaffected first-, second- and third-degree relatives had an increased risk for AN. This was statistically significant for first-degree relatives (both female and male) and second-degree relatives (female only), and at a trend level for third-degree relatives (Table 4). The magnitude of this risk tended to increase as genetic proximity increased, though the confidence intervals overlapped.

## Twin analyses

The final sample for twin modeling included 8,550 female twins: 1,724 monozygotic pairs with complete data, 177 monozygotic pairs with incomplete data, 1,170 dizygotic pairs with complete data, and 117 dizygotic pairs with incomplete data. In addition, there were 1,035 monozygotic and 1,139 dizygotic individuals without co-twin information. The mean age of these participants was 33.1 years (SD=7.6).

The modeling results for the twin analyses were: $a^2$=66% (95% CI: 54%-76%) and $e^2$=34% (95% CI: 24%-46%) for OCD; and $a^2$=38% (95% CI: 20%-54%) and $e^2$=62% (95% CI: 46%-80%) for AN. The correlation between additive genetic factors for OCD and AN was $r_a$=.52 (95% CI: .26 to .81), while for unique environmental factors it was $r_e$=.11 (95% CI: −.18 to .39). Figure 2 shows the percent of variance attributed to genetic and unique environmental influences that are specific to each disorder and that are shared by the disorders.

**Table 4** Risk of anorexia nervosa in unaffected relatives of individuals with OCD (exposed), compared with relatives of individuals without OCD (unexposed)

| | Risk of anorexia nervosa | | |
|---|---|---|---|
| | Exposed | Unexposed | RR (95% CI) |
| **First-degree relatives** | | | |
| Females + males | 108 (0.2%) | 548 (0.1%) | **1.9 (1.6-2.4)** |
| Females | 102 (0.3%) | 526 (0.2%) | **1.9 (1.6-2.4)** |
| Males | 6 (0.02%) | 22 (0.01%) | **2.6 (1.1-6.2)** |
| **Second-degree relatives** | | | |
| Females + males | 68 (0.1%) | 536 (0.1%) | 1.3 (1.0-1.6) |
| Females | 65 (0.2%) | 503 (0.1%) | **1.3 (1.1-1.6)** |
| Males | 3 (0.01%) | 33 (0.01%) | 0.9 (0.3-2.8) |
| **Third-degree relatives** | | | |
| Females + males | 151 (0.2%) | 1,381 (0.2%) | 1.1 (0.9-1.3) |
| Females | 142 (0.4%) | 1,311 (0.3%) | 1.1 (0.9-1.3) |
| Males | 60 (0.2%) | 492 (0.1%) | 1.3 (0.7-2.5) |

OCD – obsessive-compulsive disorder, RR – risk ratio, CI – confidence interval
Significant RRs are highlighted in bold

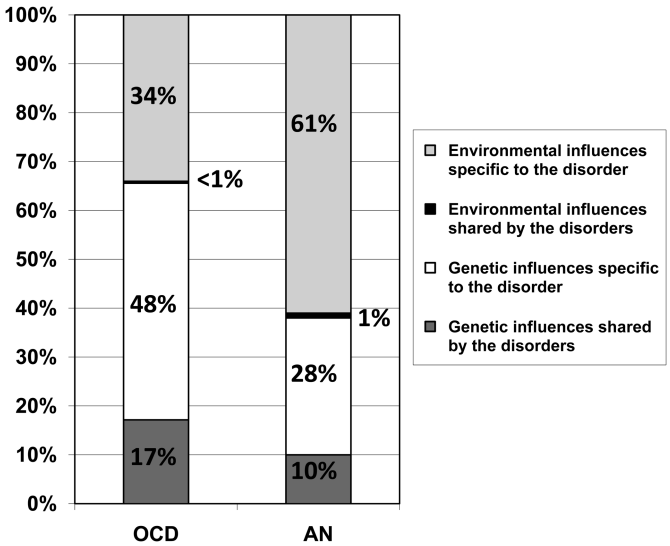


**Figure 2** Proportion of the variance accounted for by common vs. disorder-specific genetic and environmental factors across obsessive-compulsive disorder (OCD) and anorexia nervosa (AN)

## DISCUSSION

Our results extend previous reports by documenting that AN is far more common (17 times) in individuals with OCD than would be expected by chance. This was particularly true for male OCD patients, for whom the risk was increased by 37 times. In longitudinal analyses, we found that an initial diagnosis of OCD increased the risk of a later diagnosis of AN, and vice versa. Again, these longitudinal risks were substantially greater in males. The family analyses showed familial links between OCD and AN, and the bivariate twin analyses further confirmed a moderate degree of genetic overlap between these disorders. However, most of the genetic variance was disorder-specific.

Compared to unaffected individuals, patients first diagnosed with OCD were approximately 4 times more likely to

later develop AN, confirming largely retrospective reports from clinical samples (1,2,5-7) and suggesting that OCD is a risk factor for the development of AN. However, the possibility that subtle eating disorder symptoms were overlooked at initial assessment cannot be ruled out. A previous longitudinal study of pediatric OCD patients showed that those who developed an eating disorder were more likely to be female and to have a family history of an eating disorder (5). In that study, a total of 30% of those who developed an eating disorder at follow-up had eating disorder symptoms or food-related obsessions/compulsions at baseline. This suggests that the nature of OCD symptoms at presentation may assist in identifying individuals at highest risk for developing an eating disorder, and encourages eating disorders symptom screening in individuals seeking help for OCD.

Interestingly, the risk of receiving a diagnosis of OCD after an initial diagnosis of AN was much greater (approximately 10 times) than the risk of receiving a diagnosis of AN after an initial diagnosis of OCD (approximately 4 times). It may be that the diagnosis of AN, which often requires hospitalization, increases the surveillance and thus the detection of OCD. It is also possible that the progressive changes in cognitive function and neurobiology brought on by prolonged periods of starvation and low weight (21) may increase the risk of developing OCD. The relatively long gap between the two diagnoses (a median of over 2 years) is compatible with this interpretation. Although the possibility of misdiagnosis cannot be fully ruled out, our findings suggest that AN may be a more important risk factor for the development of OCD than previously recognized. Given the substantial challenges facing clinicians who manage OCD patients who are very underweight, early detection and management of incipient OCD symptoms in this population are warranted.

Our family analyses provided a rigorous, albeit not probative, test of the possible etiological link between OCD and AN. Indeed, AN was significantly more common in unaffected relatives of probands with OCD, compared with relatives of matched controls. Furthermore, the risks tended to be higher for first-degree relatives, compared to second- and third-degree relatives. Taken together, these findings suggest that shared genetic risk factors underlie the overlap between OCD and AN. This interpretation was further supported in separate bivariate twin analyses, revealing a moderate genetic correlation between self-reported OCD and AN ($r_a$=.52) and minimal overlap in unique environmental influences ($r_e$=.11). Future cross-disorder analyses of genome-wide association data should provide confirmation of these analytic results (22).

Although our results are consistent with a genetic overlap between OCD and AN, disorder-specific genetic and environmental risk factors also seem to contribute to the etiology of each disorder. Our twin results suggest that the majority of genetic variance is disorder-specific and that non-shared environmental influences are largely unique to each condition. These findings may explain the obvious clinical differences between the two conditions (12). Identification of disorder-specific environmental risk factors and genome-wide investigations at cross-cutting dimensional levels will be important next steps.

The increased comorbidity and longitudinal risk in males is intriguing and, to our knowledge, previously unreported. Several interpretations are possible. First, males with AN in general may be less likely to seek treatment (23,24); however, those presenting with complex comorbidities, such as OCD, may be more likely to do so. This would result in an over-representation of males with both AN and OCD in the patient register. Another, not incompatible, explanation is that males could require a greater familial etiologic load to manifest the AN phenotype. Because, as we show in this study, AN and OCD share genetic factors, this would result in a greater comorbidity and sequential risk in males. Similar arguments have been employed to explain the striking male preponderance in autism spectrum disorder (25). Unfortunately, our study was underpowered to conduct twin analyses in males to shed further light on potential differences in heritability of AN between the genders.

Some limitations to our study should be considered when evaluating the results. First, both OCD and AN are underrepresented in the Swedish National Patient Register, particularly OCD. This is largely due to the fact that OCD rarely requires hospitalization (outpatients were only included in the register from 2001) and that many sufferers do not seek treatment. Therefore, OCD patients with severe comorbidities (e.g., AN) may be more likely to be represented in the register, thus inflating the true comorbidity rates and longitudinal associations. Our family based analyses are less likely to be affected by this limitation, as the relatives of patients with OCD did not have a lifetime diagnosis of that disorder. It is still possible, however, that some relatives may have had OCD but did not seek treatment, or had sub-threshold symptoms. The twin analyses, conducted in a general population of twins, were largely unaffected by this limitation. Second, the diagnoses in STAGE were based on self-report and the OCD diagnosis was based on a single item. Finally, due to the low prevalence of AN in men, we were unable to examine possible gender differences in our bivariate twin models.

To conclude, the high comorbidity, sequential risk, and shared familial risks between OCD and AN suggest partially shared genetic etiological mechanisms between these disabling mental disorders, although the majority of the genetic variance was unique to each disorder. Clinicians should be aware that having one disorder might increase the risk of developing the other, even several years after the initial diagnosis. Our results underscore the importance of screening for the other disorder or nascent symptoms at clinical presentation and throughout treatment. Research into the optimal management of these complex comorbidities is warranted.

## Acknowledgements

This study was supported by the Swedish Council for Working Life and Social Research, the Swedish Research

Council, and the Anorexia Nervosa Genetics Initiative of the Klarman Family Foundation.

## References


1. Godart NT, Flament MF, Perdereau F et al. Comorbidity between eating disorders and anxiety disorders: a review. Int J Eat Disord 2002;32:253-70.
2. Kaye WH, Bulik CM, Thornton L et al. Comorbidity of anxiety disorders with anorexia and bulimia nervosa. Am J Psychiatry 2004;161:2215-21.
3. Rubenstein CS, Pigott TA, L'Heureux F et al. A preliminary investigation of the lifetime prevalence of anorexia and bulimia nervosa in patients with obsessive compulsive disorder. J Clin Psychiatry 1992;53:309-14.
4. Rasmussen SA, Eisen JL. The epidemiology and differential diagnosis of obsessive compulsive disorder. J Clin Psychiatry 1994; 55(Suppl. 10):5-10.
5. Micali N, Hilton K, Natatani E et al. Is childhood OCD a risk factor for eating disorders later in life? A longitudinal study. Psychol Med 2011;41:2507-13.
6. Bulik CM, Sullivan PF, Fear JL et al. Eating disorders and antecedent anxiety disorders: a controlled study. Acta Psychiatr Scand 1997;96:101-7.
7. Deep AL, Nagy LM, Weltzin TE et al. Premorbid onset of psychopathology in long-term recovered anorexia nervosa. Int J Eat Disord 1995;17:291-7.
8. Lilenfeld LR, Kaye WH, Greeno CG et al. A controlled family study of anorexia nervosa and bulimia nervosa: psychiatric disorders in first-degree relatives and effects of proband comorbidity. Arch Gen Psychiatry 1998;55:603-10.
9. Bellodi L, Cavallini MC, Bertelli S et al. Morbidity risk for obsessive-compulsive spectrum disorders in first-degree relatives of patients with eating disorders. Am J Psychiatry 2001;158:563-9.
10. Bienvenu OJ, Samuels JF, Riddle MA et al. The relationship of obsessive-compulsive disorder to possible spectrum disorders: results from a family study. Biol Psychiatry 2000;48:287-93.
11. Bienvenu OJ, Samuels JF, Wuyek LA et al. Is obsessive-compulsive disorder an anxiety disorder, and what, if any, are spectrum conditions? A family study perspective. Psychol Med 2012;42:1-13.
12. Phillips KA, Kaye WH. Relationship of body dysmorphic disorder and eating disorders to obsessive-compulsive disorder. In: Hollander E, Zohar J, Sirovatka PJ et al (eds). Obsessive-compulsive spectrum disorders. Refining the research agenda for DSM-V. Arlington: American Psychiatric Association, 2011:33-56.
13. Ludvigsson JF, Andersson E, Ekbom A et al. External review and validation of the Swedish national inpatient register. BMC Public Health 2011;11:1-16.
14. Rück C, Larsson KJ, Lind K et al. Validity and reliability of chronic tic disorder and obsessive-compulsive disorder diagnoses in the Swedish National Patients Register. BMJ Open 2015;5:e007520.
15. Birgegard A, Bjorck C, Clinton D. Quality assurance of specialised treatment of eating disorders using large-scale Internet-based collection systems: methods, results and lessons learned from designing the Stepwise database. Eur Eat Disord Rev 2010;18: 251-9.
16. Lichtenstein P, Sullivan PF, Cnattingius S et al. The Swedish Twin Registry in the third millennium: an update. Twin Res Hum Genet 2006;9:875-82.
17. Pisetsky EM, Thornton LM, Lichtenstein P et al. Suicide attempts in women with eating disorders. J Abnorm Psychol 2013;122: 1042-56.
18. Szatmari P, White J, Merikangas KR. The use of genetic epidemiology to guide classification in child and adult psychopathology. Int Rev Psychiatry 2007;19:483-96.
19. Bulik CM, Thornton LM, Root TL et al. Understanding the relation between anorexia nervosa and bulimia nervosa in a Swedish national twin sample. Biol Psychiatry 2010;67:71-7.
20. Mataix-Cols D, Boman M, Monzani B et al. Population-based, multigenerational family clustering study of obsessive-compulsive disorder. JAMA Psychiatry 2013;70:709-17.
21. Katzman DK, Christensen B, Young AR et al. Starving the brain: structural abnormalities and cognitive impairment in adolescents with anorexia nervosa. Semin Clin Neuropsychiatry 2001;6:146-52.
22. Bulik-Sullivan BK, Loh PR, Finucane HK et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 2015;47:291-5.
23. O'Dea JA, Abraham S. Eating and exercise disorders in young college men. J Am Coll Health 2002;50:273-8.
24. Olivardia R, Pope HG Jr., Mangweth B et al. Eating disorders in college men. Am J Psychiatry 1995;152:1279-85.
25. Robinson EB, Lichtenstein P, Anckarsater H et al. Examining and interpreting the female protective effect against autistic behavior. Proc Natl Acad Sci USA 2013;110:5258-62.


DOI 10.1002/wps.20251

# Risk of metabolic syndrome and its components in people with schizophrenia and related psychotic disorders, bipolar disorder and major depressive disorder: a systematic review and meta-analysis

Davy Vancampfort[1,2], Brendon Stubbs[3], Alex J. Mitchell[4,5], Marc De Hert[1], Martien Wampers[1], Philip B. Ward[6], Simon Rosenbaum[6], Christoph U. Correll[7,8]

[1]UPC KU Leuven, Campus Kortenberg, Department of Neurosciences, KU Leuven - University of Leuven, Kortenberg, Belgium; [2]Department of Rehabilitation Sciences, KU Leuven - University of Leuven, Leuven, Belgium; [3]School of Health and Social Care, University of Greenwich, Eltham, London, UK; [4]Department of Psycho-oncology, Leicestershire Partnership NHS Trust, Leicester, UK; [5]Department of Cancer and Molecular Medicine, University of Leicester, Leicester, UK; [6]School of Psychiatry, University of New South Wales, Sydney, NSW, Australia; [7]Zucker Hillside Hospital, Glen Oaks, NY, USA; [8]Hofstra North Shore LIJ School of Medicine, Hempstead, NY, USA

*Metabolic syndrome (MetS) and its components are highly predictive of cardiovascular diseases. The primary aim of this systematic review and meta-analysis was to assess the prevalence of MetS and its components in people with schizophrenia and related psychotic disorders, bipolar disorder and major depressive disorder, comparing subjects with different disorders and taking into account demographic variables and psychotropic medication use. The secondary aim was to compare the MetS prevalence in persons with any of the selected disorders versus matched general population controls. The pooled MetS prevalence in people with severe mental illness was 32.6% (95% CI: 30.8%-34.4%; N = 198; n = 52,678). Relative risk meta-analyses established that there was no significant difference in MetS prevalence in studies directly comparing schizophrenia versus bipolar disorder, and in those directly comparing bipolar disorder versus major depressive disorder. Only two studies directly compared people with schizophrenia and major depressive disorder, precluding meta-analytic calculations. Older age and a higher body mass index were significant moderators in the final demographic regression model ($z = -3.6$, $p = 0.0003$, $r^2 = 0.19$). People treated with all individual antipsychotic medications had a significantly ($p<0.001$) higher MetS risk compared to antipsychotic-naïve participants. MetS risk was significantly higher with clozapine and olanzapine (except vs. clozapine) than other antipsychotics, and significantly lower with aripiprazole than other antipsychotics (except vs. amisulpride). Compared with matched general population controls, people with severe mental illness had a significantly increased risk for MetS (RR = 1.58; 95% CI: 1.35-1.86; $p<0.001$) and all its components, except for hypertension ($p = 0.07$). These data suggest that the risk for MetS is similarly elevated in the diagnostic subgroups of severe mental illness. Routine screening and multidisciplinary management of medical and behavioral conditions is needed in these patients. Risks of individual antipsychotics should be considered when making treatment choices.*

**Key words:** Metabolic syndrome, severe mental illness, schizophrenia, bipolar disorder, major depressive disorder, antipsychotics

***(World Psychiatry 2015;14:339–347)***

People with severe mental illness (SMI), including schizophrenia and related psychotic disorders, bipolar disorder and major depressive disorder (MDD), experience a two-three times higher mortality rate than the general population (1,2). This mortality gap translates into a 10-20 year shortened life expectancy (3,4) and appears to be widening (5). About 60% of the excess mortality observed in SMI is due to physical comorbidities, predominantly cardiovascular diseases (CVD) (6). Factors predisposing people with SMI to CVD include antipsychotic medication and unhealthy lifestyles (7) as well as their reduced likelihood to receive standard levels of medical care (8-12).

To assist clinicians in identifying and treating patients at an increased risk of CVD, the concept of metabolic syndrome (MetS) has been introduced. MetS is defined by a combination of central obesity, high blood pressure, low high-density lipoprotein (HDL) cholesterol, elevated triglycerides and hyperglycaemia. In the general population, these clustered risk factors have been associated with the development of CVD and excess mortality (13-15). Current definitions (16-19) for MetS are aimed at being easy to use in clinical settings and share similar diagnostic thresholds (20). However, the role of abdominal obesity is central to the MetS definition of the International Diabetes Federation (IDF) (18), with provision of ethnic specific thresholds for waist circumference, while central obesity is not a mandatory criterion in the MetS definition of the National Cholesterol Education Program (NCEP) Adult Treatment Panel III (ATP III) (16,17). As a prevalent condition and predictor of CVD across racial, gender and age groups, MetS provides the opportunity to identify high-risk populations and prevent the progression of some major causes of morbidity and mortality (20).

Previous meta-analyses (21-24) documented that people with SMI have an increased risk for developing MetS compared with the general population. A brief meta-analytic report comparing MetS frequencies in patients with schizophrenia and bipolar disorder found that these populations are at similar risk (25). However, these findings should be interpreted with caution, since comparisons were performed at study level and not limited to studies directly comparing the two populations, and patient samples were not matched

for age and illness duration (26). Meta-analytic comparisons of schizophrenia and related psychotic disorders or bipolar disorder with major depressive disorder are currently lacking. In the same way, meta-analytic data including all major diagnostic SMI subgroups (i.e., schizophrenia and related psychotic disorders versus bipolar disorder versus major depressive disorder) are absent in the literature.

Large-scale pooled analyses in the SMI population are highly relevant, as they enable investigation of risk factors across large numbers of studies and participants, dissecting risk factors for MetS associated with SMI from those independent of it. Pooling data across major diagnostic categories allows for investigation of the effect of demographic variables (age, illness duration, gender, setting, geographical region) and treatments (particularly mood stabilizers and antipsychotics, as well as polypharmacotherapy versus monotherapy). If risk stratification is observed, this could potentially help guide clinicians in monitoring and treatment.

We conducted a systematic review and meta-analysis to assess pooled prevalences of MetS and its components in people with schizophrenia and related psychotic disorders, bipolar disorder and major depressive disorder, selecting studies directly comparing subjects with different disorders and taking into account demographic variables and medication use. Our secondary aim was to compare the MetS prevalence in persons with any of the selected disorders versus matched general population controls.

## METHODS

### Inclusion and exclusion criteria

This systematic review was conducted in accordance with the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) guidelines (27) and in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) standard (28). We included observational studies (cross-sectional, retrospective and prospective studies) in adults that fulfilled the following criteria: a) a diagnosis of schizophrenia or a related psychotic disorder, bipolar disorder or major depressive disorder according to the DSM-IV or ICD-10, irrespective of clinical setting (inpatient, outpatient or mixed); and b) a MetS diagnosis according to non-modified ATP-III (16), ATP-III-A (17), IDF (18) or World Health Organization (19) standards. For a randomized control trial, we extracted the variables of interest at baseline. There were no language or time restrictions.

For estimation of the prevalence of MetS, we excluded studies with: a) non-standardized diagnoses, b) non-standardized definitions of MetS, c) insufficient data for extraction of MetS frequencies, d) restriction to patients at risk for or without cardiovascular diseases, and e) restriction to children and/or adolescents. In the case of multiple publications from the same study, only the most recent paper or the article with the largest sample was included. When required, we contacted the primary/corresponding authors of potential studies to confirm eligibility, or to acquire the variables of interest if they were not available in the publication.

### Search criteria, study selection and critical appraisal

Two independent authors (DV, BS) searched MEDLINE, PsycARTICLES, EMBASE and CINAHL from database inception to January 1, 2015. Key words used were "metabolic syndrome" AND "severe mental illness" OR "schizophrenia" OR "bipolar disorder" OR "depression" OR "depressive disorder" in the title, abstract or index term fields. Manual searches were also conducted using the reference lists from recovered articles and recent meta-analyses (21-24).

After the removal of duplicates, the reviewers screened the titles and abstracts of all potentially eligible articles. Both authors applied the eligibility criteria, and a list of full text articles was developed through consensus. The reviewers then considered the full texts of these articles and the final list of included articles was reached through consensus. A third reviewer (AJM) was available for mediation throughout this process.

Methodological appraisal was performed according to PRISMA standards (28), including evaluation of bias (confounding, overlapping data, publication bias). Publication bias was tested using the Egger's regression method (29) and Begg-Mazumdar test (30), with a p value $<0.05$ suggesting the presence of bias.

### Statistical analyses

We pooled individual study data using the DerSimonian-Laird proportion method with StatsDirect (31). The trim-and-fill approach (32) was used to adjust the overall estimate for funnel plot asymmetry. Due to anticipated heterogeneity, a random effects meta-analysis was employed. Heterogeneity was measured with the Q statistic, yielding a chi-square p value, with $p<0.05$ indicating significant heterogeneity of the pooled results. We calculated the relative risk (RR) to investigate the prevalence of MetS and its components within and across SMI subgroups, the latter only in those studies directly comparing diagnostic subgroups. Moreover, we compared the prevalence of MetS between people with schizophrenia, bipolar disorder and major depressive disorder versus general population control groups that were matched on age and sex, also only using data from studies in which they were directly compared. In both analyses, only comparisons of specific SMI groups or an SMI group with a matched general population group were included that had been performed within the same study, in order to minimize variability of MetS frequencies due to different sampling and assessment procedures.

In order to increase homogeneity of compared samples and eliminate smaller studies with less precise point estimates,

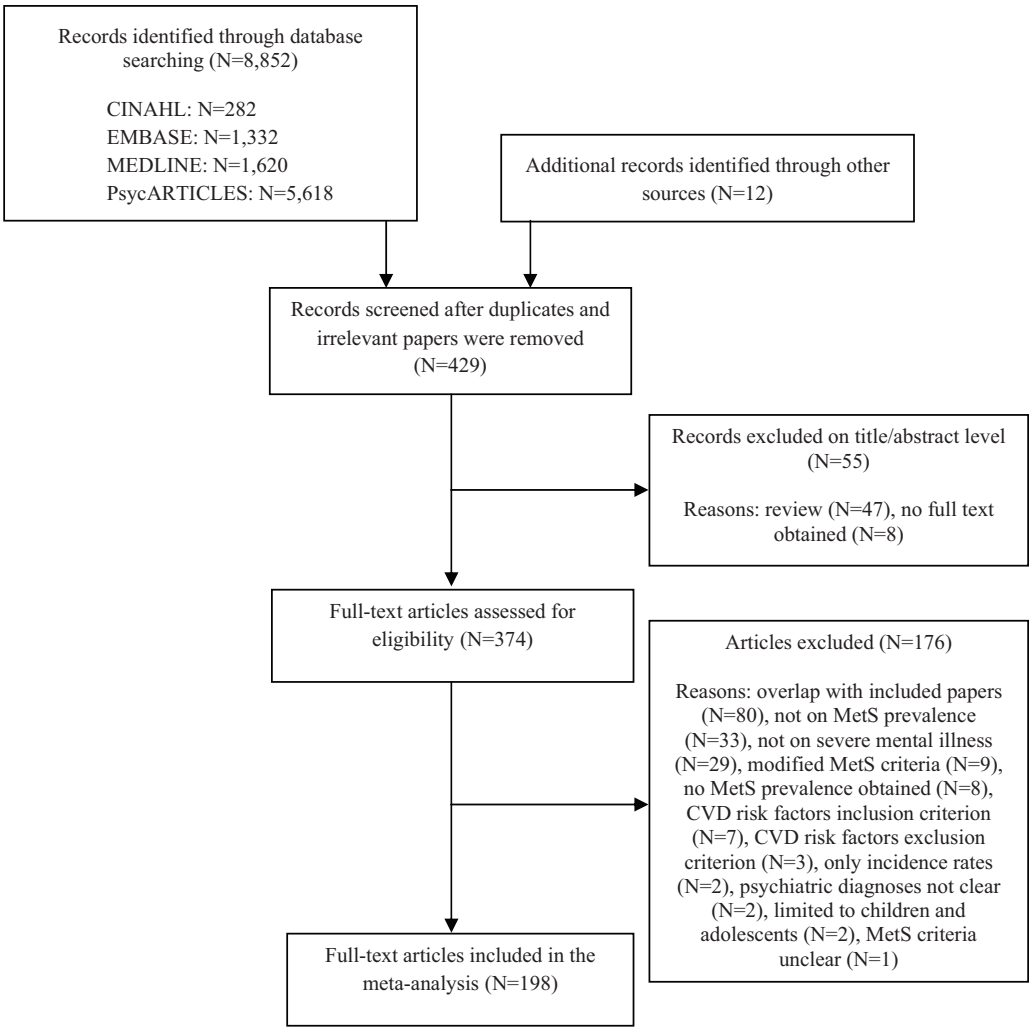


**Figure 1** Flow diagram for the search strategy. MetS – metabolic syndrome, CVD – cardiovascular disease

we also conducted sensitivity analyses, restricting the sample to large, population-based studies. Furthermore, in the entire dataset, we conducted subgroup analyses (including $\chi^2$ tests, t tests, odds ratios) to investigate differences between the three main diagnostic subgroups and between first episode and multi-episode illness, gender differences, and differences across medication regimes (antipsychotics, antidepressants, mood stabilizers, monotherapy versus polypharmacotherapy) and geographical regions. In order to reduce heterogeneity, we did not calculate diagnostic and gender differences across studies, but pooled only data of studies that compared these differences on a patient level.

Further, we conducted meta-regression analyses to investigate potential moderators (age, percentage of males, illness duration, body mass index, smoking rates) with Comprehensive Meta Analysis (version 3). Finally, since patients with a first episode of schizophrenia and those with chronic schizophrenia differ significantly in age, and since older age is a significant moderator of higher MetS rates, we also conducted a multivariable meta-regression analysis, adding both first versus multi-episode schizophrenia and age as variables into the analysis.

## RESULTS

### Search results and included participants

Our search yielded 429 publications, of which 198 met inclusion criteria (Figure 1). The list of included and excluded studies (with reasons) is available upon request. The final sample comprised 52,678 unique persons with SMI. Sample sizes ranged from 14 to 3,568 participants, with a mean sample size of 264. Mean age was 41.3 years (range 22.2-73.2), and mean illness duration was 12.4 years. Fifty-seven studies (n = 12,560) reported smoking frequencies, and half of the included participants (50.4%, 95% CI: 46.7%-54.0%, $Q = 1192.0$, $p<0.001$) smoked. The mean body mass index of the sample was 27.3 (SD = 2.7).

## Prevalence of metabolic syndrome and its components

The estimated weighted mean prevalence of MetS was 32.6% (95% CI: 30.8%-34.4%, Q = 3696, p<0.001, n= 52,678). The Begg-Mazumdar (Kendall's tau b = 0.15, p= 0.0015) and Egger test (bias = 1.46, 95% CI: 0.15-2.77, p=0.0292) indicated some publication bias. The trim-and-fill method demonstrated that adjusting for publication bias had little effect on the pooled MetS estimate, which was virtually identical (32.5%, 95% CI: 30.8%-34.2%, Q = 2991, p<0.01, n = 52,678). Restricting the analysis to population-based studies (N = 29, n = 18,594), the overall weighted mean prevalence of MetS was 35.9% (95% CI: 31.8%-40.0%, Q = 934.8, p<0.001).

Sixty-five studies reported on obesity frequency defined as waist circumference >102 cm in males and >88 cm in females (ATP-III or ATP-III-A), while 14 studies reported the obesity frequency following the ethnicity-specific IDF criteria. Overall, the proportion of patients with abdominal obesity was 50.3% by the ATP definitions (n = 20,210; 95% CI: 46.9%-53.7%, Q = 1.6, p<0.001) and 63.2% according to IDF (n = 3,789; 95% CI: 53.6%-72.3%, Q = 480.9, p< 0.001). In studies reporting on hyperglycaemia, the frequency was 18.8% (N = 56, n = 17,508; 95% CI: 16.6%-21.2%, Q = 906.9, p<0.001) when the threshold was ≥110 mg/dl (ATP-III), while it was 23.0% (N = 28, n = 8,205; 95% CI: 17.3%-29.2%, Q = 1.1, p<0.001) when the threshold was ≥100 mg/dl (ATP-III-A and IDF). Hypertriglyceridemia was present in 36.2% (N = 87, n = 26,577; 95% CI: 33.1%-39.3%, Q = 2.7, p<0.001). Low HDL cholesterol was present in 39.1% (N = 86, n = 26,193; 95% CI: 36.4%-41.9%, Q = 1.9, p<0.001). Hypertension (ATP-III, ATP-III-A and IDF) was present in 39.3% (N = 88, n = 27,441; 95% CI: 36.1%-42.5%, Q = 2.7, p<0.001).

## Subgroup analyses and predictors of metabolic syndrome

### *Diagnostic subgroups*

The pooled MetS prevalence was 33.4% (95% CI: 30.8%-36.0%, Q = 1955.0, p<0.001) in people with schizophrenia (N = 93, n = 29,596), and 34.6% (95% CI: 29.3%-40.0%, Q = 110.2, p<0.001) in those with related psychotic disorders (N = 13, n = 2,850). Similar pooled MetS prevalences were observed in patients with bipolar disorder (31.7%, 95% CI: 27.3%-36.3%, Q = 843.5, p<0.001; N = 33, n = 5,827) and major depressive disorder (31.3%, 95% CI: 27.3%-35.5%, Q = 142.7, p<0.001; N = 19, n = 5,415). In population-based studies, the pooled prevalence of MetS was 38.9% (95% CI: 34.6%-43.4%, Q = 458.1, p<0.001; N = 20, n = 12,770) for schizophrenia and 22.7% (95% CI: 20.4%-25.1%, Q = 2.28, p = 0.31; N = 3, n = 1,503) for major depressive disorder. There were insufficient data for bipolar disorder.

The releative risk of MetS versus age- and gender-matched healthy controls was 1.87 in schizophrenia and related psychotic disorders (95% CI: 1.53-2.29; p<0.001, Q = 18.3, p = 0.03; N = 11, n = 1,413), 1.58 in bipolar disorder (95% CI: 1.24-2.03; p<0.001, Q = 6.6, p = 0.25; N = 6, n = 1,125) and 1.57 in major depressive disorder (95% CI: 1.38-1.79, p<0.001, Q = 19.0, p = 0.26; N = 17, n = 5,267).

Relative risk meta-analyses established that there was no significant difference in MetS in studies directly comparing schizophrenia (39.2%, 95% CI: 30.5%-48.3%; n = 2,338) versus bipolar disorder (35.5%, 95% CI: 27.0-44.3%; n = 2,077) (N = 10, RR = 0.92, 95% CI: 0.79%-1.06%; $\chi^2 = 1.33$, p = 0.24; Q = 21.3, p<0.011). Similarly, there were no differences in the study directly comparing bipolar disorder (29.2%, 95% CI: 14.5%-46.2%; n = 137) versus major depressive disorder (34.0%, 95% CI: 19.4%-50.3%; n = 176) (N = 4; RR = 0.87, 95% CI: 0.48- 1.55; $\chi^2 = 0.21$, p = 0.64; Q = 7.73, p = 0.0518). Only two studies directly compared MetS in people with schizophrenia and major depressive disorder, precluding meta-analytic calculations.

Comparing MetS in first versus multi-episode patients within illness subgroups, first episode psychosis patients (13.7%, 95% CI: 10.4%-16.9%, Q = 8.659, p = 0.034; N = 4, n = 424) had a significantly lower MetS risk than those with multi-episode schizophrenia (34.2%, 95% CI: 30.8%-36.0%, Q = 1,955, p<0.001; N = 105, n = 29,596) (z = −8.9, p< 0.001). In order to assess if the difference in MetS rates remained significant when age was entered into the analyses, we conducted a multivariable meta-regression analysis. Within this, we pooled the prevalence of MetS in first and multi-episode schizophrenia and found that, although mean age predicted MetS prevalence (coefficient = 0.0296; 95% CI: 0.013 to 0.0463, z = 3.49, p = 0.005), first episode was also a unique predictor of lower MetS (coefficient = −0.7517; 95% CI: −1.4877 to −0.0157; z = −2, p = 0.04; $r^2 = 0.24$). There were no data in first-episode bipolar disorder or major depressive disorder patients, precluding a comparison with multi-episode patients.

### *Demographic variables*

A relative risk meta-analysis across 64 studies directly comparing MetS frequencies in male (33.5%, 95% CI: 30.0%-36.7%, Q = 814, p<0.001; n = 10,798) versus female (33.4%, 95% CI: 31.5%-38.4%, Q = 615, p<0.001; n = 8,027) participants with SMI found no gender differences (RR = 0.94; 95% CI: 0.85-1.02; $\chi^2 = 2.06$, p = 0.15; Q = 232.0, p<0.011).

Separate meta-regression analyses revealed that higher MetS frequencies were moderated by older age (coefficient = 0.0278; 95% CI: 0.0178-0.0379, z = 5.5, p<0.0001), longer illness duration (coefficient = 0.0339; 95% CI: 0.0115-0.0564, z = 2.96, p = 0.003) and higher body mass index (coefficient = 0.1537; 95% CI: 0.095-0.2123, z = 5.14, p< 0.0001), but not by smoking status (p = 0.49). When all

**Table 1** Geographical differences in pooled metabolic syndrome (MetS) prevalence

| Region | No. studies | Pooled MetS prevalence | Cochran Q |
|---|---|---|---|
| Australia and New Zealand* | 6 | 50.2% (95% CI: 35.3%-65.0%) | 73.8, p<0.001 |
| Middle-East | 6 | 35.3% (95% CI: 31.3%-39.5%) | 1287.6, p<0.001 |
| North-America | 46 | 32.4% (95% CI: 24.7%-40.8%) | 38.0, p<0.001 |
| Europe | 81 | 32.0% (95% CI: 29.4%-34.7%) | 1226.4, p<0.001 |
| Asia | 50 | 31.0% (95% CI: 27.7%-34.4%) | 691.3, p<0.001 |
| South-America | 10 | 25.8% (95% CI: 20.7%-31.3%) | 42.3, p<0.001 |
| **Country** | **No. studies** | **Pooled MetS prevalence** | **Cochran Q** |
| Australia | 5 | 50.2% (95% CI: 32.9%-67.4%) | 72.7, p<0.001 |
| South Korea | 7 | 38.9% (95% CI: 30.8%-47.3%) | 103.3, p<0.001 |
| The Netherlands | 11 | 36.5% (95% CI: 29.0%-44.4%) | 167.3, p<0.001 |
| USA | 38 | 36.4% (95% CI: 32.0%-40.9%) | 1217.8, p<0.001 |
| Croatia | 7 | 33.1% (95% CI: 24.6%-42.3%) | 39.1, p<0.001 |
| Spain | 12 | 31.0% (95% CI: 24.5%-37.9%) | 210.3, p<0.001 |
| Finland | 5 | 30.4% (95% CI: 21.8%-39.8%) | 17.9, p<0.001 |
| Taiwan | 13 | 29.8% (95% CI: 24.7%-35.1%) | 124.1, p<0.001 |
| Germany | 6 | 28.7% (95% CI: 19.2%-39.2%) | 62.8, p<0.001 |
| Canada | 5 | 27.4% (95% CI: 17.3%-38.7%) | 44.2, p<0.001 |
| India | 16 | 26.3% (95% CI: 19.0%-34.3%) | 193.0, p<0.001 |
| Brazil | 8 | 25.4% (95% CI: 18.5%-32.9%) | 39.4, p<0.001 |

*Significantly higher than in other regions, p<0.01

significant predictors were entered in one meta-regression model, body mass index (coefficient = 0.142, 95% CI: 0.0438-0.2405, z = 2.83, p = 0.004) and age (coefficient = 0.0556, 95% CI: 0.0025-0.1087, z = 2.05, p = 0.04) remained significant predictors, whilst illness duration did not (p = 0.19). Overall, the final model was a significant predictor of the variance in MetS (z = −3.6, p = 0.0003; $r^2 = 0.19$).

Pooled MetS prevalences per geographical region and country (if N≥5) can be found in Table 1. The MetS prevalence was significantly higher in Australia and New Zealand compared with all other regions (p<0.001). Pooled MetS prevalences per country ranged from 25.4% (95% CI: 18.5%-32.9%) in Brazil to 50.2% (95% CI: 32.9%-67.4%) in Australia.

### *Medication use*

Data from five studies demonstrated a trend for lower pooled MetS prevalence in participants receiving monotherapy (30.4%, 95% CI 25.4%-35.5%, Q = 15.2, p = 0.004; n = 1,364) versus polytherapy (35.2%, 95% CI: 23.8%-47.5%, Q = 18.8, p = 0.008; n = 313) (RR = 0.81; 95% CI: 0.66-1.01; $\chi^2 = 3.41$, p = 0.065; Q = 5.87, p = 0.21).

Forty-eight papers including 147 analyses reported on antipsychotics (monotherapy and N≥5). The prevalence of MetS was lowest in antipsychotic-naïve participants (10.2%, 95% CI: 6.8%-14.3%). Among those receiving antipsychotics, participants taking aripiprazole had the lowest MetS prevalence (19.4%, 95% CI:8.0%-34.2%; N = 6), whilst those taking clozapine had the highest (47.2%, 95% CI: 42.0%-52.6%; N = 30). Patients treated with amisulpride, typical antipsychotics, risperidone, olanzapine and quetiapine had MetS frequencies of 22.8% (95% CI: 7.6%-43.2%; N = 5), 28.0% (95% CI: 19.8%-37.2%; N = 15), 30.7% (95% CI: 23.7%-38.1%; N = 20), 36.2% (95% CI: 31.8%-40.9%; N = 26) and 37.3% (95% CI: 27.4-47.8%; N = 11), respectively.

An overview of the odds ratios comparing individual medications (if monotherapy and N≥5) with each other (at study level) is presented in Table 2. Patients treated with all individual antipsychotic medications had significantly (p<0.001) higher MetS risk compared to antipsychotic-naïve participants. Those treated with clozapine consistently had significantly (p<0.001) higher MetS prevalence than those treated with any other individual antipsychotic medication. Those treated with olanzapine had significantly higher MetS prevalence than those treated with amisulpride (p<0.05), aripiprazole (p<0.001), risperidone (p<0.01) and typical antipsychotic medications (p<0.05). Those treated with aripiprazole had significantly lower odds of MetS compared to other antipsychotic medications (except vs. amisulpride). There were insufficient data to compare the MetS prevalence between antipsychotic-naïve persons and those treated with specific antidepressants or mood stabilizers in similar populations.

**Table 2** Odds ratios for metabolic syndrome risk for individual antipsychotic medications (if monotherapy and N≥5)

| Medication | Antipsychotic-naïve | Amisulpride | Aripiprazole | Clozapine | Olanzapine | Quetiapine | Risperidone |
|---|---|---|---|---|---|---|---|
| Amisulpride | 3.86*** (↑)<br>(2.54-5.84)<br>N = 15; n = 999 | / | / | / | / | / | / |
| Aripiprazole | 3.25*** (↑)<br>(2.36-4.49)<br>N = 16; n = 1,319 | 0.84 (↔)<br>(0.57-1.25)<br>N = 11; n = 692 | / | / | / | / | / |
| Clozapine | 7.81*** (↑)<br>(6.02-10.22)<br>N = 22; n = 2,398 | 2.02*** (↑)<br>(1.45-2.83)<br>N = 17; n = 1,177 | 2.40*** (↑)<br>(1.91-3.03)<br>N = 18; n = 2,091 | / | / | / | / |
| Olanzapine | 5.87*** (↑)<br>(4.53-7.67)<br>N = 22; n = 2,633 | 1.52* (↑)<br>(1.08-2.16)<br>N = 15; n = 2,006 | 1.81*** (↑)<br>(1.44-2.27)<br>N = 16; n = 2,326 | 0.75*** (↓)<br>(0.65-0.86)<br>N = 22; n = 3,405 | / | / | / |
| Quetiapine | 5.14*** (↑)<br>(3.75-7.07)<br>N = 21; n = 1,266 | 1.33 (↔)<br>(0.90-1.97)<br>N = 16; n = 639 | 1.58*** (↑)<br>(1.19-2.11)<br>N = 17; n = 959 | 0.66*** (↓)<br>(0.53-0.82)<br>N = 23; n = 2,038 | 0.88 (↔)<br>(0.70-1.09)<br>N = 22; n = 2,273 | / | / |
| Risperidone | 4.57*** (↑)<br>(3.48-6.03)<br>N = 30; n = 2,025 | 1.18 (↔)<br>(0.83-1.69)<br>N = 25; n = 1398 | 1.40*** (↑)<br>(1.10-1.79)<br>N = 26; n = 1,718 | 0.58*** (↓)<br>(0.50-0.68)<br>N = 32; n = 2,797 | 0.78** (↓)<br>(0.66-0.91)<br>N = 30; n = 3,032 | 0.89 (↔)<br>(0.70-1.12)<br>N = 31; n = 1,665 | / |
| Typical antipsychotics | 4.97*** (↑)<br>(3.83-6.51)<br>N = 17; n = 2,525 | 1.28 (↔)<br>(0.91-1.83)<br>N = 12; n = 1,898 | 1.53*** (↑)<br>(1.23-1.91)<br>N = 13; n =2,218 | 0.64*** (↓)<br>(0.55-0.73)<br>N = 19; n = 3,297 | 0.85* (↓)<br>(0.74-0.97)<br>N = 17; n = 3,532 | 0.97 (↔)<br>(0.77-1.21)<br>N = 18; n = 2,165 | 1.09 (↔)<br>(0.93-1.28)<br>N = 27; n = 2,924 |

*Two-sided p<0.05, two-sided p<0.01, ***two-sided p<0.001
↑ = higher risk, ↓ = lower risk, ↔ = no significant risk difference

### Risk of metabolic syndrome and its components in persons with various disorders compared with general population controls

Thirty studies also provided data on MetS prevalence in healthy control subjects. In a pooled relative risk meta-analysis, persons with SMIs (n = 6,610; 29.2%, 95% CI: 25.9%-32.6%; Q = 230, p<0.001), compared with general population controls (n = 101,223; 18.1%, 95% CI: 15.8%-20.5%, Q = 230, p<0.001), had significantly increased risk of MetS (RR = 1.58, 95% CI: 1.35-1.86, p<0.001; Q = 62, p = 0.003).

People with severe mental illness had significantly increased risk for abdominal obesity (N = 18; RR = 1.43, 95% CI: 1.23-1.66, p<0.001; Q = 198.8, p<0.001), low HDL cholesterol (N = 19; RR = 1.33, 95% CI: 1.15-1.54, p<0.001; Q = 114.7, p<0.001), hypertriglyceridemia (N = 19; RR = 1.49, 95% CI: 1.28-1.73, p<0.001; Q = 91.2, p< 0.001), and hyperglycaemia (N = 20; RR = 1.51, 95% CI: 1.24-1.84, p<0.001; Q = 94.4, p<0.001), with a statistical trend for hypertension (N = 12; RR = 1.12, 95% CI: 0.99-1.28, p = 0.07; Q = 127.1, p<0.001).

## DISCUSSION

To our knowledge, this is the first meta-analysis of MetS and its components including and comparing data from the main SMIs: schizophrenia and related psychotic disorders, bipolar disorder and major depressive disorder. Approximately one third, 32.6% (95% CI: 30.8%-34.4%), of this population had MetS and the relative risk was 1.58 times higher than in the respective general population. MetS prevalences were consistently elevated for each of the three diagnostic subgroups compared to the general population, and comparative meta-analyses found no significant differences across schizophrenia, bipolar disorder and major depressive disorder. Importantly, we also showed for the first time on a large meta-analytic scale that MetS risk differs significantly across commonly used antipsychotic medications.

Knowledge of factors associated with the highest MetS risk can help identify individuals at greatest need for intensive monitoring and intervention. Consistent with population studies (33,34), we found no significant difference between men and women. Our results confirm earlier meta-analyses (22,35) in that MetS prevalence was higher in individuals with multi-episode schizophrenia compared with persons in their first episode. The current meta-analysis adds to the literature that a first episode diagnosis is even a unique predictor of lower MetS prevalence independent of mean age.

Also in line with data in the general population (36) and earlier work in people with schizophrenia (23), increasing age was a key predictor of MetS. When age and illness duration were entered into the same model, age was a more important determinant of MetS. However, this may also be due to the limited data available for illness duration compared to age data. Since age is a relevant risk factor for MetS in the general population too, the relative MetS risk

compared to the general population is greatest in younger people with SMI and those treated with antipsychotics (37,38). Considering the current meta-analytic data, it appears that a cumulative long-term effect of poor health behaviors and psychotropic medication use places people with SMI at the greatest risk for cardiometabolic disorders, more so than psychiatric diagnosis *per se*.

Our data suggest that patients receiving all individual antipsychotic medications are at higher MetS risk when compared to those who are antipsychotic-naïve. In line with the available literature (11,32,39-41), MetS risk was significantly higher with clozapine, followed by olanzapine. Moreover, MetS risk was significantly lower with aripiprazole than with each other antipsychotic for which data were available, including pooled typical antipsychotics, with the only exception of amisulpride. The lowest MetS prevalence for aripiprazole is noteworthy, as antipsychotics with lower cardiometabolic risk profiles in short-term studies are often prescribed for higher risk patients in clinical care, which may lead to a not reduced or even increased cardiometabolic risk in naturalistic settings (42).

Our meta-analysis also highlighted geographical differences in MetS, which indicates the possible influence of lifestyle and other environmental factors with or without genetic risk differences. This finding may, however, be somewhat affected by different MetS criteria, with IDF criteria, which are often used in Australian studies, being associated with the highest MetS prevalences. Nevertheless, people with SMI are more likely than the general population to have unhealthy lifestyle behaviors, such as being sedentary (43), smoking (44) and having diets that are high in saturated fats and refined sugars, while low in fruit and vegetables (45), placing them at higher risk for MetS and CVD than the general population. Thus, screening for and trying to minimize risk factors (including adverse lifestyle factors and antipsychotic medication choice and use) should be a key priority in the multidisciplinary treatment of people with SMI (46-49).

Whilst this is the most comprehensive and thorough meta-analysis of MetS in people with SMI conducted to date, we acknowledge some limitations that are largely related to the primary data. First, there was considerable methodological heterogeneity across studies. Second, because our study findings were based on cross-sectional rather than longitudinal data, directionality of the association between medication use and observed metabolic parameters cannot be deduced with certainty; that is, it is possible that those with inherently higher metabolic risk factors may be more likely to receive antipsychotics. Third, variables such as clinical subtypes of major depressive and bipolar disorder and concomitant or previous use of antidepressants and mood stabilizers were not reported or were insufficiently reported or controlled for in most available studies. Fourth, a threat to the validity of any meta-analysis is publication bias and heterogeneity, which we encountered in several of our analyses. However, although the main findings were heterogeneous, they were also highly robust and not influenced by publication bias, being virtually unaltered after applying the trim-and-fill method. Fifth, there were inadequate data on ethnic distribution and lifestyle behaviors, precluding meta-analytic assessment of these factors as moderating or mediating variables. Despite the above-mentioned caveats, this is the largest study of MetS risk and its moderators in people with SMI, and the first meta-analysis pooling and comparing all available data across patients with schizophrenia, bipolar disorder and major depressive disorder, comparing MetS risk across different antipsychotics and comparing the pooled risk of the three main SMI categories as well as the individual diagnostic groups with concurrently assessed, matched general population control groups.

Since antipsychotic medications are increasingly used as frontline treatments for bipolar disorder (50) and major depressive disorder (51), research on the underlying mechanisms for the development of metabolic abnormalities after pharmacotherapy initiation is urgently needed. Future studies should also examine whether different clinical subtypes of depression (i.e., melancholic or atypical) and bipolar disorder (e.g., type 1 or 2, mixed, cyclothymic), different mood states (manic, depressive or euthymic), or different antidepressant or mood stabilizers significantly modulate MetS risk. For example, previous studies (52) found that some antidepressants may, in some circumstances, reduce hyperglycaemia, normalize glucose homeostasis and also increase insulin sensitivity, whereas others, including tricyclic antidepressants, may exacerbate glycaemic dysfunction or have little effect on glucose homeostasis (53,54). Further, persons with atypical depression have significantly higher levels of inflammatory markers, body mass index, waist circumference and triglycerides, and lower HDL cholesterol than those with melancholic depression (55).

The pathophysiology underlying the association between SMI and MetS is complex and not well understood, requiring further investigation. Emerging evidence (56-59) suggests that they share pathophysiological features, including hypothalamic-pituitary-adrenal and mitochondrial dysfunction, neuro-inflammation, common genetic links and epigenetic interactions. Future research should comprehensively assess MetS risk factors and evaluate the optimal monitoring regimen and interventions. Moreover, long-term follow-up is required to accurately document the emergence of more distal outcomes, such as diabetes, ischemic heart disease, medical costs, and premature mortality (58).

## Acknowledgements

The authors are very grateful to the following researchers for sending additional data: K. Blank (Hartford Hospital, Hartford, CT, USA); R.K. Chadda (All India Institute of Medical Sciences, Ansari Nagar, New Delhi, India); R. Chengappa (University of Pittsburgh, School of Medicine, Pittsburgh, PA, USA); H.-W. Chiu (Taipei Medical University, Taipei, Taiwan); D. Cohen (Geestelijke Gezondheidszorg Noord -

Holland Noord, The Netherlands); T. Cohn (University of Toronto, Toronto, Canada); J. Crilly and J.S. Lamberti (University of Rochester Medical Center, Rochester, NY, USA); V. Ellingrod (University of Michigan College of Pharmacy, Ann Arbor, MI, USA); S. Grover (Postgraduate Institute of Medical Education and Research, Chandigarh, India); T. Heiskanen and H. Koponen (Kuopio University Hospital, Kuopio, Finland); P. Ifteni (Transilvania University, Brasov, Romania); F. Lamers (VU University Medical Centre Amsterdam, Amsterdam, The Netherlands); G.J. L'Italien (Yale University School of Medicine, New Haven, CT, USA); P. Mackin (Newcastle University, Newcastle upon Tyne, UK); J. Meyer (University of California, San Diego, CA, USA); H. Mulder (Utrecht University, Utrecht and Wilhelmina Hospital, Assen, The Netherlands); J.K. Patel (University of Massachusetts Medical School, Worcester, MA, USA); T. Sanchez-Araña Moreno (Hospital de la Merced, Osuna, Spain); K. Taxis (University of Groningen, Groningen, The Netherlands); P.J. Teixeira (Instituto de Previdência dos Servidores do Estado de Minas Gerais, Belo Horizonte, Brazil); S. Tirupati (James Fletcher Hospital, Newcastle, New South Wales, Australia); B. Vuksan (University Hospital Centre Zagreb, Zagreb, Croatia).

## References


1. Osborn DPJ, Levy G, Nazareth I et al. Relative risk of cardiovascular and cancer mortality in people with severe mental illness from the United Kingdom's General Practice Research Database. Arch Gen Psychiatry 2007;64:242-49.
2. Reininghaus U, Dutta R, Dazzan P et al. Mortality in schizophrenia and other psychoses: a 10-year follow-up of the ÆSOP first-episode cohort. Schizophr Bull 2015;41:664-73.
3. Chang CK, Hayes RD, Perera G et al. Life expectancy at birth for people with serious mental illness from a secondary mental health care case register in London, UK. PLoS One 2011;6:e19590.
4. Lawrence D, Hancock KJ, Kisely S. The gap in life expectancy from preventable physical illness in psychiatric patients in Western Australia: retrospective analysis of population based registers. BMJ 2013;346:f2539.
5. Saha S, Chant D, McGrath J. A systematic review of mortality in schizophrenia. Arch Gen Psychiatry 2007;64:1123-31.
6. Hoang U, Goldacre MJ, Stewart R. Avoidable mortality in people with schizophrenia or bipolar disorder in England. Acta Psychiatr Scand 2013;127:195-201.
7. De Hert M, Correll CU, Bobes J et al. Physical illness in patients with severe mental disorders. I. Prevalence, impact of medications and disparities in health care. World Psychiatry 2011;10:52-77.
8. Mitchell AJ, Lord O. Do deficits in cardiac care influence high mortality rates in schizophrenia? A systematic review and pooled analysis. J Psychopharmacol 2010;24(Suppl. 4):69-80.
9. Mitchell AJ, Lord O, Malone D. Differences in the prescribing of medication for physical disorders in individuals with v. without mental illness: meta-analysis. Br J Psychiatry 2012;201:435-43.
10. Mitchell AJ, Malone D, Doebbeling CC. Quality of medical care for people with and without comorbid mental illness and substance misuse: systematic review of comparative studies. Br J Psychiatry 2009;194:491-9.
11. De Hert M, Vancampfort D, Correll CU et al. Guidelines for screening and monitoring of cardiometabolic risk in schizophrenia: systematic evaluation. Br J Psychiatry 2011;199:99-105.
12. Mitchell AJ, Delaffon V, Vancampfort D et al. Guideline concordant monitoring of metabolic risk in people treated with antipsychotic medication: systematic review and meta-analysis of screening practices. Psychol Med 2012;42:125-47.
13. Gami AS, Witt BJ, Howard DE et al. Metabolic syndrome and risk of incident cardiovascular events and death: a systematic review and meta-analysis of longitudinal studies. J Am Coll Cardiol 2007;49:403-14.
14. Galassi A, Reynolds K, He J. Metabolic syndrome and risk of cardiovascular disease: a meta-analysis. Am J Med 2006;119:812-9.
15. Mottillo S, Filion KB, Genest J et al. The metabolic syndrome and cardiovascular risk: a systematic review and meta-analysis. J Am Coll Cardiol 2010;56:1113-32.
16. Expert Panel on Detection and Evaluation of High Blood Cholesterol in Adults. Executive summary of the third report of the expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III). JAMA 2001;285:2486-97.
17. Grundy SM, Cleeman JI, Daniels RS et al. Diagnosis and management of the metabolic syndrome: an American Heart/National Heart, Lung, and Blood Institute Scientific Statement. Circulation 2005;112:2735-52.
18. Alberti KG, Zimmet P, Shaw P. The metabolic syndrome, a new worldwide definition. A consensus statement from the International Diabetes Federation. Diab Med 2006;23:469-80.
19. World Health Organization. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus. Geneva: World Health Organization, 1999.
20. Alberti KG, Eckel RH, Grundy SM et al. A joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. Circulation 2009;120:1640-5.
21. Mitchell AJ, Vancampfort D, Sweers K et al. Prevalence of metabolic syndrome and metabolic abnormalities in schizophrenia and related disorders – a systematic review and meta-analysis. Schizophr Bull 2013;39:306-18.
22. Mitchell AJ, Vancampfort D, De Herdt A et al. Is the prevalence of metabolic syndrome and metabolic abnormalities increased in early schizophrenia? A comparative meta-analysis of first episode, untreated and treated patients. Schizophr Bull 2013;39: 295-305.
23. Vancampfort D, Vansteelandt K, Correll CU et al. Metabolic syndrome and metabolic abnormalities in bipolar disorder: a meta-analysis of prevalence rates and moderators. Am J Psychiatry 2013;170:265-74.
24. Vancampfort D, Correll CU, Wampers M et al. Metabolic syndrome and metabolic abnormalities in patients with depression: a meta-analysis of prevalence rates and moderators. Psychol Med 2014;94:2017-28.
25. Bartoli F, Carrà G, Crocamo C et al. Bipolar disorder, schizophrenia, and metabolic syndrome. Am J Psychiatry 2013;170:927-8.
26. Vancampfort D, Mitchell AJ, Correll CU et al. Response to Bartoli et al. Am J Psychiatry 2013;170:928-9.
27. Stroup DF, Berlin JA, Morton SC et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000;283:2008-12.
28. Moher D, Liberati A, Tetzlaff J et al. The PRISMA 660 Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. PLoS Med 2009;6:e1000097.
29. Egger M, Davey SG, Schneider M et al. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;315:629-34.
30. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics 1994;50:1088-101.

31. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7:177-88.
32. Duvall S, Tweedie R. A non-parametric 'trim and fill' method for assessing publication bias in meta-analysis. J Am Stat Assoc 2000; 95:89-98.
33. Ford ES, Giles WH, Dietz WH. Prevalence of the metabolic syndrome among US adults: findings from the third National Health and Nutrition Examination Survey. JAMA 2002;287:356-9.
34. Park YW, Zhu S, Palaniappan L et al. The metabolic syndrome: prevalence and associated risk factor findings in the US population from the Third National Health and Nutrition Examination Survey, 1988-1994. Arch Intern Med 2003;163:427-36.
35. Vancampfort D, Wampers M, Mitchell AJ et al. A meta-analysis of cardio-metabolic abnormalities in drug naïve, first-episode and multi-episode patients with schizophrenia versus general population controls. World Psychiatry 2013;12:240-50.
36. North BJ, Sinclair DA. The intersection between aging and cardiovascular disease. Circ Res 2012;110:1097-108.
37. De Hert M, van Winkel R, Van Eyck D et al. Prevalence of diabetes, metabolic syndrome and metabolic abnormalities in schizophrenia over the course of the illness: a cross-sectional study. Clin Pract Epidemiol Ment Health 2006;2:14.
38. De Hert M, Detraux J, van Winkel R et al. Metabolic and cardiovascular adverse effects associated with antipsychotic drugs. Nat Rev Endocrinol 2011;8:114-26.
39. Hammerman A, Dreiher J, Klang SH et al. Antipsychotics and diabetes: an age-related association. Ann Pharmacother 2008;42: 1316-22.
40. Nielsen J, Skadhede S, Correll CU. Antipsychotics associated with the development of type 2 diabetes in antipsychotic-naïve schizophrenia patients. Neuropsychopharmacology 2010;35:1997-2004.
41. Correll CU, Robinson DG, Schooler NR et al. Cardiometabolic risk in first episode schizophrenia-spectrum disorder patients: baseline results from the RAISE-ETP Study. JAMA Psychiatry 2014;71:1350-63.
42. Kessing LV, Thomsen AF, Mogensen UB et al. Treatment with antipsychotics and the risk of diabetes in clinical practice. Br J Psychiatry 2010;197:266-71.
43. Vancampfort D, Probst M, Knapen J et al. Associations between sedentary behaviour and metabolic parameters in patients with schizophrenia. Psychiatry Res 2012;200:73-8.
44. Dickerson F, Stallings CR, Origoni AE et al. Cigarette smoking among persons with schizophrenia or bipolar disorder in routine clinical settings, 1999-2011. Psychiatr Serv 2013;64:44-50.
45. Bly MJ, Taylor SF, Dalack G et al. Metabolic syndrome in bipolar disorder and schizophrenia: dietary and lifestyle factors compared to the general population. Bipolar Disord 2014;16:277-88.
46. De Hert M, Dekker JM, Wood D et al. Cardiovascular disease and diabetes in people with severe mental illness position statement from the European Psychiatric Association (EPA), supported by the European Association for the Study of Diabetes (EASD) and the European Society of Cardiology (ESC). Eur Psychiatry 2009;24:412-24.
47. McIntyre RS, Alsuwaidan M, Goldstein BI et al. The Canadian Network for Mood and Anxiety Treatments (CANMAT) task force recommendations for the management of patients with mood disorders and comorbid metabolic disorders. Ann Clin Psychiatry 2012;24:69-81.
48. Vancampfort D, De Hert M, Skjerven LH et al. International Organization of Physical Therapy in Mental Health consensus on physical activity within multidisciplinary rehabilitation programmes for minimising cardio-metabolic risk in patients with schizophrenia. Disabil Rehabil 2012;34:1-12.
49. Gierisch JM, Nieuwsma JA, Bradford DW et al. Pharmacologic and behavioral interventions to improve cardiovascular risk factors in adults with serious mental illness: a systematic review and meta-analysis. J Clin Psychiatry 2014;75:424-40.
50. Pillarella J, Higashi A, Alexander GC et al. Trends in use of second-generation antipsychotics for treatment of bipolar disorder in the United States, 1998-2009. Psychiatr Serv 2012;63: 83-6.
51. Davidson JR. Major depressive disorder treatment guidelines in America and Europe. J Clin Psychiatry 2010;71:e04.
52. Hennings JM, Schaaf L, Fulda S. Glucose metabolism and antidepressant medication. Curr Pharm Des 2012;18:5900-19.
53. Mojtabai R. Antidepressant use and glycemic control. Psychopharmacology 2013;227:467-77.
54. Correl CU, Detraux J, De Lepeleire J et al. Effects of antipsychotics, antidepressants and mood stabilizers on risk for physical diseases in patients with schizophrenia, depression and bipolar disorder. World Psychiatry 2015;14:119-36.
55. Lamers F, Vogelzangs N, Merikangas KR et al. Evidence for a differential role of HPA-axis function, inflammation and metabolic syndrome in melancholic versus atypical depression. Mol Psychiatry 2013;18:692-9.
56. Nousen EK, Franco JG, Sullivan EL. Unraveling the mechanisms responsible for the comorbidity between metabolic syndrome and mental health disorders. Neuroendocrinology 2013;98:254-66.
57. Manu P, Correll CU, Wampers M et al. Markers of inflammation in schizophrenia: association vs. causation. World Psychiatry 2014;13:189-92.
58. Rethorst CD, Bernstein I, Trivedi MH. Inflammation, obesity, and metabolic syndrome in depression: analysis of the 2009-2010 National Health and Nutrition Examination Survey (NHANES). J Clin Psychiatry 2014;75:1428-32.
59. Correll CU, Joffe BI, Rosen LM et al. Cardiovascular and cerebrovascular risk factors and events associated with second-generation antipsychotic compared to antidepressant use in a non-elderly adult sample: results from a claims-based inception cohort study. World Psychiatry 2015;14:55-62.

DOI 10.1002/wps.20252

# Ketamine for depression: evidence, challenges and promise

Carlos A. Zarate Jr., Mark J. Niciu

Experimental Therapeutics and Pathophysiology Branch, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892, USA

Major depressive disorder and bipolar depression are among the most prevalent and disabling mental disorders worldwide. Real-world effectiveness trials in major depressive disorder have underscored that most pharmacological options target monoamines, which are involved in a minority (15-20%) of synaptic contacts in the mammalian brain.

Most synapses (~50%) use the amino acid glutamate as their primary neurotransmitter, and preclinical models of depression have implicated aberrant glutamatergic neurotransmission for 25 years (1). More recently, the N-methyl-D-aspartate (NMDA) glutamate receptor antagonist ketamine was shown to produce rapid and robust antidepressant effects in patients with treatment-resistant major depressive disorder and bipolar depression (2-7).

## EVIDENCE

Ketamine is a non-competitive NMDA receptor antagonist that works as an open channel pore blocker at the phencyclidine binding site, thereby preventing the flux of cations (primarily calcium) and neuronal excitation/depolarization.

Several randomized, placebo-controlled trials of sub-anesthetic dose ketamine infusions (0.5 mg/kg for 40 min) have been conducted in individuals with major depressive disorder, including those with treatment-resistant depression (2-4). Sub-anesthetic dose ketamine also has similar antidepressant efficacy in treatment-resistant bipolar depression subjects maintained on mood stabilizers (5), and has not demonstrated increased affective switching to hypo/mania over placebo (8).

Ketamine has also been shown to rapidly reduce suicidal thinking (6,7). Because few evidence-based treatments for suicidality exist – none of which have rapid onset – ketamine may be a promising rapid-acting antidepressant treatment option in emergency and acute inpatient psychiatry.

Finally, repeated sub-anesthetic dose ketamine infusions have demonstrated preliminary efficacy and safety/tolerability in brief trials (9).

All the aforementioned placebo-controlled trials have used racemic ketamine mixtures. The S-enantiomer (S-ketamine/esketamine) has three- to four-fold greater affinity for the NMDA receptor, which may permit lower dosing to preserve antidepressant efficacy while limiting undesirable side effects.

Non-intravenous modes of ketamine administration have also been studied (intramuscular, subcutaneous, oral, sublingual and intranasal), with mixed efficacy but typically fewer side effects than intravenous infusion.

Specific ketamine metabolites have also been shown to correlate with antidepressant response (10); some have affinity for non-NMDA receptors (e.g., antagonism of alpha 7 nicotinic acetylcholine receptors) (11), which may also contribute to their antidepressant mechanism of action.

In that regard, ketamine's antidepressant mechanism has been an active topic of preclinical and clinical investigation. Ketamine-induced NMDA receptor antagonism of inhibitory gamma-aminobutyric acid (GABA)ergic cortical interneurons has been shown to release tonic inhibition of excitatory (glutamatergic) pyramidal neurons to increase synaptic glutamate release (acute "glutamate surge") (12). Because postsynaptic NMDA receptors are blocked, synaptic glutamate can then preferentially bind to and activate alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors (13). Postsynaptic membrane depolarization then initiates intracellular second messenger/signal transduction cascades, resulting in some or all of the following: mammalian target of rapamycin (mTOR) activation (14), increased brain-derived neurotrophic factor (BDNF) translation and secretion (15), and glycogen synthase kinase-3 (GSK-3) inhibition (16). These acute molecular and cellular responses to ketamine stimulate synaptic plasticity.

Clinical neurobiological studies have focused on more amenable units of analysis, namely genetics, peripheral measures, and neuroimaging (including sleep electroencephalography). Some of them may be critical mediators of antidepressant response to ketamine, including BDNF genotype, circulating BDNF levels, acute changes in central glutamate levels, synaptic potentiation, and circuit-level de/synchronization (17).

## CHALLENGES

Although a handful of psychiatrists and anesthesiologists are currently administering ketamine in office-based outpatient settings, significant challenges exist for ketamine's potential broader dissemination in treating major depression. First, ketamine is not approved by the U.S. Food and Drug Administration (FDA) for any depressive disorder; this lack of indication may hamper dissemination, with concomitant public health implications.

Yet, for ketamine to be more widely disseminated clinically, standardization of best practices must be put in place for optimal mode of administration, dosing and frequency. Several studies are currently investigating alternative modes

of administration (e.g., intranasal esketamine), but none are comparing different modes head-to-head.

For optimal dosing, with the exception of a small (N=4), placebo-controlled, crossover study (18), all randomized controlled trials in both treatment-resistant major depressive disorder and bipolar depression have used the same dose (0.5 mg/kg). However, ketamine's antidepressant dose-response is currently being investigated in a multi-site, psychoactive placebo-controlled, parallel-design trial with midazolam 0.045, ketamine 0.1, ketamine 0.2, ketamine 0.5, and ketamine 1.0 mg/kg infusions.

It should also be noted that, because antidepressant response to ketamine is typically short-lived, evidence-based strategies to maintain response/prevent relapse are critical for clinical practice. The most logical and potentially efficacious strategy is repeated dosing ("boosters"), similar to maintenance therapy in electroconvulsive therapy (ECT). However, few published studies exist of repeated ketamine infusions in active major depressive disorder (9,19,20), and these have typically only offered <10 infusions over 12-21 days. Notably, the risk for abuse and potential long-term side effects – for instance, cognitive sequelae, urinary cystitis – may increase manifold with repeated dosing. These risks necessitate close clinical follow-up and/or appropriate consultation.

Oral glutamatergic modulators are also reasonable strategies to prevent relapse. In a four-week, randomized, placebo-controlled trial, the oral glutamatergic modulator riluzole did not maintain ketamine's antidepressant effect over placebo, but the effect size was large (d=0.78), suggesting that the study might have been underpowered (21). D-cycloserine, a partial agonist at the NMDA receptor glycine site, demonstrated preliminary efficacy for treatment-refractory bipolar depression in a small (N=7), open-label, eight-week study in which a daily escalating dose of the drug followed the administration of sub-anesthetic dose ketamine (22). In addition, ketamine co-administration with standard oral antidepressants/mood stabilizers has shown preliminary efficacy in preclinical studies (23,24), but has not yet been investigated in controlled studies in humans.

In initial trials, ketamine had a large-to-very-large antidepressant effect, with maximal efficacy within 24 hours and relapse often within one week. Nevertheless, ketamine's antidepressant efficacy difference was exaggerated by minimal response to the inert placebo (intravenous saline).

In the largest randomized controlled trial to date with a psychoactive placebo (intravenous midazolam), ketamine still separated at 24 hours post-infusion. Ketamine's drug difference, however, was not as great, due to the more typical placebo response for a major depressive disorder trial (4). Although a better placebo than saline, midazolam also has its flaws – for instance, minimal acute dissociative side effects – that may compromise the integrity of the blind in savvy patients. Future research challenges include developing a better control condition than midazolam and formally assessing randomization expectancies.

Another potential hurdle is the identification and replication of enriched subgroups with augmented antidepressant response to ketamine. Our group has identified several non-overlapping clinical predictors of ketamine's antidepressant efficacy, including increased body mass index, family history of alcohol use disorder in a first-degree relative, and dimensional anxious depression (25).

In addition to these clinical descriptive parameters, several genetic, central neurobiological, and peripheral measures have also been shown to correlate with ketamine's antidepressant efficacy (26,27). Nevertheless, few studies have combined measures/datasets in order to increase predictive power and detect smaller effects. Due to the heterogeneity of major depressive disorder, this combinatorial approach may best be undertaken by forming a multi-site ketamine depression consortium to maximize the sample size of enriched subgroups for prospective mechanistic studies.

Another issue of concern is that a sensitive and specific human biomarker of glutamate function – for instance, an NMDA receptor subunit positron emission tomography (PET) ligand – has yet to be developed. In an *in vitro* model system, intracellular second messenger/signal transduction mediators and effectors hypothesized to be critical to ketamine's antidepressant efficacy (e.g., mTOR phosphorylation and inhibition of eukaryotic elongation factor-2 kinase (eEF2)) may also reflect glutamate receptor engagement. Again, such systems would not only improve our understanding of ketamine's antidepressant mechanisms, but would also prove very useful for glutamate-based drug screening.

Although several depression-like induction paradigms have been developed in healthy volunteers – for instance, acute monoamine reduction (reserpine, dietary tryptophan depletion) and "sickness syndrome"-like induction (lipopolysaccharide) – these models have yet to be tested and/or reported as ketamine-responsive.

Ketamine also enhances resiliency to stress in rodent models of despair (28,29) and may have analogous effects in humans. This may facilitate the rapid screening of candidate glutamate-based drugs in healthy volunteers, thereby reducing resource allocation to drugs likely to fail early in the clinical pipeline.

In summary, a sensitive and specific ketamine-responsive model system remains a substantial challenge for future research.

## PROMISE

The discovery of ketamine's rapid and robust antidepressant efficacy has provided hope for patients with treatment-resistant depression and depression researchers alike. This promise is two-fold: a) the identification of novel glutamate-based mechanisms of disease and treatment response in depressive disorders; and b) the availability of a first-in-class, rapid-acting antidepressant medication. Ketamine's preliminary efficacy for suicidality, where swift onset and significant

response are absolutely vital, also provides promise as a prototypical anti-suicidal treatment.

Finally, in addition to adult treatment-resistant major depressive disorder and bipolar depression, ketamine has also demonstrated preliminary efficacy and/or is currently being studied in other disorders, which may ultimately extend its utility in clinical practice. These include pediatric/adolescent depression and behavioral dysregulation, autism, obsessive-compulsive disorder, post-traumatic stress disorder, and depression comorbid with alcohol dependence.


## Acknowledgements

Funding for this work was provided by the Intramural Research Program at the National Institute of Mental Health, National Institutes of Health, a NARSAD Independent Investigator grant, and a Brain & Behavior Mood Disorders Research Award to C.A. Zarate. The authors thank the 7SE research unit and staff for their support. I. Henter provided excellent editorial assistance.


## References


1. Skolnick P, Popik P, Trullas R. Glutamate-based antidepressants: 20 years on. Trends Pharmacol Sci 2009;30:563-9.
2. Berman RM, Cappiello A, Anand A et al. Antidepressant effects of ketamine in depressed patients. Biol Psychiatry 2000;47:351-4.
3. Zarate CA Jr., Singh JB, Carlson PJ et al. A randomized trial of an N-methyl-D-aspartate antagonist in treatment-resistant major depression. Arch Gen Psychiatry 2006;63:856-64.
4. Murrough JW, Iosifescu DV, Chang LC et al. Antidepressant efficacy of ketamine in treatment-resistant major depression: a two-site randomized controlled trial. Am J Psychiatry 2013;170:1134-42.
5. Diazgranados N, Ibrahim L, Brutsche NE et al. A randomized add-on trial of an N-methyl-D-aspartate antagonist in treatment-resistant bipolar depression. Arch Gen Psychiatry 2010;67:793-802.
6. Diaz Granados N, Ibrahim LA, Brutsche NE et al. Rapid resolution of suicidal ideation after a single infusion of an N-methyl-D-aspartate antagonist in patients with treatment-resistant major depressive disorder. J Clin Psychiatry 2010;71:1605-11.
7. Price RB, Iosifescu DV, Murrough JW et al. Effects of ketamine on explicit and implicit suicidal cognition: a randomized controlled trial in treatment-resistant depression. Depress Anxiety 2014;31:335-43.
8. Niciu MJ, Luckenbaugh DA, Ionescu DF et al. Subanesthetic dose ketamine does not induce an affective switch in three independent samples of treatment-resistant major depression. Biol Psychiatry 2013;74:e23-4.
9. Murrough JW, Perez AM, Pillemer S et al. Rapid and longer-term antidepressant effects of repeated ketamine infusions in treatment-resistant major depression. Biol Psychiatry 2013;74:250-6.
10. Zarate CA Jr., Brutsche N, Laje G et al. Relationship of ketamine's plasma metabolites with response, diagnosis, and side effects in major depression. Biol Psychiatry 2012;72:331-8.
11. Moaddel R, Abdrakhmanova G, Kozak J et al. Sub-anesthetic concentrations of (R,S)-ketamine metabolites inhibit acetylcholine-evoked currents in alpha7 nicotinic acetylcholine receptors. Eur J Pharmacol 2013;698:228-34.
12. Moghaddam B, Adams B, Verma A et al. Activation of glutamatergic neurotransmission by ketamine: a novel step in the pathway from NMDA receptor blockade to dopaminergic and cognitive disruptions associated with the prefrontal cortex. J Neurosci 1997;17:2921-7.
13. Maeng S, Zarate CA Jr., Du J et al. Cellular mechanisms underlying the antidepressant effects of ketamine: role of alpha-amino-3-hydroxy-5-methylisoxazole-4-propionic acid receptors. Biol Psychiatry 2008;63:349-52.
14. Li N, Lee B, Liu RJ et al. mTOR-dependent synapse formation underlies the rapid antidepressant effects of NMDA antagonists. Science 2010;329:959-64.
15. Autry AE, Adachi M, Nosyreva E et al. NMDA receptor blockade at rest triggers rapid behavioural antidepressant responses. Nature 2011;475:91-5.
16. Beurel E, Song L, Jope RS. Inhibition of glycogen synthase kinase-3 is necessary for the rapid antidepressant effect of ketamine in mice. Mol Psychiatry 2011;16:1068-70.
17. Zarate CA Jr., Mathews DC, Furey ML. Human biomarkers of rapid antidepressant effects. Biol Psychiatry 2013;73:1142-55.
18. Lai R, Katalinic N, Glue P et al. Pilot dose-response trial of i.v. ketamine in treatment-resistant depression. World J Biol Psychiatry 2014;15:579-84.
19. Rasmussen KG, Lineberry TW, Galardy CW et al. Serial infusions of low-dose ketamine for major depression. J Psychopharmacol 2013;27:444-50.
20. Diamond PR, Farmery AD, Atkinson S et al. Ketamine infusions for treatment resistant depression: a series of 28 patients treated weekly or twice weekly in an ECT clinic. J Psychopharmacol 2014;28:536-44.
21. Niciu MJ, Luckenbaugh DA, Ionescu DF et al. Ketamine's antidepressant efficacy is extended for at least four weeks in subjects with a family history of an alcohol use disorder. Int J Neuropsychopharmacol 2014;18:10.1093.
22. Kantrowitz JT, Halberstam B, Gangwisch J. Single-dose ketamine followed by daily D-cycloserine in treatment-resistant bipolar depression. J Clin Psychiatry 2015;76:737-8.
23. Zhang GF, Liu WX, Qiu LL et al. Repeated ketamine administration redeems the time lag for citalopram's antidepressant-like effects. Eur Psychiatry 2015;30:504-10.
24. Melo A, Kokras N, Dalla C et al. The positive effect on ketamine as a priming adjuvant in antidepressant treatment. Transl Psychiatry 2015;5:e573.
25. Niciu MJ, Luckenbaugh DA, Ionescu DF et al. Clinical predictors of ketamine response in treatment-resistant major depression. J Clin Psychiatry 2014;75:e417-23.
26. Niciu MJ, Henter ID, Luckenbaugh DA et al. Glutamate receptor antagonists as fast-acting therapeutic alternatives for the treatment of depression: ketamine and other compounds. Annu Rev Pharmacol Toxicol 2014;54:119-39.
27. Ionescu DF, Luckenbaugh DA, Niciu MJ et al. Effect of baseline anxious depression on initial and sustained antidepressant response to ketamine. J Clin Psychiatry 2014;75:e932-8.
28. Yilmaz A, Schulz D, Aksoy A et al. Prolonged effect of an anesthetic dose of ketamine on behavioral despair. Pharmacol Biochem Behav 2002;71:341-4.
29. Brachman RA, McGowan JC, Perusini JN et al. Ketamine as a prophylactic against stress-induced depressive-like behavior. Biol Psychiatry (in press).


DOI 10.1002/wps.20269

# The impact of war on mental health: lest we forget

ALEXANDER C. MCFARLANE

Centre for Traumatic Stress Studies, University of Adelaide, Adelaide, Australia

The often-unconscious and enduring impact of war is one of the driving forces of history. Yet these terrible costs and the lessons learned by psychiatry tend to be forgotten (1). At a time when many nations are remembering the legacy of World War I, the greatest military conflagration in history, it is timely to reflect on what has been learnt about the impacts of war on mental health.

Ironically, it is only since the inclusion of post-traumatic stress disorder (PTSD) in the DSM-III in 1980 (2) that the field of traumatic stress has blossomed and been subsequently underpinned by a major body of neuroscience and clinical research.

Despite the slow development of interest into the long-term consequences of the traumatic stress of war, many of the developments in mental health care in the 20th century emerged from the innovations demanded by the need to deal more effectively with the flood of mental casualties amongst the combatants of World Wars I and II. The model of community psychiatry was adapted from the model of forward psychiatry developed by the military to deal with acute combat stress reactions; this model was underpinned by the principles of the provision of early treatment close to the battle front with the expectancy of recovery and return to service (1). Crisis intervention, group therapy and therapeutic communities were innovations that evolved out of the military medical corps (1).

However, psychiatrists who served in the military were often conflicted by powerful and potentially competing value systems concerning whether their primary responsibility was to the soldier or to the war effort (3). The prevailing attitudes would tend to indicate that individuals' interests often lost out – the veteran who broke down in battle was generally stigmatized. The diagnoses promulgated by the profession, such as compensation neurosis, *lack of moral fibre* and inadequate personality reflected how the problem was seen to be caused by moral weakness and vulnerability (4). In this characterization, the causal role of the horrors of combat were minimized by psychiatry, in contrast to compensation seeking and vulnerability.

## THE IMPACT OF THE VIETNAM WAR

It was in the ferment of the protests against the Vietnam War in the U.S. that veterans, partly as a consequence of their political activism, were able to lobby for an independently conducted study of the impact of their war service. The National Vietnam Veterans' Readjustment Study was a turning point in defining the psychological costs of war, with 18.7% having a lifetime history of PTSD (5). These costs of traumatic war stress extended beyond PTSD, to the increased risk for depression, personality disorders, suicide, and alcohol abuse (6).

The Vietnam veterans' battle to gain recognition for their psychological injuries fostered an acceptance of the diagnosis of PTSD and the development of the field of traumatic stress studies. This knowledge, in turn, led to recognition of the plight of the psychological welfare of civilian casualties of war internationally, such as refugees and victims of torture, and their special needs for care.

Activist psychiatrists in the Vietnam veterans' movement also documented the impact of other horrors of war, including the bombing of Hiroshima and the brain washing of prisoners of war (7). This work contributed to the medical profession's important role in the antinuclear movement.

## THE CYCLES OF VIOLENCE OF WAR

There has been an increasing awareness of the moral injury suffered by combatants and the particular costs of the act of killing on mental health, which involves a fundamental violation of an inbuilt prohibition, overridden by military training (8). The violence associated with PTSD impacts on veterans' families, as well as on the broader society (9). Tragically, it is these psychological costs that can lead to cycles of violence, both within the communities that have been at war and between nations seeking revenge and reparation.

These enduring effects of violence have become of particular concern with the current conflicts in the Middle East and the associated terrorist movements that seek to lure young men and women to their violent cause. Psychiatry has a responsibility to contribute to more sophisticated understandings of these cycles of hatred and how to stop violence propagating itself. These destabilizing consequences of war lead to major humanitarian crises and enduring psychological legacies that contribute to poverty in nations such as Rwanda and Somalia.

## EPIDEMIOLOGICAL STUDIES OF WAR

In the last three decades, there has been a major research effort to understand the broad costs of military service in combat, often in response to lobbying by the veteran community in Western nations. Fears about the health consequences of exposures to chemical, nuclear and biological weapons, for example after the first Persian Gulf War, have been

significant drivers to these research programs (10). However, the consistent finding has been the increased rates of psychiatric disorder due to combat exposure. These long-term studies of veterans have also demonstrated that there is a frequent pattern of delayed onset PTSD, confirming the reality of the prolonged risk arising from combat exposure (11).

More recently, a debate has emerged about the apparent differences in the prevalence of psychiatric disorder in the U.S. and U.K. veterans from recent wars in the Middle East (12). This controversy has arisen, in part, because of the different methodologies used in studying veterans. However, when self-reported combat exposure is taken into account, most of the differences in the reported prevalence of PTSD between the U.S. and U.K. forces no longer exist (12).

The long-term cost of combat was examined over a 43 year period of pension entitlement records of an entire cohort of 60,228 Australian Vietnam veterans, documenting that 47.9% had accepted claims for a mental health condition (13). The persistent risk of emerging disorder has also been demonstrated in a longitudinal study of Israeli combat veterans, which also showed the benefits of frontline treatment for combat stress reactions (11). These findings suggest that studies reporting mental health outcomes relatively soon after deployment are likely to underestimate the total cost of war.

A further challenge in interpreting the impact of combat exposure on rates of psychiatric disorder relates to the "healthy warrior" effect (14). This phenomenon has been shown in representative samples, where those who deploy are more resilient and psychologically healthy prior to deploying than those who do not deploy (14). These differences make epidemiological comparisons of disorder prevalence between deployed and non-deployed groups and the community more difficult to interpret, as these background differences hide the detrimental effects of combat. The demonstrated gradient between the intensity and duration of combat exposure and its adverse mental health impacts is the critical issue (5).

## THE SOMATIC AND BIOLOGICAL CONSEQUENCES OF COMBAT EXPOSURE

Post-deployment studies have highlighted the importance of the somatic manifestations of psychological distress. Veterans often fear that their physical symptoms are indicative of exposure to environmental toxins, and are reluctant to accept that they relate to psychological trauma of war. Combat-related PTSD has been found to increase the risk of a range of chronic diseases (15). Importantly, there appears to be both a direct effect of the stress of combat exposure on the presence of chronic disease and mortality, as well as this being amplified by the presence of PTSD (16). There has also been considerable interest in the morbidity of mild traumatic brain injury, both as a separate problem as well as a risk factor for PTSD (17).

Prospective designs have been used to investigate the neurobiological effects of combat and have shown how perceived combat stress modifies amygdala coupling with the insular and dorsal anterior cingulate circuits, which are related to fear reactivity and somatic self-awareness (18). Abnormalities of the hypothalamic-pituitary-adrenal (HPA) axis have also been examined, with studies demonstrating that higher numbers of glucocorticoid receptors pre-deployment predict the risk of developing PTSD symptoms (18). These studies have provoked considerable interest in the possibility of developing biomarkers for PTSD and their ability to predict the long-term emergence of a disorder (18).

These enduring effects of war have also been shown to transmit inter-generationally in the offspring of Holocaust survivors, through the maternal transmission of increased glucorticoid receptor sensitivity, a risk factor for PTSD (19). Hence, war impacts on the next generation neurobiologically, as well as through the impaired attachment behaviour of PTSD sufferers (19).

## CONCLUSIONS

Documenting the psychological costs of war is important, as it powerfully argues for the need to globally improve the treatment services for veterans and effected civilians alike. The substantial research effort into studying veteran populations has also contributed broadly to the understanding and acceptance of the effects of traumatic stress in society and focus attention on the need for improved services. However, despite advances in evidence-based care, substantial morbidity remains, highlighting the need for innovation in treatments and rehabilitation.

Political leaders need to remember these long-term indelible consequences when they consider declaring war. The ultimate method of prevention is to stop war, an aspiration that is tragically at odds with human nature.

## References


1. Glass AJ. Mental health programs in the armed forces. In: Arieti S (ed). American handbook of psychiatry, 2nd ed. New York: Basic Books, 1974.
2. American Psychiatric Association. Diagnostic and statistical manual of mental disorders, 3rd ed. Washington: American Psychiatric Association, 1980.
3. Camp NM. The Vietnam War and the ethics of combat psychiatry. Am J Psychiatry 1993;150:1000-10.
4. Jones E, Palmer I, Wessely S. War pensions (1900-1945): changing models of psychological understanding. Br J Psychiatry 2002; 180:374-9.
5. Dohrenwend BP, Turner JB, Turse NA et al. The psychological risks of Vietnam for US veterans: a revisit with new data and methods. Science 2006;313:979-82.
6. Tanielian TL, Jaycox L. Invisible wounds of war: psychological and cognitive injuries, their consequences, and services to assist recovery (Vol. I). Santa Monica: Rand Corporation, 2008.

7. Lifton RJ. Witness to an extreme century: a memoir. New York: Free Press, Simon and Schuster, 2011.
8. Maguen S, Lucenko BA, Reger MA et al. The impact of reported direct and indirect killing on mental health symptoms in Iraq war veterans. J Trauma Stress 2010;23:86-90.
9. MacManus D, Dean K, Jones M et al. Violent offending by UK military personnel deployed to Iraq and Afghanistan: a data linkage cohort study. Lancet 2013;381:907-17.
10. Institute of Medicine (US) Committee on Gulf War and Health. Health effects of serving in the Gulf War, update 2009. Gulf War and Health: Volume 8: Update of health effects of serving in the Gulf War. Washington: National Academies Press (US), 2010.
11. Solomon Z, Mikulincer M. Trajectories of PTSD: a 20-year longitudinal study. Am J Psychiatry 2006;163:659-66.
12. Sundin J, Herrell RK, Hoge CW et al. Mental health outcomes in US and UK military personnel returning from Iraq. Br J Psychiatry 2014;204:200-7.
13. Clarke PM, Gregory R, Salomon JA. Long-term disability associated with war-related experience among Vietnam veterans: retrospective cohort study. Med Care 2015;53:401-8.
14. Larson GE, Highfill-McRoy RM, Booth-Kewley S. Psychiatric diagnoses in historic and contemporary military cohorts: combat deployment and the healthy warrior effect. Am J Epidemiol 2008; 167:1269-76.
15. McFarlane AC. The long-term costs of traumatic stress: intertwined physical and psychological consequences. World Psychiatry 2010;9:3-10.
16. Boscarino JA. Posttraumatic stress disorder and mortality among U.S. Army veterans 30 years after military service. Ann Epidemiol 2006;16:248-56.
17. Rosenfeld JV, McFarlane AC, Bragge P et al. Blast-related traumatic brain injury. Lancet Neurol 2013;12:882-93.
18. Vermetten E, Baker D, Yehuda R. New findings from prospective studies. Psychoneuroendocrinology 2015;51:441-3.
19. 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.

DOI 10.1002/wps.20253

# Problem Management Plus (PM+): a WHO transdiagnostic psychological intervention for common mental health problems

Katie S. Dawson[1], Richard A. Bryant[1], Melissa Harper[2,3], Alvin Kuowei Tay[1,4], Atif Rahman[5], Alison Schafer[6], Mark van Ommeren[2]

[1]University of New South Wales, Sydney, Australia; [2]World Health Organization, Geneva, Switzerland; [3]University of Geneva, Geneva, Switzerland; [4]Psychiatry Research and Teaching Unit, Liverpool Hospital, Sydney, Australia; [5]University of Liverpool, Liverpool, UK; [6]World Vision Australia/International

Mental disorders are significant contributors to the global burden of disease (1). While they occur across all levels of socio-economic status, the majority of populations in low- and middle-income countries (LMICs) do not have access to effective psychological and pharmacological interventions (2). Key barriers to sustainable delivery of psychological therapies in LMICs include limited mental health funding and infrastructure, chronic shortage of mental health professionals, lack of treatments adapted to the local context, and challenges associated with training and supervision. Implementation of low-intensity psychological interventions by trained para-professionals is one potential solution to this problem (3,4) which is receiving significant attention as part of global mental health research agendas (e.g., 5).

A number of low-intensity interventions have demonstrated clinical benefit and utility in high-income settings. For example, early analyses of the UK's Improving Access to Psychological Therapies programme (IAPT, 6) found a substantial reduction in depression and anxiety in people who attended at least two sessions of low-intensity interventions. Additionally, a recent meta-analysis challenged conventional thinking and provided support for low-intensity interventions as an effective treatment even for individuals with symptoms of severe depression (7).

Evidence for the applicability of psychological interventions by non-specialists in LMICs is mounting (8,9). For instance, group interpersonal psychotherapy facilitated by local para-professionals has been shown to be effective in rural Uganda among depressed adults compared to usual care at six month follow-up (10). In rural Pakistan, Rahman et al (11) found that local community health workers could effectively deliver a locally adapted cognitive-behavioural intervention for perinatal depression. Mothers receiving the treatment demonstrated significant clinical improvement on depression symptoms, showed less disability and better overall and social functioning. Finally, a comparatively more intensive transdiagnostic intervention, the Common Elements Treatment Approach (CETA), has shown promising results for the treatment of symptoms of depression, anxiety and post-traumatic stress in Burmese refugees when delivered by para-professionals (12).

To fill the gap between mental health needs and access to quality care, and extend the current research on low-intensity interventions in LMICs, the World Health Organization (WHO) – as part of its Mental Health Gap Action Programme (mhGAP) – has begun to develop and test low-intensity psychological interventions. The current paper focuses on one such intervention, named Problem Management Plus (PM+).

## THERAPEUTIC FOUNDATIONS FOR PM+

PM+ is for adults suffering from symptoms of common mental health problems (e.g., depression, anxiety, stress or grief), as well as self-identified practical problems (e.g., unemployment, interpersonal conflict). It is not suitable for people presenting with severe mental health problems (e.g., those with psychosis or at imminent risk for suicide).

One of the core features underpinning PM+ is adherence to a transdiagnostic approach. Transdiagnostic treatments are "those that apply the same underlying principles across mental disorders, without tailoring the protocol to specific diagnoses" (13). This approach can be very useful because most people present with comorbidity. Addressing multiple problems at one time through shared emotional mechanisms is more efficient (14).

Considerable momentum has developed in high-income settings for the use of transdiagnostic treatments, with initial evidence pointing to their efficacy in alleviating common mental health problems (15). A key advantage of these approaches for LMICs is that they reduce the need for and challenge of making differential diagnoses and learning multiple treatment manuals for different disorders (16).

Reflecting this therapeutic approach, PM+ has integrated problem-solving and behavioural treatment techniques that demonstrate amenability to low-intensity delivery while seeking to preserve a strong evidence base (3,17). Following review by 24 international experts, four core therapeutic strategies (described below) were included in the manual. There was a strong emphasis on behavioural (as opposed to cognitive) techniques, as these would likely be easier to learn by individuals and lay helpers.

## DESCRIPTION OF PM+

The name "Problem Management Plus" is intended to reflect the therapeutic aims of the intervention: to improve one's management over practical problems (e.g., unemployment, interpersonal conflict) and associated common mental health problems. The term "problem management" is used rather than "problem solving" to highlight that many practical problems encountered by people living in adversity may not necessarily be solvable. The "plus" refers to the evidence-based behavioural strategies that enhance one's capacity to adaptively manage emotional problems.

In PM+, subjects are seen on an individual, face-to-face basis for five weekly sessions with a lay helper. The length of the sessions is 90 min, to allow adequate time for explanation of a strategy and application to client-identified problems. Independent practice of strategies between sessions is encouraged and reviewed in subsequent sessions, thus enhancing learning through repetition.

In addition to the four core strategies, PM+ includes a psychoeducation component, delivered in session one. Individuals are educated about common reactions to adversity and receive a general overview and rationale of the intervention. A brief motivational interviewing component is included to enhance one's commitment to being actively engaged in PM+.

## CORE STRATEGIES OF PM+

### Managing stress

Lay helpers introduce a simple stress management strategy in session one, to optimize initial mastery of stress and anxiety symptoms as well as enhance relaxation. Stress management has been identified as an effective strategy in the treatment of post-traumatic stress disorder and depression, albeit less effective than high-intensity intervention strategies, such as cognitive-behavioural therapy (18-21). Within PM+, slow breathing is taught, given its ease of learning, likelihood of being acceptable in different cultural contexts, and potential to be delivered succinctly.

### Managing problems

From session two, people are taught basic skills to help them address practical problems. In PM+, this strategy extends the traditional six-step problem solving format (22) to emulate Bowen et al's problem solving approach (23). It includes categorizing problems as solvable, unsolvable and unimportant prior to selecting a target problem. This step aims to support individuals to take control of their problems by determining what is important to them and investing solely in those problems considered of significance to their lives. This approach has been shown to have promising results in randomized controlled trials in high-income contexts (24) and also in a South African pilot study (8).

### Get going, keep doing

This behavioural activation strategy aims to increase the opportunity for positive reinforcement from the environment and directly address inertia, a distinguishing feature of depression (25,26). Numerous studies have demonstrated that behavioural activation is an effective means to reduce depressed mood (27). In PM+, individuals are encouraged to re-engage gradually with pleasant and task-oriented activities to improve mood and functionality. This strategy is introduced in the second session.

### Strengthening social support

A distinct strategy to promote social support was retained in the final intervention manual. It aims to optimize a person's capacity to re-engage in the community, elicit support (e.g., emotional, practical) from others or specific agencies, and provide support him/herself. By the end of session three, the individual has likely gained some personal locus of control and mastery over his/her symptoms, and so strengthening his/her social support is considered. Perceived social support has been found to be a robust construct associated with better psychological outcome in a variety of populations, including those exposed to traumatic events (28-30).

## RELAPSE PREVENTION

Relapse prevention education is delivered in the final session. This involves identifying personal warning signs of relapse, gently testing people's knowledge of strategies, including how best to apply them to manage specific problems, and identifying future goals.

## GROUP ADAPTATION

To enhance cost-effectiveness and accessibility, PM+ has recently been adapted to be delivered in a group setting (and plans are underway for development and testing in e-format). All core treatment components, session content and frequency have been retained in the group version. However, treatment sessions last three hours, to accommodate the unique dynamics of delivering an intervention to a group (e.g., group discussions) and include group rituals and breaks. A ratio of no greater than one facilitator to eight participants has been recommended. Facilitators are expected to have a similar profile as that of individual PM+ lay helpers (discussed below) and undergo a brief training course

specific to group PM+. This variation of the intervention is currently being tested in rural Pakistan.

## WHO PROVIDES PM+?

Upholding a task shifting approach, PM+ is intended to be delivered by lay helpers who have completed at least high school but without previous mental health training. Some disparity exists in the literature with regard to the duration of training of lay helpers. While some interventions have adopted longer training programmes, such as six weeks (31) and two months (32), the majority of studies of this nature have demonstrated effective outcomes after one to four weeks of training (e.g., 33-37). Briefer training periods are more feasible in many communities with resource- and time-related restrictions.

PM+ thus far implements an eight-day training programme, followed by a two to three week period of in-field practice with ongoing, weekly supervision. Supervision is conducted by skilled mental health professionals who have received PM+ training and have experience in its delivery.

## IS PM+ ADAPTABLE ACROSS CONTEXTS?

Many psychological interventions are developed for delivery in high-income contexts, and socio-cultural acceptability is a critical barrier to improving access to effective treatment in LMICs. Chowdhary et al (38) have identified key components that require adaptation in different cultures (e.g. language, content, use of local idioms of distress and metaphors), for which PM+ has sought to uphold.

Socio-cultural adaptations of PM+ to local contexts through formative studies are encouraged prior to implementation of the intervention. Such studies have been carried out in Pakistan and Kenya, confirming that PM+ can provide a template that is adaptable to various contexts.

## CONCLUSIONS

The WHO and its partners have produced a low-intensity intervention aimed at reducing symptoms of common mental disorder in people living in communities affected by adversity, whether they are humanitarian settings or low-income urban settings exposed to community violence. PM+ has been developed in response to the urgent need for affordable but evidence-based interventions that are amenable to low-income settings. Specifically and importantly, it is intended for delivery by lay helpers without formal mental health training, representing a feasible psychological intervention for settings with few specialists.

It is hoped that ongoing randomized controlled trials will show that this simple intervention can provide effective care for adults with common mental health problems in communities exposed to adversity in LMICs.


## Acknowledgements

The United Nations High Commissioner for Refugees funded the conceptualization phase of the PM+. The opinions expressed in this paper are those of the authors and do not necessarily represent the decisions, policies, or views of the WHO.


## References


1. Ferrari AJ, Charlson FJ, Norman RE et al. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. PLoS Med 2013;10:e1001547.
2. World Health Organization. Task shifting: rational redistribution of tasks among health workforce teams: global recommendations and guidelines. Geneva: World Health Organization, 2008.
3. Bennett-Levy J, Richards D, Farrand P et al. Oxford guide to low intensity CBT interventions. Oxford: Oxford University Press, 2010.
4. Patel V, Chowdhary N, Rahman A et al. Improving access to psychological treatment: lessons from developing countries. Behav Res Ther 2011;49:523-8.
5. Collins PY, Patel V, Joestl SS et al. Grand challenges in global mental health. Nature 2011;475:27-30.
6. Clark DM. Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: the IAPT experience. Int Rev Psychiatry 2011;23:318-27.
7. 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;346:1-11.
8. van't Hof E, Stein DJ, Marks I et al. The effectiveness of problem solving therapy in deprived South African communities: results from a pilot study. BMC Psychiatry 2011;11:156-64.
9. Wiley-Exley E. Evaluations of community health care in low and middle income countries: a 10 year review of the literature. Soc Sci Med 2007;64:1231-41.
10. Bass J, Neugebauer R, Clougherty KF et al. Group interpersonal psychotherapy for depression in rural Uganda: 6-month outcomes: randomised controlled trial. Br J Psychiatry 2006;188:567-73.
11. Rahman A, Malik A, Sikander S et al. Cognitive Behaviour Therapy-based intervention by community health workers for mothers with depression and their infants in rural Pakistan: a cluster-randomised trial. Lancet 2008;372:902-9.
12. Bolton P, Lee C, Haroz EE et al. A transdiagnostic community-based mental health treatment for comorbid disorders: development and outcomes of a randomized controlled trial among Burmese refugees in Thailand. PLoS Med 2014;11:e1001757.
13. McEnvoy PM, Nathan P, Norton PJ. Efficacy of transdiagnostic treatments: a review of published outcome studies and future research directions. J Cogn Psychother 2009;23:20-33.
14. Wilamowska ZA, Thompson-Hollands J, Fairholme CP et al. Conceptual background, development, and preliminary data from the unified protocol for transdiagnostic treatment of emotional disorders. Depress Anxiety 2010;27:882-90.
15. Bullis JR, Fortune MR, Farchione TJ et al. A preliminary investigation of the long-term outcome of the Unified Protocol for Transdiagnostic Treatment of Emotional Disorders. Compr Psychiatry 2014;55:1920-7.
16. Murray LK, Dorsey S, Haroz EE et al. A common elements treatment approach for adult mental health problems in low- and middle-income countries. Cogn Behav Pract 2014;21:111-23.
17. Barlow DH, Allen LB, Choate ML. Toward a unified treatment for emotional disorders. Behav Ther 2004;35:205-30.
18. Bisson J, Andrew M. Psychological treatment of post-traumatic stress disorder (PTSD). Cochrane Database Syst Rev 2007;3:CD003388.
19. Dua T, Barbui C, Clark N et al. Evidence-based guidelines for mental, neurological, and substance use disorders in low- and

middle-income countries: summary of WHO recommendations. PLoS Med 2011;8:1-11.
20. Jorm AF, Morgan AJ, Hetrick SE. Relaxation for depression. Cochrane Database Syst Rev 2008;4:CD007142.
21. Tol WA, Barbui C, van Ommeren M. Management of acute stress symptoms, PTSD, and bereavement: recommendations from the WHO Guideline Development Group for Conditions Specifically Related to Stress. JAMA 2013;310:477-8.
22. D'Zurilla TJ, Goldfried MR. Problem solving and behavior modification. J Abnorm Psychol 1971;78:107-26.
23. Bowen D, Scogin F, Lyrene B. The efficacy of self-examination therapy and cognitive bibliotherapy in the treatment of mild to moderate depression. Psychother Res 1995;5:131-40.
24. Cuijpers P, van Straten A, Warmerdam L. Problem solving therapies for depression: a meta-analysis. Eur Psychiatry 2007;22:9-15.
25. Ferster CB. A functional analysis of depression. Am Psychol 1973;28:857-70.
26. Lewinsohn PM. A behavioral approach to depression. In: Freidman RJ, Katz M (eds). The psychology of depression: contemporary theory and research. Oxford: Wiley, 1974:157-78.
27. Cuijpers P, van Straten A, Warmerdam L. Behavioral activation treatments of depression: a meta-analysis. Clin Psychol Rev 2007; 27:318-26.
28. Brewin CR, Andrews B, Valentine JD. Meta-analysis of risk factors for posttraumatic stress disorder in trauma-exposed adults. J Consult Clin Psychol 2000;68:748-66.
29. Lowe SR, Rhodes JE. Trajectories of psychological distress among low-income, female survivors of Hurricane Katrina. Am J Orthopsychiatry 2013;83:398-412.
30. Schweitzer R, Melville F, Steel Z et al. Trauma, post-migration living difficulties, and social support as predictors of psychological adjustment in resettled Sudanese refugees. Aust N Zeal J Psychiatry 2006;40:179-88.
31. Neuner F, Onyut PL, Ertl V et al. Treatment of posttraumatic stress disorder by trained lay counselors in an African refugee settlement: a randomized controlled trial. J Consult Clin Psychol 2008;76:686-94.
32. Patel V, Weiss HA, Chowdhary N et al. Effectiveness of an intervention led by lay health counselors for depressive and anxiety disorders in primary care in Goa, India (MANAS): a cluster randomized controlled trial. Lancet 2010;376:2086-95.
33. Ali BS, Rahbar MH, Naeem S et al. The effectiveness of counselling on anxiety and depression by minimally trained counselors: a randomized controlled trial. Am J Psychother 2003;57:324-36.
34. Araya R, Rojas G, Fritsch R et al. Treating depression in primary care among low-income women in Santiago, Chile: a randomised controlled trial. Lancet 2003;361:995-1000.
35. Bolton P, Bass J, Neugebauer R et al. Group interpersonal psychotherapy for depression in rural Uganda: a randomized controlled trial. JAMA 2007;289:3117-24.
36. Bolton P, Bass J, Betancourt T et al. Interventions for depression symptoms amongst adolescent survivors of war and displacement in northern Uganda: a randomized controlled trial. JAMA 2007; 298:519-28.
37. Chibanda D, Mesu P, Kajawu L et al. Problem-solving therapy for depression and common mental disorders in Zimbabwe: piloting a task-shifting primary mental health care intervention in a population with a high prevalence of people living with HIV. BioMed Central Public Health 2011;11:1-10.
38. Chowdhary N, Jotheeswaran AT, Nadkarni A et al. The methods and outcomes of cultural adaptations of psychological treatments for depressive disorders: a systematic review. Psychol Med 2014; 44:1131-46.

DOI 10.1002/wps.20255

# Transition from child to adult mental health services: needs, barriers, experiences and new models of care

Swaran P. Singh, Helena Tuomainen

Division of Mental Health and Wellbeing, Warwick Medical School, University of Warwick, Warwick, UK

Transition from child to adult health care is a common experience for young people with enduring health problems who reach the age boundary between services. Transition is distinct from transfer (1), since it is more than a discrete administrative event. Good transition should be a co-ordinated, purposeful, planned and patient-centred process that ensures continuity of care, optimizes health, minimizes adverse events, and ensures that the young person attains his/her maximum potential. It starts with preparing a service user to leave a child-centred health care setting and ends when that person is received in, and properly engaged with, the adult provider (2).

In physical disorders, transition became a clinical and research priority as an increasing number of young people with previously life-threatening conditions survived into adulthood and needed ongoing care. Systematic and narrative reviews in cystic fibrosis (3), haemophilia (4), diabetes (5), congenital heart disease (6), cancer (7), cerebral palsy and spina bifida (8) and palliative care (9) have all identified transition as a risk period for disengagement and deterioration, but also a therapeutic opportunity for ensuring good outcomes into adult life.

Three broad categories of interventions have been tried: those aimed at the patient (educational programmes, skills training); those aimed at the staff (named transition co-ordinators, joint clinics run by paediatric and adult physicians); and changes in service delivery (separate young adult clinics, out of hours phone support, enhanced follow-up) (10). Yet the clinical and cost evaluation of such transition programmes is inconsistent and there are no robust and validated transition-related outcome measures (11).

## TRANSITION IN MENTAL HEALTH

Transition in mental health appears to be equally, if not more, problematic than in physical care settings. Seamless transition from child and adolescent mental health services (CAMHS) to adult mental health services (AMHS) is not the norm; instead young people with mental health problems frequently find themselves without professional support or a referral to an adult service (1,12). Alternatively, they may be referred, but the adult services are ill equipped to meet their needs (13).

Studies from the U.K. and U.S. show that mental health service use declines drastically when young people reach 16 years of age (by 24% and 45%, respectively), and even more so at the age of 18 (over 60% in the U.K.) (14,15). While young people with severe mental disorders such as psychosis are more likely to transition to adult services, those with neurodevelopmental, emotional/neurotic and personality disorders are far less likely to cross the boundary, and have more pronounced transition difficulties (15).

In the U.K., only about 15% of young people with attention deficit hyperactivity disorder (ADHD) make a transition (16); the figure for Ireland is 7% (17). In the U.S., there is the additional problem of lack of, or inconsistent, health insurance coverage for ADHD (18). Adult ADHD services are sparse or non-existent and many professionals are sceptical about the existence of ADHD in adulthood (19).

A particularly vulnerable group is represented by looked-after young people in the public care system, who are less likely to have family support but have significant mental health and social problems, including higher risk of self-harm and suicide, poorer educational achievement, and greater risk of unemployment, homelessness and incarceration (20). The labyrinthine service structures and interface means that the complex mental health needs of care-leavers remain unmet as they fall through the care gap (21) or disengage from services (22), increasing their use of crisis care (23) and ultimately leading to poor outcomes.

## MENTAL HEALTH NEEDS AND PREFERENCES OF YOUNG PEOPLE

Some key findings in recent years have changed our understanding of developmental psychopathology and age of onset of adult mental disorders. Large longitudinal epidemiological studies have confirmed continuity of childhood psychopathology into adult life, including both homotypic continuity (a disorder manifesting in the same manner across time) and sequential comorbidity (24). Our understanding of different developmental trajectories of the same disorder has improved; we know that juvenile onset disorders have poorer prognosis in adult life (25) and that sequential comorbidity may be due to a shared underlying diathesis (26).

The National Comorbidity Survey Replication from the U.S. has radically altered our understanding of the age of onset of different mental disorders (27). This large dataset allowed the authors to explore the prevalence of mental health problems, and also determine the age of onset for each recognized (DSM-IV) disorder. Overall, half of all

lifetime cases started by the age of 14, three quarters by the age of 24, with later onsets usually being comorbid conditions. The weight of evidence is such that adult mental health disorders are now being reframed as "extension of juvenile disorders" (28).

Studies across the developed world show that young people do not engage well with adult services (29). Young people may not be aware of what is available or refrain from seeking help because of stigma and unhelpful beliefs about autonomy (30). Their fluctuating clinical presentation with multiple comorbidities may not meet stringent criteria for stretched and struggling services. They may face a bewildering array of developmental and situational transitions that accompany health care transitions, such as changes in housing and relationships, gaining greater independence and moving on to adult roles (15,31).

Parents and young people find services particularly unhelpful during the transition period (30,32). Young people do not feel adequately prepared or supported during transition, lack understanding of adult services, feel insecure at the loss of the familiar and dread of the unfamiliar, and both young people and their families feel that their voices are not heard during the transition process (15,31,33). Abrupt and unplanned transition has been likened to "having to move house due to a flood" rather than a planned process determined by choice, appropriate advice and informed decision making (34). The current child-adult split in mental health services, therefore, creates weakness in the care pathway where it should be most robust (35) and is a major "design flaw" in current configuration (29).

## BARRIERS AT THE CAMHS-AMHS INTERFACE

Historically, child and adult psychiatric services have developed under very different societal needs and demands (36). In the U.S., child psychiatry dates its beginning to 1899, with the establishment of the first Juvenile Court in Chicago, when a group of influential and socially concerned women started campaigning for better understanding and management of juvenile delinquency (37). The influence of child psychoanalysts such as A. Freud, H. Hug-Hellmuth and M. Klein ensured that child psychiatry had its ideological and conceptual roots firmly in family, community and society rather than in a biological or diagnostic paradigm.

Over the subsequent decades, behavioural and educational psychologists, psychiatrists, criminologists, paediatricians, neurologists and social workers, often with starkly differing concepts about the causes and treatment of childhood mental disorders, contributed to the development of child psychiatry. Unlike adult psychiatry with its focus on individual psychopathology and diagnosis-led treatment, child psychiatry recognized early the wider influences of family and interpersonal processes in both the genesis and management of childhood mental disorders.

Over time, child and adolescent services have developed a culture, an organization and models of functioning very different from adult care, and these pre-existing differences get accentuated at the transition boundary (15). A range of obstacles hampers communication and collaboration at the CAMHS and AMHS interface (38,39). Separate funding and governance structures result in distinct systems with rigid boundaries and lack of understanding of services across the divide (40). Legal, logistic and clinical differences, combined with time and resources constraints, prevent services working together to provide parallel care, with particular concerns about where the responsibility of clinical care lies (31,36).

This lack of experience of working together contributes to limited understanding of what is needed, what is expected and the purpose of good transitional care (38). Some barriers relate to users and carers. Many young people and their families decline referral to adult services due to stigma and misperception. All these barriers contribute to a lack of referrals despite ongoing need for care, young people dropping through the care gap, and poor experience of care for those who make it to the other side (15,31).

## NEW MODELS OF CARE

Although barriers to good transition have been mapped, little has been tested to make transition better (41). Systematic reviews have identified a small number of interventions that facilitate transition, but the evidence is based on small, non-random, retrospective studies often with no comparison group (39).

A recent international Delphi study identified six essential elements of a successful transition programme: a) assuring a good coordination (such as timing of transfer, communication, follow-up, remaining available as a consultant, etc.) between child and adult professionals; b) starting planning transition at an early age (at least one year before the transfer boundary); c) discussing with patient and family about self-management; d) including young person's views and preferences in transition planning; e) if developmentally appropriate, seeing the adolescent alone at least for part of the consultation; and f) identifying an adult provider willing to take on the young patient before transfer (42).

In the looked-after population, transition support services that provide training and promote independence and self-sufficiency have been tried, but the evidence remains equivocal and the studies suffer from the same methodological limitations as identified in other reviews (43).

Identifying what is needed appears much easier than actually providing it. In current clinical practice, there is no consensus on who can be discharged on reaching the transitional boundary, who should receive transitional care, how this care should be delivered, what outcomes should be measured, what are the outcomes of those who fall through the care

gap, and what are the individual, organizational and societal costs of poor, inadequate or inappropriate transition.

Recent evidence confirming that treatment in the early stages of a disorder is likely to be both a clinically and a cost-effective strategy to reduce long-term disease burden has led to very strong arguments that the early intervention paradigm should be applied to all disorders of youth onset (44). And instead of fixing "the broken bridge" between two models of care, neither of which serves young people well, there should be a radical redesign with a seamless new pathway within a stigma free, youth friendly specialist model.

Several such models have sprung up in Australia, U.K., Ireland, Singapore and Denmark, with new ones proposed in Canada, U.S. and Israel (29). While some might argue that having a 0-25 service, as planned in Birmingham, U.K. (http://forwardthinkingbirmingham.org.uk) simply shifts the transition boundary to 25, the new pathway will be robust at the period of maximum risk both of discontinuity of care in early onset disorders and of the peak incidence of emerging mental disorders.

Meanwhile, the search for good transition models continues. MILESTONE is a European Union (EU)-funded transition project (www.milestone-transitionstudy.eu) that aims to delineate the child-adult interface, including policies, service structure and organization, and transition-related training in mental health care across Europe; identify a large (N=1000) prospective cohort of transition age youth in eight EU countries and track their journey across the transition boundary; robustly test the clinical and cost-effectiveness of a model of managed transition in improving health and social outcomes using a cluster-randomized design; and create training, commissioning and policy guidelines for improving transitional care across the EU.

## CONCLUSIONS

Young people receiving care from child mental health services are at high risk of falling through the child-adult service gap as they cross the transition boundary between services; or experience poor care, leading to high risk of disengagement from services and discontinuity of care. The transition boundary spans the maximum risk period for the emergence of serious mental disorders, hence focussing on transitional care has the potential for transforming outcomes in youth mental health.

We need to urgently develop and implement reformed service models that are specifically geared to meeting the unique needs of adolescents and young adults, are based on needs and preferences rather than strictly aligned to chronology and rigid diagnostic boundaries, and provide high quality evidence-based interventions that promote well-being, self-sufficiency, autonomy and fulfilment. Our young people deserve nothing less.

## References


1. Paul M, Ford T, Kramer T et al. Transfers and transitions between child and adult mental health services. Br J Psychiatry 2013; 202(Suppl. 54):s36-40.
2. Coleman EA, Berenson RA. Lost in transition: challenges and opportunities for improving the quality of transition of care. Ann Intern Med 2004;141:533-6.
3. Tuchman LK, Schwartz LA, Sawicki GS et al. Cystic fibrosis and transition to adult medical care. Pediatrics 2010;125:566-73.
4. Breakey VR, Blanchette VS, Bolton-Maggs PHB. Towards comprehensive care in transition for young people with haemophilia. Haemophilia 2010;16:848-57.
5. Nakhla M, Daneman D, Frank M et al. Translating transition: a critical review of the diabetes literature. J Pediatr Endocrinol Metab 2008;21:507-16.
6. Dearani JA, Connolly HM, Martinez R et al. Caring for adults with congenital cardiac disease: successes and challenges for 2007 and beyond. Cardiol Young 2007;17:87-96.
7. Freyer DR. Transition of care for young adult survivors of childhood and adolescent cancer: rationale and approaches. J Clin Oncol 2010;28:4810-8.
8. Binks JA, Barden WS, Burke TA et al. What do we really know about the transition to adult-centered health care? A focus on cerebral palsy and spina bifida. Arch Phys Med Rehabil 2007;88: 1064-73.
9. Doug M, Adi Y, Williams J et al. Transition to adult services for children and young people with palliative care needs: a systematic review. Arch Dis Child 2011;96:78-84.
10. Crowley R, Wolfe I, Lock K et al. Improving the transition between paediatric and adult healthcare: a systematic review. Arch Dis Child 2011;96:548-53.
11. Prior M, McManus M, White P et al. Measuring the "triple aim" in transition care: a systematic review. Pediatrics 2014;134:E1648-61.
12. Singh SP. Transition of care from child to adult mental healthservices: the great divide. Curr Opin Psychiatry 2009;22:386-90.
13. Dimitropoulos G, Tran AF, Agarwal P et al. Challenges in making the transition between pediatric and adult eating disorder programs: a qualitative study from the perspective of service providers. Eating Disorders 2013;21:1-15.
14. Pottick KJ, Bilder S, Vander Stoep A et al. US patterns of mental health service utilization for transition-age youth and young adults. J Behav Health Serv Res 2008;35:373-89.
15. Singh SP, Paul M, Ford T et al. Process, outcome and experience of transition from child to adult mental healthcare: multiperspective study. Br J Psychiatry 2010;197:305-12.
16. Ogundele M, Omenaka I. An audit of transitional care for adolescents with ADHD in a North West England district. Arch Dis Child 2012;97(Suppl. 1):A129.
17. McNicholas F, Adamson M, McNamara N et al. Who is in the transition gap? Transition from CAMHS to AMHS in the Republic of Ireland. Irish J Psychol Med 2015;32:61-9.
18. Montano CB, Young J. Discontinuity in the transition from pediatric to adult health care for patients with attention-deficit/hyperactivity disorder. Postgrad Med 2012;124:23-32.
19. Swift KD, Sayal K, Hollis C. ADHD and transitions to adult mental health services: a scoping review. Child: Care, Health and Development 2014;40:775-86.
20. Department for Children, Schools and Families/Office of National Statistics. Children looked after in England (including adoption and care) year ending 31 March 2009. London: Department for Children, Schools and Families, 2009.
21. Campbell F, O'Neill PM, While A et al. Interventions to improve transition of care for adolescents from paediatric services to adult services. Cochrane Database Syst Rev 2012;4:1-11.

22. Tripp J, Amiel S, Balen A et al. Bridging the Gaps: health care for adolescents. London: Royal College of Paediatrics and Child Health, 2003.
23. Meltzer H, Gatward R, Corbin T et al. The mental health of young people looked after by local authorities in England. London: Her Majesty's Stationery Office, 2003.
24. Hofstra MB, Van Der Ende J, Verhulst FC. Continuity and change of psychopathology from childhood into adulthood: a 14-year follow-up study. J Am Acad Child Adolesc Psychiatry 2000;39: 850-8.
25. Wickramaratne PJ, Greenwald S, Weissman MM. Psychiatric disorders in the relatives of probands with prepubertal-onset or adolescent-onset major depression. J Am Acad Child Adolesc Psychiatry 2000;39:1396-405.
26. Krueger RF. The structure of common mental disorders. Arch Gen Psychiatry 1999;56:921-6.
27. Kessler RC, Chiu WT, Demler O et al. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005;62: 617-27.
28. Kim-Cohen J, Caspi A, Moffitt TE et al. Prior juvenile diagnoses in adults with mental disorder: developmental follow-back of a prospective-longitudinal cohort. Arch Gen Psychiatry 2003;60: 709-17.
29. McGorry PD, Goldstone SD, Parker AG et al. Cultures for mental health care of young people: an Australian blueprint for reform. Lancet Psychiatry 2014;1:559-68.
30. Davis M, Butler M. Service system supports during the transition from adolescence to adulthood: parent perspectives. Alexandria: National Technical Assistance Center, 2002.
31. Hovish K, Weaver T, Islam Z et al. Transition experiences of mental health service users, parents, and professionals in the United Kingdom: a qualitative study. Psychiatr Rehabil J 2012;35: 251-7.
32. Jivanjee P, Kruzich J. Supports for young people with mental health conditions and their families in the transition years. Best Pract Ment Health 2011;7:115-33.
33. Fegran L, Hall EOC, Uhrenfeldt L et al. Adolescents' and young adults' transition experiences when transferring from paediatric to adult care: a qualitative metasynthesis. Int J Nurs Studies 2014; 51:123-35.
34. Wilson A, Tuffrey A, McKenzie C et al. After the flood: young people's perspectives on transition. Lancet Psychiatry 2015;2:376-8.
35. McGorry P, Bates T, Birchwood M. Designing youth mental health services for the 21st century: examples from Australia, Ireland and the UK. Br J Psychiatry 2013;202(Suppl. 54):s30-5.
36. Singh SP, Evans N, Sireling L et al. Mind the gap: the interface between child and adult mental health services. Psychiatr Bull 2005;29:292-4.
37. Schowalter JE. The history of child and adolescent psychiatry. Psychiatry 1994;2:1-13.
38. McLaren S, Belling R, Paul M et al. 'Talking a different language': an exploration of the influence of organizational cultures and working practices on transition from child to adult mental health services. BMC Health Serv Res 2013;13:254.
39. Embrett MG, Randall GE, Longo CJ et al. Effectiveness of health system services and programs for youth to adult transitions in mental health care: a systematic review of academic literature. J Ment Health Pol Econom 2015;18:S12-S3.
40. Davis M, Sondheimer DL. State child mental health efforts to support youth in transition to adulthood. J Behav Health Serv Res 2005;32:27-42.
41. Swift KD, Hall CL, Marimuttu V et al. Transition to adult mental health services for young people with Attention Deficit/Hyperactivity Disorder (ADHD): a qualitative analysis of their experiences. BMC Psychiatry 2013;13:74.
42. Suris JC, Rutishauser C, Akre C. Does talking about it make a difference? Opinions of chronically ill young adults after being transferred to adult care. Arch Pediatr 2015;22:267-71.
43. Everson-Hock ES, Jones R, Guillaume L et al. Supporting the transition of looked-after young people to independent living: a systematic review of interventions and adult outcomes. Child Care Health Develop 2011;37:767-79.
44. Birchwood M, Singh SP. Mental health services for young people: matching the service to the need. Br J Psychiatry 2013;202(Suppl. 54):s1-2.

DOI 10.1002/wps.20266

# Impact of the Germanwings plane crash on mental illness stigma: results from two population surveys in Germany before and after the incident

The Germanwings plane crash on March 24, 2015 and its wide international media coverage have prompted concerns about a possible setback in fighting mental illness stigma (1).

The influence of media coverage on mental health related attitudes has been repeatedly demonstrated (2,3). There is reason to expect that the presumed murder-suicide of the co-pilot, killing 150 persons and being linked to a diagnosis of depression, has increased perceptions of dangerousness, unpredictability, fear, anger and desire for social distance from persons with serious mental illness. In fact, a series of population surveys conducted in Germany in 1990 and 1991 before and after violent attacks on two politicians by persons with psychotic disorders demonstrated a considerable increase of stigma after the attacks. The proportion of the population being unwilling to sublet a room to a person with schizophrenia, for example, increased by 24% (4).

Using data from two consecutive representative online surveys in Germany before and after the plane crash, we examined whether and to what extent attitudes towards persons with mental illness did worsen after the incident in March 2015.

Two surveys were conducted among persons >15 years old from an established market research panel in Germany. The first survey in November 2014 was part of a survey experiment, from which we use a "no intervention" control group for the present analysis (N=598); the second survey in May 2015 was an identical replication of that condition (N=806). Quota sampling yielded two independent samples representative for the general population with respect to age, gender and region.

Respondents were randomly presented a case history of a woman, Anne, suffering from either depression or schizophrenia, without mentioning of the diagnosis (5). Afterwards, they answered questions on perceived dangerousness, blame, continuity beliefs, emotional reactions, support for structural discrimination, and desire for social distance. Responses were recorded on five-point Likert scales, which we combined into three categories: agree or likely, undecided, disagree or unlikely.

We then calculated multinomial logit regression models for all items, comparing the predicted probability for choosing each category between surveys. Analyses controlled for type of vignette, respondents' gender, age, and years of education. To establish significance, we computed 95% confidence intervals (CI) for the predicted difference between surveys with the delta method. We multiplied probabilities by 100, so they can be read as percentages endorsing each category. All analyses were conducted using STATA, version 13.

Two items showed significant differences between surveys. After the plane crash, 24% of respondents regarded Anne as unpredictable, compared to 17% before the incident (change in predicted probability: 7%, CI: 3 to 11). On the other hand, 22% compared to 27% endorsed the statement "To some extent, most people will experience problems that are similar to those of Anne" (−5%, CI: −10 to 0). Agreement to other items related to dangerousness increased by smaller and not significant amounts: "Anne is a danger to other people" by 3% (CI: 0 to 6); "Anne is a danger to herself" by 5% (CI: −1 to 11).

Emotional reactions like fear, anger or sympathy, support for restrictions like compulsory treatment or withdrawal of the driving license, and desire for social distance (move next door, spend an evening socializing, make friends, work closely on a job, marry into family) did not differ significantly between surveys (changes in predicted probability: −2% to 3%).

An analysis of interaction effects for type of disorder did not show significant interactions, suggesting that the observed changes were not illness specific.

These results suggest that the plane crash did have a measurable impact on public attitudes towards persons with mental disorder. Increased perceptions of unpredictability and reduced notions of similarity between a person with mental illness and most other persons seem to be related to the flight incident and the suspected murder-suicide of the co-pilot. However, considering the horrible facts that have been made public about the incident, its broad media coverage and its frequent explicit linking to a mental disorder of the co-pilot, the observed changes were surprisingly small. In particular, emotional reactions towards a person with mental illness did not change, and the desire for social distance did not increase. It seems as if the public has largely resisted the impulse of generalizing negative stereotypes and reactions to all persons suffering from mental illness.

Probably, the intensity of attitude changes would have been stronger if elicited with regard to a person resembling more closely the co-pilot, for example depicting a young male person or even a pilot of a passenger airplane with mental illness. Still, our data suggest that it might be premature to complain about a general "resurgence" of mental illness stigma after the plane crash (6).

A limitation of our study is its restriction to an online sample. Online samples are usually better educated than

the general population, and the online population was presumably exposed to even stronger media coverage of the plane crash. While we did control all analyses reported here for educational achievement, we did not record the amount and type of media consumption of our samples.

Combating public stigma of mental illness has been proven a difficult task (7). In contrast to the early 1990s (4), media reporting of a single, extremely disturbing incident seems not to have caused a large scale shift in public attitudes, suggesting that the public may have become more resistant to negative generalizations about mental illness.

***Georg Schomerus[1,2], Susanne Stolzenburg[1,2], Matthias C. Angermeyer[3,4]***

*[1]Department of Psychiatry, University of Greifswald, Greifswald, Germany; [2]HELIOS Hanseklinikum Stralsund, Stralsund, Germany; [3]Department of Public Health, University of Cagliari, Cagliari, Italy; [4]Center for Public Mental Health, Gösing am Wagram, Austria*

## References

1. Torjesen I. The pilot, depression, and the salacious headlines that feed stigma. BMJ 2015;350:h1874.
2. Angermeyer MC, Schulze B. Reinforcing stereotypes: how the focus on forensic cases in news reporting may influence public attitudes towards the mentally ill. Int J Law Psychiatry 2001;24: 469-86.
3. Pirkis J, Blood RW, Francis C et al. On-screen portrayals of mental illness: extent, nature, and impacts. J Health Commun 2006; 11:523-41.
4. Angermeyer MC, Matschinger H. Violent attacks on public figures by persons suffering from psychiatric disorders. Their effect on the social distance towards the mentally ill. Eur Arch Psychiatry Clin Neurosci 1995;245:159-64.
5. Schomerus G, Matschinger H, Angermeyer MC. Continuum beliefs and stigmatizing attitudes towards persons with schizophrenia, depression and alcohol dependence. Psychiatry Res 2013;209:665-9.
6. Schmidt L, Reif A. Talking to a psychiatrist: "Our patients are seen as crazy, not ill". Frankfurter Allgemeine Zeitung, June 10, 2015.
7. Angermeyer MC, Matschinger H, Schomerus G. Attitudes towards psychiatric treatment and people with mental illness: changes over two decades. Br J Psychiatry 2013;203:146-51.

DOI 10.1002/wps.20257

# Influence of early life characteristics on psychiatric admissions and impact of psychiatric disease on inflammatory biomarkers and survival: a Danish cohort study

Most psychiatric research has focused on the identification of etiologic and prognostic factors for specific psychiatric diagnostic categories, in particular depression and schizophrenia. A limited number of studies have examined risk factors such as composite measures of cognitive ability (IQ) or mortality across different psychiatric diseases (1-3). Furthermore, although mental disorders seem to have their roots early in life (4,5), few studies have explored the influence of early life characteristics on these disorders over the life course and potential underlying mechanisms such as systemic inflammation (6).

In a cohort of Danish men, we examined the impact of social, mental and physical characteristics assessed in childhood, young adulthood and midlife on the incidence of all psychiatric admissions and of schizophrenia, depression, alcohol and drug abuse. We further explored the influence of the above psychiatric diseases on inflammatory biomarkers and survival.

The information used was extracted from birth registers (birth weight and father's socioeconomic position); a school survey in 1965 (IQ); conscript examinations (IQ, education and body mass index); a health survey in 2004 (Major Depression Inventory, body mass index and smoking) (7); and a follow-up examination in 2010 (adult socioeconomic position, IQ, Major Depression Inventory, body mass index, smoking and inflammatory biomarkers) (8). Biomarkers included high-sensitivity C-reactive protein (hsCRP), interleukin (IL)-6 and IL-18, IL-10, tumor necrosis factor alpha (TNF-alpha) and interferon gamma (IFN-gamma).

Participants were recruited from the Metropolit cohort, consisting of 11,532 men born in 1953 in the Copenhagen Metropolitan area. Data from birth certificates were available for all members of the cohort. For 11,108 men, additional information was collected from conscript board examinations at about age 20 years. Further, 7,987 cohort members participated in a school-based survey in 1965 around age 12 years, and 6,292 were followed up by mailed questionnaire in 2004 around age 51 years (7). In 2010, around age 57 years, 7,799 cohort members living in the Eastern part of Denmark were invited to The Copenhagen Ageing Midlife Biobank, and 2,486 participated in a health examination including blood sampling and psychological tests (8).

Information on any admission to a psychiatric ward from 1972 to 2009 was obtained by linkage with the Danish Psychiatric Central Registry. All-cause mortality was followed from 1968 to 2009 by register linkage with the Danish civil registration system.

Data were analyzed using chi-square test, t-test, linear and Cox regression analysis in STATA version 12. Psychiatric admission was entered as time-dependent variable in the survival analysis.

Of the cohort members, 1,640 (14.2%) had ever been admitted to a psychiatric ward between age 19 and 56 years. The most frequent diagnosis (34.5%) was alcohol or drug abuse, while 18.2% were diagnosed with schizophrenia and 17.1% had an affective disorder.

Men with any psychiatric admission had lower socioeconomic position and lower IQ from childhood to middle age. They had lower mean birth weight and lower body mass index at age 20 and in middle age, and were more often smokers. In adjusted regression analysis, low IQ at age 20 increased the likelihood of developing schizophrenia (hazard ratio, HR per SD decrease = 1.39, 95% CI: 0.79-1.61), but decreased that of developing depression (HR per SD decrease = 0.81, 95% CI: 0.68-0.96). Birth weight and low education at age 20 were associated with alcohol or drug abuse (HR per 100 g increase = 0.98, 95% CI: 0.97-1.00; and HR low versus high = 1.51, 95% CI: 1.16-1.97, respectively).

Among the 2,486 men who participated in the health examination in 2010, the 242 men with a psychiatric admission had significantly higher levels of hsCRP, IL-6 and IL-18, while IL-10, TNF-alpha and IFN-gamma were not associated with psychiatric morbidity. The regression coefficients for having any psychiatric admission were $\beta = 0.37$ (95% CI: 0.16-0.57) for hsCRP; $\beta = 0.16$ (95% CI: 0.02-0.30) for IL-6; and $\beta = 0.09$ (95% CI: 0.01-0.18) for IL-18.

Analyses of the relation between psychiatric diagnosis and biomarkers showed that alcohol or drug abusers ($\beta = 0.79$, 95% CI: 0.42-1.17) and those with affective disorders ($\beta = 0.44$, 95% CI: 0.00-0.89) had higher levels of hsCRP. Men with an abuse diagnosis also had higher IL-6 levels ($\beta = 0.52$, 95% CI: 0.28-0.76).

During the follow-up period, 1,392 (12.7%) of the 11,532 cohort members died. Of them, 511 (37%) had a psychiatric diagnosis. Men with a psychiatric admission had higher mortality rates at age 55 years (HR = 5.43, 95% CI: 4.76-6.20), after adjustment for early life characteristics. The analyses also showed increased mortality in

all four psychiatric diagnostic categories. The highest HRs were observed for alcohol or drug abuse (8.23, 95% CI: 6.98-9.68) and schizophrenia (6.43, 95% CI: 5.20-8.12).

These findings suggest that low birth weight, socioeconomic position and IQ early in life increase the risk of psychiatric disease, in particular of alcohol or drug abuse, in adult men. Alcohol or drug abuse is strongly associated with inflammatory biomarkers and poor survival.

***Merete Osler[1], Egill Rostrup[2], Merete Nordentoft[3], Erik Lykke Mortensen[4], Helle Bruunsgaard[5], Birgitte Fagerlund[6]***

*[1]Research Center for Prevention and Health, Glostrup Hospital, Glostrup, Denmark; [2]Functional Imaging Unit, Department of Diagnostics, Glostrup Hospital, Glostrup, Denmark; [3]Mental Health Centre Copenhagen, University of Copenhagen, Copenhagen, Denmark; [4]Institute of Public Health and Center for Healthy Aging, University of Copenhagen, Copenhagen, Denmark; [5]Department of Clinical Immunology, Rigshospitalet, Copenhagen, Denmark; [6]Center for Neuropsychiatric Schizophrenia Research and Lundbeck Foundation Centre for Clinical Intervention and Neuropsychiatric Schizophrenia Research, Psychiatric Centre Glostrup, Glostrup, Denmark*

## Acknowledgements

The authors thank all those who initiated and/or continued the Metropolit study: K. Svalastoga, E. Høgh, P. Wolf, T. Rishøj, G. Strande-Sørensen, E. Manniche, B. Holten, I.A. Weibull and A. Ortman. They are also grateful to the staff at the Department of Public Health and the National Research Center for the Working Environment who undertook the data collection for the Copenhagen Aging and Midlife Biobank. Further thanks to H. Bruunsgaard, N.-E. Fiehn, Å.M. Hansen, P. Holm-Pedersen and R. Lund, who initiated and established the Copenhagen Aging and Midlife Biobank, together with K. Avlund, E.L. Mortensen and M. Osler. The study has been supported by a grant from the Velux Foundation.

## References

1. Mortensen EL, Sørensen HJ, Jensen HH et al. IQ and mental disorder in young men. Br J Psychiatry 2005;187:407-15.
2. Urfer-Parnas A, Mortensen EL, Saebye D et al. Pre-morbid IQ in mental disorders. A Danish draft-board study of 7486 psychiatric patients. Psychol Med 2010;40:547-56.
3. Nordentoft M, Wahlbeck K, Hällgren J et al. Excess mortality, cause of death and life expectancy in 270,770 patients with recent onset of mental disorders in Denmark, Finland and Sweden. PLoS One 2013;8:e55176.
4. Harris C, Barraclough B. Excess mortality of mental disorder. Br J Psychiatry 1998;173:11-53.
5. Papachristou E, Frangou S, Reichenberg A. Expanding conceptual frameworks: life course risk modelling for mental disorders. Psychiatry Res 2013;206:140-5.
6. Raedler TJ. Inflammatory mechanisms in major depressive disorder. Curr Opin Psychiatry 2011;24:519-25.
7. Osler M, Lund R, Christensen U et al. Cohort profile: the Metropolit 1953 Danish male birth cohort. Int J Epidemiol 2006;35: 541-6.
8. Avlund K, Osler M, Mortensen EL et al. Copenhagen Ageing and Midlife Biobank (CAMB). An introduction. J Ageing Health 2014; 26:5-20.

DOI 10.1002/wps.20258

# Treating major depressive episodes with antidepressants can induce or worsen metabolic syndrome: results of the METADAP cohort

Recent data (1-4) show a high comorbidity between major depressive disorder and metabolic syndrome (MetS) (5), a cluster of risk factors for cardiovascular diseases and type 2 diabetes including high waist circumference, high blood pressure, hypertriglyceridemia, low high-density lipoprotein (HDL) cholesterol, and high fasting plasma glucose.

In a context of increasing prescription of antidepressant medication (6) and evidence of weight gain induced by antidepressants (7), the impact of antidepressant treatment on MetS has to be clarified. Indeed, there has been no prospective study of reasonable sample size and duration addressing the incidence of MetS in patients with major depressive episode treated with antidepressants.

This question was addressed in the METADAP, a 6-month prospective, multicentric, real-world treatment observational cohort study of 624 patients with a diagnosis of major depressive disorder and a current major depressive episode. Data were collected from November 2009 to March 2013 in six university psychiatry departments in France.

Consecutive in- or out-patients, aged 18 to 65 years, with a current major depressive episode in a context of major depressive disorder (with a minimum score of 18 at the Hamilton Depression Rating Scale-17, HDRS-17) were assessed for MetS at the start of the index antidepressant treatment (M0), and one (M1), three (M3) and six (M6) months later. All of them provided their written informed consent.

Patients with psychotic symptoms, bipolar disorders, psychotic disorders, eating disorders, current substance abuse or dependence, pregnancy, organic brain syndromes or severe unstable medical conditions were not included. Patients receiving antipsychotics or mood stabilizers before inclusion and/or for 4 months or more during the last year were also excluded. Antipsychotics, mood stabilizers and stimulants were not permitted during the study, because of their metabolic effects. Benzodiazepines at the minimum effective dose and for the minimum time period and psychotherapies were allowed. The index antidepressant treatment had to be a monotherapy. The drug and its dose were left to the treating psychiatrist, using "real world" treatment options.

MetS was diagnosed according to the International Diabetes Federation definition (8). Participants had to have fasted and abstained from strenuous physical activity for 8 hours before examination. Triglycerides, HDL cholesterol and fasting plasma glucose levels were assessed using routine standardized laboratory methods. Thereafter, an assistant investigator blind to the major depression assessment measured waist circumference and blood pressure.

Mixed-effects multivariate models were used, because they are a well-accepted method for analyzing longitudinal clinical data in which missing or mistimed observations are present (9). All regression models included main effects for time since initiation of current antidepressant treatment, age, gender, HDRS-17 score at baseline, lifetime duration of prior major depressive disorder, lifetime duration of prior antidepressant medication, antidepressant-free period before inclusion, and current antidepressant classes.

Of 689 pre-included patients, 643 were included, of whom 19 had major deviations to the protocol. Thus, 624 patients were analyzed. Six had missing data for MetS at baseline.

Patients' mean age was 45.6±13.2 years; 68.7% were women, 87.5% were inpatients at baseline. Their mean HDRS-17 score at baseline was 24.7±5.0. Their mean number of previous major depressive episodes was 1.9±2.1. The average lifetime duration of major depressive disorder before inclusion was 11.5±12.2 years. The lifetime duration of antidepressant drug treatment before inclusion was 2.3±4.1 years.

Upon inclusion, 22.7% of patients were antidepressant naïve. The administered antidepressant was a selective serotonin reuptake inhibitor (SSRI) in 38.9% of cases, a serotonin norepinephrine reuptake inhibitor (SNRI) in 38.3%, a tricyclic antidepressant (TCA) in 8.8%, and another one in 14.0%. The mean duration of follow-up was 4.9±4.6 months. The drop-out rate was 25.9% before M1, 21.8% between M1 and M3, and 14.3% later. The main reasons for drop-out were antidepressant change (28.4%), prescription of antipsychotics or mood stabilizers (29.4%), and lost to follow-up (20.4%).

In patients without MetS at baseline (N=442, 70.8%), the incidence of MetS was 11.7% at M3 and 16.5% at M6. This increase was significant (mixed-effect multivariate logistic regression: OR=2.29, 95% CI: 1.69-3.10, $p<0.0001$). It was observed within both the SSRI (0% to 16.2%, $p<0.001$) and the SNRI group (0% to 16.1%, $p=0.001$). This increase was independent from other factors, such as age, lifetime duration of prior antidepressant medication, and presence of an antidepressant-free period at baseline.

The number of altered components of MetS significantly increased with time (M0: 1.2±0.9, M3: 1.3±1.1, M6: 1.5±1.2; mixed-model multivariate Poisson regression: incident risk ratio, IRR=1.06, 95% CI: 1.02-1.09, $p<0.0001$). It

was significantly higher in patients treated with SNRIs than in those treated with SSRIs (IRR=1.45, 95% CI: 1.16-1.80, p=0.001), and it was lower amongst patients who were antidepressant-free at baseline (IRR=0.81, 95% CI: 0.65-0.99, p=0.03). These effects were independent from each other, from age and gender.

In patients with MetS at baseline, mixed-effect multivariate linear regressions showed significant increases over time of supine blood pressure (M0: 123.2±16.4 mmHg, M3: 124.8±13.9 mmHg, M6: 126.8±15.0 mmHg, $p<0.05$) and fasting plasma glucose (M0: 0.98±0.29 g/l, M3: 1.07±0.48 g/l, M6: 1.03±0.31 g/l, $p<0.01$), which were independent from other factors.

The highlight of this study is the early and significant incidence of MetS after initiation of treatment with antidepressants. The majority of cases occurred in the first three months of treatment. A significant worsening of MetS was also observed in patients who already had the syndrome at baseline.

Taken together, these results suggest that treating major depressive episodes with antidepressants can induce or worsen MetS. Specific recommendations for the prevention of MetS in patients with major depressive disorder receiving antidepressant medication are needed. Further studies assessing the underlying mechanisms of this phenomenon are warranted.

***Emmanuelle Corruble[1-4], Khalil El Asmar[2], Severine Trabado[1,3-5], Céline Verstuyft[1,3,4,6], Bruno Falissard[1-3,7], Romain Colle[1-4], Anne-Cécile Petit[2], Florence Gressier[1-4], Sylvie Brailly-Tabard[1,3-5], Florian Ferreri[8,9], Jean-Pierre Lépine[10], Emmanuel Haffen[11], Mircea Polosan[12], Céline Bourrier[2], Gabriel Perlemuter[1,3,13,14], Philippe Chanson[1,3-5,15], Bruno Feve[9,16,17], Laurent Becquemont[1,3,4,6]***

*[1]University Paris-Sud, Le Kremlin Bicêtre, France; [2]Institut National de la Santé et de la Recherche Médicale UMR-1178, Le Kremlin Bicêtre, France; [3]Hôpitaux Universitaires Paris-Sud, Le Kremlin Bicêtre, France; [4]Hôpital de Bicêtre, Le Kremlin Bicêtre, France; [5]Institut National de la Santé et de la Recherche Médicale UMR-S1185, Le Kremlin Bicêtre, France; [6]Institut National de la Santé et de la Recherche Médicale U1184, Le Kremlin Bicêtre, France; [7]Hôpital Paul Brousse, Villejuif, France; [8]Université Pierre et Marie Curie, Paris, France; [9]Service de Psychiatrie, Hôpital Saint-Antoine, Paris, France; [10]Université Paris Diderot, UMR-S1144, Paris, France; [11]Université de Franche-Comté, Besançon, France; [12]Grenoble Institut de Neurosciences, Grenoble, France; [13]Institut National de la Santé et de la Recherche Médicale UMR-996, Clamart, France; [14]Hôpital Antoine-Béclère, Clamart, France; [15]Institut National de la Santé et de la Recherche Médicale U1185, Le Kremlin Bicêtre, France; [16]Institut Hospitalo-Universitaire ICAN, Paris, France; [17]Service d'Endocrinologie, Hôpital Saint-Antoine, Paris, France*

## Acknowledgement

This study was funded by the Programme Hospitalier de Recherche Clinique National of the French Ministry of Health (AOM06022).

## References

1. Pan A, Keum N, Okereke OI et al. Bidirectional association between depression and metabolic syndrome: a systematic review and meta-analysis of epidemiological studies. Diabetes Care 2012;35:1171-80.
2. Vancampfort D, Correll CU, Wampers M et al. Metabolic syndrome and metabolic abnormalities in patients with major depressive disorder: a meta-analysis of prevalences and moderating variables. Psychol Med 2013;21:1-12.
3. Rhee SJ, Kim EY, Kim SH et al. Subjective depressive symptoms and metabolic syndrome among the general population. Prog Neuropsychopharmacol Biol Psychiatry 2014;54:223-30.
4. Rethorst CD, Bernstein I, Trivedi MH. Inflammation, obesity, and metabolic syndrome in depression: analysis of the 2009-2010 National Health and Nutrition Examination Survey (NHANES). J Clin Psychiatry 2014;75:e1428-32.
5. Eckel RH, Grundy SM, Zimmet PZ. The metabolic syndrome. Lancet 2005;365:1415-28.
6. Mojtabai R, Olfson M. National trends in long-term use of antidepressant medications: results from the U.S. National Health and Nutrition Examination Survey. J Clin Psychiatry 2014;75:169-77.
7. Blumenthal SR, Castro VM, Clements CC et al. An electronic health records study of long-term weight gain following antidepressant use. JAMA Psychiatry 2014;71:889-96.
8. Alberti KG, Zimmet P, Shaw J et al. The metabolic syndrome – a new worldwide definition. Lancet 2005;366:1059-62.
9. Fitzmaurice GM, Laird NM, Shneyer L. An alternative parameterization of the general linear mixture model for longitudinal data with non-ignorable drop-outs. Stat Med 2001;20:1009-21.

DOI 10.1002/wps.20260

# Are we studying the right populations to understand suicide?

Suicide ranks 13th among leading causes of years of life lost, with more than 800,000 deaths worldwide annually (1). Particularly alarming is that 5.3% of deaths among those 15-49 years old are self-inflicted (1). The highest suicide rates are reported in Eastern Europe and East Asian countries, while the lowest are reported in Latin American and Muslim countries (2), and suicide rates may vary sharply across neighboring countries. Although effective suicide prevention policies exist, they may need to be adjusted for these large geographic differences in suicidal behavior, possibly related to culture, but largely still unexplained (3).

Given the socio-economic and personal impact of suicide, one might expect countries with high suicide rates to invest in suicide research and prevention. This has been the case in the Scandinavian countries, which have made large investments to understand and prevent suicidal behavior and have reduced their suicide rates (4). Of note, Swedish suicide prevention policy plans were relatively less strongly implemented than Danish or Finnish ones, and suicide prevention among Swedish males with mental disorders failed (4). Certainly, causality cannot be inferred from temporal association, but the data are intriguing.

Nevertheless, the worldwide distribution of research investment in suicide shows little correspondence with actual suicide rates. As in most domains, suicide research productivity is clustered in North America and Europe. In fact, of 19,440 published articles recorded in the Web of Science with the keyword "suicide" during 2010-2014, 5802 (37.3%) were from U.S. institutions and 6944 (44.6%) from European Union institutions. In contrast, 6.0% of recent suicide studies are from India and China, although these countries account for more than one third of the world population and almost half of world's suicides (5,6).

An analogous pattern is found when examining European Union and U.S. suicide research in more detail. Almost half of the scientific production regarding suicide in the European Union is from countries with low base suicide rates (<10 suicides per 100,000), such as Italy, the Netherlands, Spain and the UK, that represent about one third of the total European Union population. Similarly, the geographic distribution of suicide researchers within the U.S. does not follow suicide rates at the population level. Most research is carried out in institutions of the East and West coast (notably the Northeast), while the states in the West, where suicide rates are highest, produce far less suicide research.

In the same vein, most suicide studies are carried out with urban samples, but the highest suicide rates are usually found in rural areas (7). An inverse example of this relationship is the effect of urbanization in China, which seems to explain the declining rates of suicides along the last decade (8). Although the vast majority of suicides still occur in rural areas of low- and middle-income countries (5,6), the theoretical models of suicide, the recommended preventive interventions and the evidence about their effectiveness almost all come from urban institutions in high-income countries. The appropriateness of these models and interventions for low- and middle-income countries is uncertain.

Thus, most suicide research seems to be conducted in areas where suicide risk is lowest. As in any other field, regional differences in scientific output are correlated with research budgets and the size of the country's economy. However, because current suicide research is focused on low-risk populations, our capacity to build generalizable predictive and preventive models may be hindered. The limitations of suicide studies focused on a specific community can be illustrated by several facts.

First, the effect of life events on suicide risk is influenced by environmental or cultural factors. Losing a close relative or having financial problems seem to prompt different consequences for suicide risk depending on social networks, cultural reactions and even the economic climate in each country (2). This can be readily observed in the variability of suicide rates over time in different countries. For instance, South Korea has seen a dramatic increase in suicide rates, which have tripled (from <10 to around 30 suicides per 100,000) since the nineties (2,9). This escalation occurred in the context of economic growth, with country-specific factors – notably the unequal distribution of wealth affecting the elderly, the sensationalist media coverage of suicides, and the low rates of antidepressant treatment – appearing to play an important role in stoking the rise.

Second, the heritability of a broad suicidal phenotype including ideation, plans and attempts has been estimated to range from 30 to 50% (10). This variability likely reflects environmental effects, posited to modulate genetic predisposition to suicidal and other behaviors, but usually studied at the level of the individual's exposure to adversities in the environment (e.g., early childhood adversity) as opposed to more general environmental effects. Most of the growing literature on gene-environment interactions in suicidal behaviors focuses on individual life experiences in a particular community. However, the influence of social climate cannot be accurately measured if we do not compare distinct environments. For example, does corporal punishment have a different effect on children raised where it is culturally accepted compared to children raised where it is prohibited? Indeed, the effect of socio-cultural contexts on putative suicide risk factors, such as ethnicity or unemployment, may depend on ethnic density or employment rates, respectively (11). Moreover, risk factors for suicide may differ in high- and low-income countries (5),

but relevant site-specific findings may be disregarded because they are not disseminated in international scientific networks.

Third, the complexity of suicidal behavior is unlikely to be reflected in just a few variables, and studies combining factors in different dimensions to predict suicidal behavior have obtained discouraging results (12). Thus, the development of a robust model of suicidal behavior may require studies that include large samples and high-risk populations, most probably affected by gene-environment interactions. Unfortunately, multicentric studies including urban and rural areas are frequently hampered by unreliable data sources, disparate definitions of cause of death and, probably connected with social taboos and stigma, an underestimation of suicide deaths in many countries (13). In fact, one of the few cross-national studies on suicidal behavior, the World Health Organization's multisite intervention SUPRE-MISS, suggests that site-specific approaches to suicide prevention are needed given differences in prevalence of suicidal ideation and attempts (14).

In sum, if suicide research is only conducted in low-risk areas, the translation of these efforts into a global model of suicide behavior might prove problematic. International collaborations to boost suicide research are already under way, but so far they have been hindered by the use of divergent methodologies for the assessment of suicidal behavior. Collaborative approaches, consensual definitions and international expertise could foster suicide research and facilitate investigations in high-risk countries lacking resources and know-how.

***Jorge Lopez-Castroman[1], Hilario Blasco-Fontecilla[2], Philippe Courtet[1], Enrique Baca-Garcia[3,4], Maria A. Oquendo[4]***

*[1]Centre Hospitalier Universitaire Régional Montpellier and University of Montpellier, Montpellier, France; [2]Department of Psychiatry, IDIPHIM-Puerta de Hierro University Hospital, CIBERSAM, Madrid, Spain; [3]Department of Psychiatry, IIS-Jimenez Diaz Foundation, CIBERSAM, Madrid, Spain; [4]Department of Psychiatry, Columbia University/New York State Psychiatric Institute, New York, NY, USA*

## References

1. Lozano R, Naghavi M, Foreman K et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012;380:2095-128.
2. Blasco-Fontecilla H, Perez-Rodriguez MM, García-Nieto R et al. Worldwide impact of economic cycles on suicide trends over 3 decades: differences according to level of development. A mixed effect model study. BMJ Open 2012;2:e000785.
3. While D, Bickley H, Roscoe A et al. Implementation of mental health service recommendations in England and Wales and suicide rates, 1997-2006: a cross-sectional and before-and-after observational study. Lancet 2012;379:1005-12.
4. Wahlbeck K, Westman J, Nordentoft M et al. Outcomes of Nordic mental health systems: life expectancy of patients with mental disorders. Br J Psychiatry 2011;199:453-8.
5. Phillips MR, Cheng HG. The changing global face of suicide. Lancet 2012;379:2318-9.
6. Bertolote JM, Fleischmann A. Suicide and psychiatric diagnosis: a worldwide perspective. World Psychiatry 2002;1:181-5.
7. Fontanella CA, Hiance-Steelesmith DL, Phillips GS et al. Widening rural-urban disparities in youth suicides, United States, 1996-2010. JAMA Pediatr 2015;169:466-73.
8. Wang CW, Chan CL, Yip PS. Suicide rates in China from 2002 to 2011: an update. Soc Psychiatry Psychiatr Epidemiol 2014;49:929-41.
9. Hong J, Knapp M. Impact of macro-level socio-economic factors on rising suicide rates in South Korea: panel-data analysis in East Asia. J Ment Health Policy Econ 2014;17:151-62.
10. Mann JJ, Arango VA, Avenevoli S et al. Candidate endophenotypes for genetic studies of suicidal behavior. Biol Psychiatry 2009;65:556-63.
11. Neeleman J, Wilson-Jones C, Wessely S. Ethnic density and deliberate self harm; a small area study in south east London. J Epidemiol Community Health 2001;55:85-90.
12. Lee B-H, Kim Y-K. Potential peripheral biological predictors of suicidal behavior in major depressive disorder. Prog Neuropsychopharmacol Biol Psychiatry 2011;35:842-7.
13. Reynders A, Scheerder G, Van Audenhove C. The reliability of suicide rates: an analysis of railway suicides from two sources in fifteen European countries. J Affect Disord 2011;131:120-7.
14. Bertolote JM, Fleischmann A, De Leo D et al. Suicide attempts, plans, and ideation in culturally diverse sites: the WHO SUPRE-MISS community survey. Psychol Med 2005;35:1457-65.

DOI 10.1002/wps.20261

# International Society for Nutritional Psychiatry Research consensus position statement: nutritional medicine in modern psychiatry

In recent years, there has been an unprecedented growth in both the quantity and methodological quality of research directed at exploring the relationship between nutrition and mental health. Indeed, the strength of data has now afforded nutritional medicine a place in the mainstream psychiatric discourse (1).

Robust associations have been established between nutritional quality and mental health, with the bulk of this evidence indicating a protective effect of healthy diets on depressed mood (2), and the newest research supporting a detrimental impact of unhealthy diets on the mental health of young people (3,4) and adults (5,7).

There are also convincing data supporting the application of certain nutrient-based supplements (nutraceuticals) as monotherapy or combined therapy (8), or as augmentation therapy (9).

Although the growth in scientific research related to nutrition in psychiatry may be recent, it is now at a stage where it can no longer be ignored. In light of this, we aim to provide a platform to move towards a new integrated paradigm in psychiatry whereby nutritional considerations (both educational and prescriptive) can be considered "mainstream" (1). To this end, we present a consensus position statement from the International Society for Nutritional Psychiatry Research (ISNPR).

In brief, the ISNPR was formed in 2013 with the aim to advance research and communication on nutritional medicine in the field of psychiatry. One of its first goals was to formulate a position statement that embodied the principles of the organization, allowing for codification of the society's underpinning tenets.

In order to develop this, we employed a Delphi-based model by which ISNPR researcher and clinician members could vote on a select list of 110 statements created by an expert steering committee.

The committee provided a list of sub-statements concerning three main topics/areas: the current general needs and challenges in psychiatry; key elements of diet and nutraceutical evidence related to mental health/psychiatry; potential public health and clinical applications. These were transcribed and tabulated in Survey Monkey for online voting by the wider ISNPR membership.

A Likert scale (0–10) was used for each statement (0=don't include, 5=don't know/depends, 10=definitely include), and statements that received a mean score of >6.5/10 by ISNPR members were reviewed by the steering committee for inclusion in the position statement, which is presented below.

*Present treatment of mental disorders is achieving sub-optimal outcomes; in addition little attention is given to preventative efforts. Due to the immense burden of mental disorders, there is now an urgent need to identify modifiable targets to reduce the incidence of these disorders. Diet and nutrition offer key modifiable targets for the prevention of mental disorders and have a fundamental role in the promotion of mental health.*

*Epidemiological data, basic science, and clinical evidence suggest that diet influences both the risk for and outcomes of mental disorders. As such, we advocate that evidence-based nutritional change should be regarded as an efficacious and cost-effective means to improve mental health.*

*In addition to dietary modification, we recognize that nutrient-based (nutraceutical) prescription has the potential to assist in the management of mental disorders at the individual and population level. Many of these nutrients have a clear link to brain health, including: omega-3s, B vitamins (particularly folate and B12), choline, iron, zinc, magnesium, S-adenosyl methionine (SAMe), vitamin D, and amino acids. While we advocate for these to be consumed in the diet where possible, additional select prescription of these as nutraceuticals may also be justified.*

*Ongoing research (including randomized controlled trials) in the area is recognized as critical, using methodologically rigorous designs. Further explication of the biological pathways affected by nutritional modification is also required. Clinical trials of nutraceuticals should include assessment of biomarkers in tandem with clinical outcomes. Global research and health promotion activities focused on improving population health should also include mental health parameters as priority targets and measured outcomes.*

*Importantly, the activities of the food industry need to be examined at a governmental level and relevant policies designed to reduce the global burden of physical and mental ill-health attributable to poor diet. Such policies are advised to stimulate significant public change in dietary habits back towards a traditional wholefood diet (dependent on the culture). Further, there is now a vital need for better public and clinician education to communicate current research findings from the field.*

*In summary, nutrition and nutraceuticals should now be considered as mainstream elements of psychiatric practice, with research, education, policy, and health promotion reflecting this new paradigm.*

As detailed in our consensus statement, we advocate for the pursuit of an integrative psychiatric model, with diet as a key element. Further, the select use of evidence-based nutraceuticals should be a mainstay of treatment as either standalone therapies (mainly in cases of less severe mental disorders, non-tolerance to medication, nutrient deficiencies, or patient choice), or as adjunctive interventions with psychotropic medications to augment treatment efficacy. We recognize the importance of clinician and public education regarding evidence-based nutrition and nutraceuticals to drive mainstream acknowledgement of their impact on mental health.

It is the intention that this position statement and the ongoing work of ISNPR will assist in facilitating a transformation in psychiatry to better address the substantial global burden of mental illness, recognizing and embracing diet and nutrition as central determinants of both physical and mental health.

***Jerome Sarris[1,2], Alan C. Logan[3], Tasnime N. Akbaraly[4,5], G. Paul Amminger[6], Vicent Balanzá-Martínez[7], Marlene P. Freeman[8], Joseph Hibbeln[9], Yutaka Matsuoka[10], David Mischoulon[11], Tetsuya Mizoue[12], Akiko Nanri[12], Daisuke Nishi[13], Natalie Parletta[14], Drew Ramsey[15], Julia J. Rucklidge[16], Almudena Sanchez-Villegas[17,18], Andrew Scholey[2], Kuan-Pin Su[19,20], Felice N. Jacka[21-24]***

*[1]Department of Psychiatry, University of Melbourne, Melbourne Clinic, Melbourne, Australia; [2]Centre for Human Psychopharmacology, Swinburne University of Technology, Hawthorn, Australia; [3]CAMNR, Calabasas, CA, USA; [4]Inserm U1198, Montpellier, France; EPHE, Paris, France; University Montpellier 2, Montpellier, France; [5]Department of Epidemiology and Public Health, University College London, London, UK; [6]Orygen, The National Centre of Excellence in Youth Mental Health, Australia; [7]Teaching Unit of Psychiatry and Psychological Medicine, La Fe University and Polytechnic Hospital, University of Valencia Medical School, CIBERSAM, Valencia, Spain; [8]Perinatal and Reproductive Psychiatry, Massachusetts General Hospital, Boston, MA, USA; [9]National Institute on Alcohol Abuse and Alcoholism, NIH, Bethesda, MD, USA; [10]Department of Clinical Epidemiology, Translational Medical Center, National Center of Neurology and Psychiatry, Tokyo, Japan; [11]Depression Clinical and Research Program, Massachusetts General Hospital, Boston, MA, USA; [12]Department of Epidemiology and Prevention, National Center for Global Health and Medicine, Tokyo, Japan; [13]Department of Mental Health Policy and Evaluation, National Institute of Mental Health, Japan; [14]School of Population Health, University of South Australia, Adelaide, Australia; [15]Columbia University College of Physicians and Surgeons, New York, NY, USA; [16]Department of Psychology, University of Canterbury, Christchurch, New Zealand; [17]Research Institute of Biomedical and Health Sciences, University of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain; [18]CIBEROBN, Instituto de Salud Carlos III, Madrid, Spain; [19]Department of Psychiatry and Mind-Body Interface Laboratory, China Medical University Hospital, Taichung, Taiwan; [20]Graduate Institute of Neural and Cognitive Sciences, China Medical University, Taichung, Taiwan; [21]IMPACT SRC, School of Medicine, Deakin University, Geelong, 3220, Australia; [22]Murdoch Children's Research Centre, Parkville, Australia; [23]Department of Psychiatry, University of Melbourne, Royal Melbourne Hospital, Melbourne, Australia; [24]Black Dog Institute, Road Prince of Wales Hospital, Randwick, Australia*

## References

1. Sarris J, Logan A, Akbaraly T et al. Nutritional medicine as mainstream in psychiatry. Lancet Psychiatry 2015;2:271-4.
2. Lai JS, Hiles S, Bisquera A et al. A systematic review and meta-analysis of dietary pattern and depression in community-dwelling adults. Am J Clin Nutr 2013;99:181-97.
3. O'Neil A, Quirk SE, Housden S et al. Relationship between diet and mental health in children and adolescents: a systematic review. Am J Public Health 2014;104:e31-42.
4. Jacka FN, Ystrom E, Brantsaeter AL et al. Maternal and early postnatal nutrition and mental health of offspring by age 5 years: a prospective cohort study. J Am Acad Child Adolesc Psychiatry 2013;52:1038-47.
5. Akbaraly TN, Brunner EJ, Ferrie JE et al. Dietary pattern and depressive symptoms in middle age. Br J Psychiatry 2009;195: 408-13.
6. Sanchez-Villegas A, Delgado-Rodríguez M, Alonso A et al. Association of the Mediterranean dietary pattern with the incidence of depression: the Seguimiento Universidad de Navarra/University of Navarra follow-up (SUN) cohort. Arch Gen Psychiatry 2009; 66:1090-8.
7. Jacka FN, Pasco JA, Mykletun A et al. Association of Western and traditional diets with depression and anxiety in women. Am J Psychiatry 2010;167:305-11.
8. Rucklidge JJ, Kaplan BJ. Broad-spectrum micronutrient formulas for the treatment of psychiatric symptoms: a systematic review. Expert Rev Neurother 2013;13:49-73.
9. Sarris J, Kavanagh D, Byrne G. Adjuvant use of nutritional and herbal medicines with antidepressants, mood stabilizers and benzodiazepines. J Psychiatr Res 2010;44:32-41.

DOI 10.1002/wps.20223

# Becoming a psychiatrist in Europe: the title is recognized across the European Union, but what are the differences in training, salary and working hours?

The professional qualification as a psychiatrist is automatically recognized across Europe if a national training program fulfils the minimum requirement of four years of training duration (1,2). This is applicable to all 28 European Union (EU) member states, as well as to other countries of the European Economic Area, such as Norway and Switzerland (2). However, what is equivalent on paper may be not in practice: patients and hospital staff increasingly encounter doctors with different educational backgrounds due to the open European labour market and the mobility of trainees and psychiatrists.

In 2014/2015, the European Federation of Psychiatric Trainees surveyed training in psychiatry by a questionnaire directed at representatives of national psychiatric trainee associations. Except Cyprus, Czech Republic, Latvia, Luxembourg and Spain, all EU countries were covered.

A medical practitioner who undergoes postgraduate training in psychiatry and qualifies as a specialist is called a psychiatrist. Only in the German speaking countries, i.e. Austria, Germany, Switzerland and Liechtenstein, the specialist holds the title of "psychiatrist and psychotherapist" (2), even though training in psychotherapy is a mandatory part of psychiatric training in most European countries (3).

Although skills in psychotherapy are widely considered essential for psychiatrists, the number of patients to whom trainees are required to deliver psychotherapy varies and can be as little as zero, as in Estonia (4). In some countries, e.g. the Netherlands, child and adolescent psychiatry is a subspecialty of "adult" psychiatry. In other countries (25 out of 31) it is a separate specialty with up to 600 trainees (as reported from UK). However, only in Belgium the title specifies that a psychiatrist is specialized in adults ("psychiatrie de l'adulte") (2). A title such as "general psychiatrist" (awarded in the UK) could help differentiate subspecialties and underline the balance of technical and non-technical elements of care (5).

In order to match the EU minimum requirement, training duration needs to be four years or longer. The maximum required training durations are seven years in Ireland and six years in Austria, Finland, Switzerland and the UK. Training is not nationally standardized in four out of 31 countries (Belgium, Finland, France and Greece), underlining the challenge of establishing a single, unified European exam. In some countries it is required to rotate in a university hospital (six months in France) or a psychiatry ward in a general hospital (six months in Greece), or to spend twelve months in another hospital (Switzerland), pushing trainees to switch workplace. Outpatient care is strongly enforced in Finland, where half of the training has to take place in outpatient care. Overall, national curricula are still mainly defined by total duration and duration of rotations in (sub)specialties, despite the benefits of competency-based training (which may also facilitate a pan-European exam).

Appropriate working conditions, including salary and working hours, are essential for high-quality clinical training. Trainees in EU countries work 35 (Bulgaria) to 65 hours (Malta, including on-call hours) per week. Non-EU countries are characterized by less working hours: 35 hours per week in Belarus, Russia, Serbia and Ukraine. Income varies from 90€ per month in Ukraine to >4,000€ in Switzerland, i.e. in some countries trainees earn 44 times more than in others. The top-five countries in terms of average monthly salaries, mostly including on-call hours, after tax deduction are Switzerland, Sweden and the UK (≥4,000€), Norway (3,400€) and Germany (2,900€), while the lowest monthly salaries are paid in Ukraine (90€), Bulgaria (140€), Belarus (150€), Russia (150-500€) and Romania (400€). In Portugal (1,200€) it is common for trainees to spend a period abroad, during which they continue to be paid by their institution. Trainees in Belgium are paid (1,900-2,400€ per month) by their supervisors, which may cause conflicts of interest. Notably, not all aspects of training (especially parts of the psychotherapy curricula) are free of charge for trainees (4), further reducing their spendable income.

In most countries (17 out of 31), too few medical practitioners choose psychiatry as their specialty, yet initiatives to increase recruitment are lacking. As a consequence, in 16 countries, not all vacant positions are being filled, and only in very few countries (e.g., Greece) demand for training positions exceeds openings.

Thus, the characteristics of psychiatric training vary widely across Europe, despite an open labour market where specialists frequently work in foreign countries. The fact that the qualification of psychiatrists is equivalent throughout Europe should stimulate international cooperation when re-designing training curricula. Guidance and support by international organizations such as the European Federation of Psychiatric Trainees (6), the European Psychiatric Association (7), the European Union of Medical Specialists (8), the World Health Organization and the

WPA are crucial in order to facilitate harmonization of curricula. To improve local implementation, an international system of training programme inspections should be established.

***Franziska Baessler[1], Florian Riese[2], Mariana Pinto da Costa[3], Liwia de Picker[4], Olga Kazakova[5], Athanasios Kanellopoulos[6], Roland Grassl[7], Thomas Gargot[8], European Federation of Psychiatric Trainees, Marisa Casanova Dias[9]***

*[1]Faculty of Medicine, University of Berne, Berne, Switzerland; [2]Division of Psychiatry Research and Psychogeriatric Medicine, Psychiatric University Hospital Zurich, Zurich, Switzerland; [3]Hospital de Magalhães Lemos, University of Porto, Porto, Portugal; [4]Collaborative Antwerp Psychiatric Research Institute, University of Antwerp, Antwerp, Belgium; [5]Psychiatric Department, Psychiatric Clinic of Minsk City, Minsk, Belarus; [6]Evgenidion Therapeftirion, National and Kapodistrian University of Athens, Athens, Greece; [7]University Clinic for Child and Adolescent Psychiatry and Psychotherapy, Vienna, Austria; [8]Service de Pédopsychiatrie, Hôpital de la Pitié Salpêtrière, Paris, France; [9]Barnet, Enfield and Haringey NHS Trust, London, UK*

## References

1. European Union. Directive 2013/55/EU of the European Parliament and of the Council of 20 November 2013 amending Directive 2005/36/EC on the recognition of professional qualifications and Regulation (EU) No. 1024/2012 on administrative cooperation through the Internal Market Information System ('the IMI Regulation'). Official Journal of the European Union, L 354, December 28, 2013.
2. European Union. Directive 2005/36/EC of the European Parliament and of the Council of 7 September 2005 on the recognition of professional qualifications. Official Journal of the European Union, L 255/25, September 30, 2005.
3. Lotz-Rambaldi W, Schäfer I, ten Doesschate R et al. Specialist training in psychiatry in Europe – Results of the UEMS-survey. Eur Psychiatry 2008;23:157-68.
4. Fiorillo A, Luciano M, Giacco D et al. Training and practice of psychotherapy in Europe: results of a survey. World Psychiatry 2011;10:238.
5. Maj M. Technical and non-technical aspects of psychiatric care: the need for a balanced view. World Psychiatry 2014;13:209-10.
6. European Federation of Psychiatric Trainees. EFPT recommendations on standards of psychiatric training. European Federation of Psychiatric Trainees, 2013.
7. Mayer S, van der Gaag RJ, Dom G et al. European Psychiatric Association (EPA) guidance on post-graduate psychiatric training in Europe. Eur Psychiatry 2014;29:101-16.
8. European Union of Medical Specialists. European framework for competencies in psychiatry. European Union of Medical Specialists, 2009.

DOI 10.1002/wps.20259

# WPA Secretariat: the global link to Member Societies

**Roy Abraham Kallivayalil**
WPA Secretary General

The WPA Secretariat is located at the Geneva University Psychiatric Hospital, in splendid surroundings, in a campus full of beautiful trees and open spaces with lush greenery. It is the headquarters of the WPA.

Ever since the establishment of the Association, the WPA Secretariat used to move with the incumbent office bearers. As the size of our office grew, the need for a permanent secretariat became increasingly evident. Besides administering the WPA, we needed a place to keep our valuable documents in safe custody.

There were several reasons for which Geneva was chosen as the location of the Secretariat. It was the town where the headquarters of the World Health Organization (WHO) were situated, and working in physical proximity and in collaboration with the WHO on mental health matters was an obvious advantage. But the most important factor was the offer from the Geneva University Hospital to provide free space and facilities to WPA for its Secretariat in their campus. In many ways, this can be considered as a gift to the psychiatrists of the world by that university. The permanent secretariat of WPA at Geneva came into being when an "accord of collaboration" between the WPA and the Geneva University Hospital was signed by M.B. Gruson (Director General of the hospital) and A. Okasha (WPA President) on September 6, 2004. This accord is valid for an initial period of 20 years and is subject to renewal thereafter.

The WPA Secretariat provides the global link to 135 Member Societies and more than 200,000 members. Member Societies are the most important components of the WPA, and the Secretariat keeps them informed about the discussions and decisions of the Executive Committee, other Committees and the WPA General Assembly. Similarly, the Secretary General relates the opinions and concerns of the Member Societies to these bodies and conveys decisions and actions back to them.

The Secretariat makes every effort to keep in constant touch with the Member Societies and provide information sought by them. It co-ordinates the admission process of Member Societies, Affiliated Associations and individual affiliated members. It gives logistical support to the WPA President and its leadership and makes arrangements for its business meetings.

The WPA Secretary General is in charge of the WPA Secretariat and is responsible for the administrative tasks of the WPA. We have 18 Zonal Representatives who constitute the WPA Board. The Board advises the Executive Committee and the General Assembly on the work of the WPA and helps to strengthen collaboration between Member Societies and in the implementation of the WPA Action Plan. Through the Secretariat, the Secretary General co-ordinates the work of the Zonal Representatives and serves as a liaison between them and the WPA governing bodies.

The services provided by the Secretariat include the following:

- *WPA News*. This is a quarterly publication with issues in March, June, September and December every year. It publishes news and photos received from the Member Societies, Executive Committee members, Zonal Representatives, Scientific Sections, Affiliated Associations, etc.. Other highlights are the message from the WPA President and an update on educational activities, WPA publications and forthcoming meetings. It is edited by the Secretary General, and digital copies and a limited number of print copies are sent to all Member Societies and office bearers.
- *Directory of WPA components*. This database includes all information including postal addresses, telephone numbers, e-mail addresses, etc. of all WPA office bearers, Presidents and Secretaries of Member Societies. A printed copy is also made available during every triennium.
- *WPA Library*. This is located at the Secretariat in Geneva and has several new books and journals. Entrance to the library is unrestricted for the Member Societies and office bearers.
- *WPA archives room*. This is located at the basement of the Geneva University Hospital. We have space constraints, and efforts are now on towards e-archiving of important documents.
- *WPA central files*. These are maintained in the Secretariat in both digital and print formats.
- *WPA information folders*. They are edited and updated every three years for public relations and promotional activities.
- *WPA general survey*. This is prepared every triennium under the guidance of the Executive Committee and then distributed to all WPA components. It analyses the achievements and deficiencies and helps us to chart new directions.
- *Manual of Procedures*. This is updated every three years, reflecting changes in the Statutes and Bylaws adopted at each General Assembly, and prescribes the mode of functioning of the Secretariat and the WPA components.
- *Visitors*. The Secretariat encourages visits by Member Societies and office bearers. Their impressions are recorded in a visitor's diary kept in the office.

The WPA Secretary General is the head of the Secretariat and is ably assisted by the Administrator and the Deputy Administrator. The Administrator is in charge of handling all activities related to staff, contacts with the Swiss authorities, attending WPA Executive Committee meetings and

WPA General Assembly, liaison with the Geneva University Hospital, maintaining financial records, monitoring budgets for all Standing and Operational Committees, etc.. The Deputy Administrator handles correspondence with WPA components and replies to general inquiries and requests under the guidance of the Secretary General, updates the mailing lists, manages mass mailings of the WPA News, WPA Directory, etc.. Some tasks, such as electronic and hard copy filing and archiving, progressive organization of the materials in the archives room, are jointly performed by them.

The Secretariat works towards achieving the aims, objectives and mission of the WPA and ensuring success of the Action Plan 2014-2017 (1). We are obliged to President D. Bhugra, President-Elect H. Herrman, the members of the Executive Committee, the Zonal Representatives and the Member Societies for their constant help and support.

Making the Secretariat a global link for the psychiatrists of the world, responsive to their needs and aspirations, is our goal. We hope to work committedly towards this end!

## Reference

1. Bhugra D. The WPA Action Plan 2014-2017. World Psychiatry 2014;13:328.

DOI 10.1002/wps.20265

# The ICD-11 beta draft is available online

**Mario Luciano**
WHO Collaborating Centre for Research and Training in Mental Health, Naples, Italy

The beta draft of the chapter on mental and behavioural disorders of the eleventh revision of the International Classification of Diseases (ICD-11) is now available online at http://apps.who.int/classifications/icd11/browse/l-m/en.

In addition to reading the contents, registered users can actively contribute to the development of the chapter by: a) commenting on the available materials and responding to the comments that have already been made; b) answering some questions about the quality of the materials; c) making proposals of changes or additions to the classification.

The ICD-11 Clinical Descriptions and Diagnostic Guidelines for each disorder will finally contain the following elements: a definition, a set of inclusion and exclusion terms, a description of the essential (required) features, a characterization of the boundary of the disorder with normality (threshold for the diagnosis) and with other disorders (differential diagnosis), a series of coded qualifiers/subtypes, and a description of course features, associated clinical presentations, culture-related features, developmental presentations, and gender-related features (see 1). At present, the beta draft includes the definitions of the various disorders (summary statements of about 100-125 words each), the inclusion and exclusion terms, and, in some cases, the definitions of qualifiers/subtypes.

From the available materials, registered users are able to appreciate several features of the revised classification that have been already extensively discussed in the scientific literature (e.g., 2-13).

Among them is the introduction of the grouping of disorders specifically associated with stress, including the new categories of complex post-traumatic stress disorder and prolonged grief disorder, and an extensively revised category of adjustment disorder. Acute stress reaction is now characterized as a non-disordered response and classified among "conditions associated with psychosocial circumstances" (see 6).

The definitions and subtyping of personality disorders and bodily distress disorder have also been extensively revised and simplified (see 2,13), and are being lively discussed on the beta draft platform. The grouping of impulse control disorders now includes also pathological gambling and compulsive sexual behaviour disorder (see 11). A new name ("disorders of intellectual development") and characterization is provided for those conditions that were subsumed under the heading "mental retardation" in the ICD-10 (see 10).

In the definition of schizophrenia, disturbances of self-experience are highlighted in addition to those of thinking, perception, cognition, volition and affect. The one month duration criterion is kept, and functional impairment is not mentioned as a mandatory criterion, contrary to the DSM-5. Qualifiers referring to the course of the disorder are introduced. Schizoaffective disorder is characterized cross-sectionally as a disorder in which the diagnostic requirements for schizophrenia and a mood episode are met within the same episode of illness, either simultaneously or within a few days, contrary to the longitudinal characterization of the DSM-5 (see 3).

In the grouping of mood disorders, the concept of mixed episode, characterized by either a mixture or a very rapid alternation of prominent manic and depressive symptoms on most days during a period of at least two weeks, is kept, contrary to the DSM-5 (see 4). The categories of bipolar type II disorder and premenstrual dysphoric disorder are introduced (see 4), and the definition provided for the latter is already being debated on the beta draft platform.

In the grouping of feeding and eating disorders, subtypes of anorexia nervosa "with dangerously low body weight" and "with significantly low body weight" have been included, and the new category of avoidant-restrictive food intake disorder has been introduced (see 5).

Internet-based and clinic-based field studies of the new classification are now

ongoing (see 1). The former are being implemented through the Global Clinical Practice Network, currently including about 12,000 practitioners from all regions of the world. Psychiatrists can register to this network in any of nine languages at www.globalclinicalpractice.net.

The possibility of an interaction between the ICD-11 and the Research Domain Criteria (RDoC) projects is also being considered. Indeed, the main objectives of the two projects (i.e., improving the clinical utility of psychiatric diagnoses for the former; exploring in an innovative way the etiopathogenetic underpinnings of psychopathology for the latter) can be regarded as complementary, and much can be done to reduce the current gap between the RDoC constructs and some clinical phenomena that psychiatrists encounter in their ordinary clinical practice, especially in the area of psychoses (see 14-26).

## References


1. First MB1, 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.
2. Creed F, Gureje O. Emerging themes in the revision of the classification of somatoform disorders. Int Rev Psychiatry 2012; 24:556-67.
3. Gaebel W. Status of psychotic disorders in ICD-11. Schizophr Bull 2012;38:895-8.
4. Maj M, Reed GM. The development of the ICD-11 classification of mood and anxiety disorders. World Psychiatry 2012; 11(Suppl. 1).
5. Al-Adawi S, Bax B, Bryant-Waugh R et al. Revision of ICD: status update on feeding and eating disorders. Adv Eat Disord 2013; 1:10-20.
6. Maercker A, Brewin CR, Bryant RA et al. Diagnosis and classification of disorders specifically associated with stress: proposals for ICD-11. World Psychiatry 2013;12:198-206.
7. Maj M. "Clinical judgment" and the DSM-5 diagnosis of major depression. World Psychiatry 2013;12:89-91.
8. Wakefield JC. DSM-5 grief scorecard: assessment and outcomes of proposals to pathologize grief. World Psychiatry 2013; 12:171-3.
9. Wakefield JC, Schmitz MF. When does depression become a disorder? Using recurrence rates to evaluate the validity of proposed changes in major depression diagnostic thresholds. World Psychiatry 2013;12:44-52.
10. Bertelli MO, Salvador-Carulla L, Scuticchio D et al. Moving beyond intelligence in the revision of ICD-10: specific cognitive functions in intellectual developmental disorders. World Psychiatry 2014;13: 93-4.
11. Grant JE, Atmaca M, Fineberg NA et al. Impulse control disorders and "behavioural addictions" in the ICD-11. World Psychiatry 2014;13:125-7.
12. Wakefield JC, Schmitz MF. Uncomplicated depression is normal sadness, not depressive disorder: further evidence from the NESARC. World Psychiatry 2014;13: 317-9.
13. Tyrer P, Reed GM, Crawford MJ. Classification, assessment, prevalence, and effect of personality disorder. Lancet 2015;385: 717-26.
14. Reed GM. Toward ICD-11: improving the clinical utility of WHO's international classification of mental disorders. Prof Psychol Res Pract 2010;41:457-64.
15. International Advisory Group for the Revision of ICD-10 Mental and Behavioural Disorders. A conceptual framework for the revision of the ICD-10 classification of mental and behavioural disorders. World Psychiatry 2011;10:86-92.
16. Maj M. Mental disorders as "brain diseases" and Jaspers' legacy. World Psychiatry 2013;12:1-3.
17. McGorry PD. The next stage for diagnosis: validity through utility. World Psychiatry 2013;12:213-5.
18. Cuthbert BN. The RDoC framework: facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. World Psychiatry 2014;13:28-35.
19. Cuthbert BN; NIMH RDoC Workgroup. The RDoC framework: continuing commentary. World Psychiatry 2014;13:196-7.
20. First MB. Preserving the clinician-researcher interface in the age of RDoC: the continuing need for DSM-5/ICD-11 characterization of study populations. World Psychiatry 2014;13:54-5.
21. Jablensky A, Waters F. RDoC: a roadmap to pathogenesis? World Psychiatry 2014; 13:43-4.
22. Keshavan MS, Ongur D. The journey from RDC/DSM diagnoses toward RDoC dimensions. World Psychiatry 2014;13:44-6.
23. Maj M. Keeping an open attitude towards the RDoC project. World Psychiatry 2014; 13:1-3.
24. Parnas J. The RDoC program: psychiatry without psyche? World Psychiatry 2014; 13:46-7.
25. Sartorius N. The only one or one of many? A comment on the RDoC project. World Psychiatry 2014;13:50-1.
26. Stein DJ. An integrative approach to psychiatric diagnosis and research. World Psychiatry 2014;13:51-3.


DOI 10.1002/wps.20262

**Acknowledgement**
This publication has been partially supported by an unrestricted educational grant from Lundbeck Italia SpA, which is hereby gratefully acknowledged.





http://www.arabpsynet.com/Journals/WJ/WP-October2015.Eng.pdf