DSM-5, psychiatric epidemiology and the false positives problem

J. C. Wakefield

School of Social Work and Department of Psychiatry, New York University, New York, New York, USA

The revision effort leading to the publication of the fifth edition of the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (DSM-5) was flawed in process, goals and outcome. The revision process suffered from lack of an adequate public record of the rationale for changes, thus shortchanging future scholarship. The goals, such as dimensionalising diagnosis, incorporating biomarkers and separating impairment from diagnosis, were ill-considered and mostly abandoned. However, DSM-5's greatest problem, and the target of the most vigorous and sustained criticism, was its failure to take seriously the false positives problem. By expanding diagnosis beyond plausible boundaries in ways inconsistent with DSM-5's own definition of disorder, DSM-5 threatened the validity of psychiatric research, including especially psychiatric epidemiology. I present four examples: increasing the symptom options while decreasing the diagnostic threshold for substance use disorder, elimination of the bereavement exclusion from major depression, allowing verbal arguments as evidence of intermittent explosive disorder and expanding attention-deficit/hyperactivity disorder to adults before addressing its manifest false positives problems.

Introduction: the false positives challenge to DSM-5

The fifth edition of the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (DSM-5) (American Psychiatric Association, 2013) has been criticised for its revision process, goals and content. The most vehement and sustained objections were aimed at the content of its diagnostic criteria. It was argued that the revised criteria illegitimately expanded psychiatric diagnosis into areas of normal-range distress and other problems in living, undermining the integrity of psychiatry as a medical discipline, obscuring the meaning of its research results and potentially leading to unwarranted and possibly harmful treatment. Emblematic of this concern was the statement by Allen Frances, who had Chaired the DSM-IV (American Psychiatric Association, 1994) revision, that due to DSM-5's changes, 'Many millions of people with normal grief, gluttony, distractibility, worries, reactions to stress, the temper tantrums of childhood, the forgetting of old age and 'behavioral addictions' will soon be mislabeled as psychiatrically sick...' (Frances, 2012). The 'false positives problem'

of mislabelling normal condition as mental disorders is the issue that most impacts psychiatric epidemiology, given its heavy reliance on DSM criteria in community studies (Wakefield & Schmitz, 2010, 2011). Consequently, after brief comments on DSM-5's process and goals, I focus on DSM-5's false positives problem, arguing that this concern is largely justified and that DSM-5 was a missed opportunity to address this problem.

Failure of the process to provide an adequate scholarly record

For scholars trying to understand and evaluate the validity of the DSM-5 Task Force's decisions, the most important problem with the DSM-5 revision process was its secrecy and lack of adequate documentation. This was a major step backward from the systematic, open approach of DSM-IV. The Task Force responded to criticism by insisting that 'the process for developing *DSM-V* has been the most open and inclusive ever' (Schatzberg *et al.* 2009), and by absurdly suggesting that critics had financial motives. However, the reality is that basic elements of public access to information were lacking, and remain so. The in-process records of the workgroups' proposals and rationales that appeared periodically on the DSM-5 website

Address for correspondence: Professor J. C. Wakefield, NYU Silver School of Social Work, 1 Washington Square North, New York, New York 10003, USA.

(Email: wakefield@nyu.edu)

have disappeared, with no final summary documents posted. The many emailed comments to the website – touted by the Task Force as demonstrating the process's openness – and the workgroups' responses have never been made public. In DSM-IV, the reasoning and evidence behind each change was documented in authoritative Source Books that have proven invaluable to scholars, whereas no comparable record is planned for DSM-5. Review papers by workgroup members exist for some areas, but they are scattered through the literature and do not necessarily represent the workgroup's final rationale.

Most egregiously, the deliberations of the DSM-5 Scientific Review Committee, formed in response to the DSM-5 controversies to evaluate the strength of the scientific evidence for each proposed change and provide recommendations to the workgroups, are being kept strictly secret. It is ironic that the deliberations of a committee formed to reassure a nervous profession and public about troubling proposals emanating from confidential workgroups should themselves be kept confidential. One imagines that the reason is to avoid embarrassment; many proposals were rejected, and workgroups sometimes overrode the committee's recommendations and went ahead with rejected proposals anyway (Kendler, 2013). Finally, the unprecedented confidentiality agreements signed by all workgroup members took an additional toll. Speaking from personal experience, a workgroup member I asked to speak to my university seminar about contemplated changes in his specialty declined, explaining 'to be honest, given the confidentiality agreement, I am not really sure of what I can and cannot say.' The needlessly secretive DSM-5 mindset, antithetical to both the appearance and reality of intellectual integrity, shortchanged future scholarship.

Overreaching goals, failed aspirations

Most of DSM-5's prominently mentioned goals, including separating impairment from diagnosis, incorporating biomarkers into diagnosis, and rethinking the definition of mental disorder (Regier *et al.* 2009, 2011), were eventually abandoned, leaving a feeling of 'much ado about nothing.' The problem was that the goals were ill-conceived to begin with (Wakefield, 2009; First & Wakefield, 2010).

The Task Force's primary initiative was dimensionalisation of diagnosis: 'we have decided that one, if not the major, difference between DSM-IV and DSM-V will be the more prominent use of dimensional measures in DSM-V' (Regier *et al.* 2009, p. 649). Bread-and-butter nosological problems, including high comorbidity, excessive not-otherwise-specified

(NOS) diagnoses, and high within-category heterogeneity, were all supposed to be addressed by a shift from categories to dimensions. Engaging in the same dimensional fantasy that Eysenck unsuccessfully attempted to realise decades before (Wakefield, 1997*b*), the Task Force imagined a system of dimensions of severity of various psychological problems (e.g., psychosomatic symptoms, anxiety and depression) replacing diagnostic categories, modelled on a proposal to replace categorical personality disorders with a system of trait dimensions (which itself proved untenable for now [Wakefield, 2008, 2013*c*]). In the imagined system, comorbidity is eliminated (everyone falls at exactly one point on the system of dimensions), heterogeneity within categories is no longer an issue (each single point in multidimensional space is a unique category), and use of not-otherwise-specified diagnoses is eliminated (everyone falls somewhere in multidimensional space, so everyone is specified).

However, the conceptual underpinnings of the proposal were lacking. The Task Force failed to address the most obvious question: in a system that places everyone on a set of continuous dimensions, what is a disorder? And, how does this approach satisfy the manual's definition of mental disorder? The idea that for DSM-5, 'it's the disorder threshold, stupid' (Regier *et al.* 2004) – an agenda that would have targeted false positives – was left in the dust. The problem of thresholds was kicked down the road in a Pascal's-wager-type faith that, as Vice-Chair Regier put it, 'statistically valid cutpoints between normal and pathological' would somehow eventually emerge from research using severity dimensions (Greenberg, 2010). As a bold high-risk research program, maybe this made sense; as an approach to revising the field's diagnostic manual in vivo, it seems overreaching to say the least.

As the DSM's definition of mental disorder and my harmful dysfunction analysis of mental disorder (Wakefield, 1992*a*, *b*, 1993, 1999*a*, *b*, 2006) maintain, severity dimensions are insufficient indicators of disorder. To be a disorder, symptoms must be caused by a dysfunction. Disorder inherently involves both harmful symptoms, tracked by severity, and dysfunction, an inferred failure of some psychological mechanism to perform its biological function. Symptom severity does not necessarily reflect dysfunction; many normal conditions are, by any objective standard, symptomatically severe (e.g., normal grief after losing a child, or pain during childbirth), and some disorders are mild. False positives most often occur when harmful symptoms are mislabelled disorders without satisfying the definition's dysfunction requirement (Wakefield, 1997*a*).

Psychiatric epidemiology and the false positives problem

False positives are both a clinical/ethical issue (Wakefield, 2010*a*, 2013*b*; 2015) and a research issue, most of all for epidemiology. In attempting to establish true prevalence of mental disorder in the community, psychiatric epidemiology depends heavily on the 'conceptual validity' (Wakefield, 1992*a*, *b*; 2013*a*, 2014) of the DSM's diagnostic criteria for mental disorders – that is, the ability of the criteria to distinguish mental disorders from similar normal-range forms of distress and deviance. Conceptual validity is not the same as construct validity, which concerns identifying one aetiologically homogeneous condition; just as reliability is necessary but not sufficient for validity, conceptual validity is necessary but not sufficient for construct validity of a disorder category because a disorder category that is conceptually valid (i.e., contains only disorders) might still encompass several different constructs (as the category of schizophrenia probably does now). However, conceptual validity at least limits a category to disorders and so eliminates normal pathways to symptoms, thus is a step towards construct validity.

As psychiatry shifted its focus over the past century from the asylum, where patients were generally manifestly disordered, to the community, with its vast amount of non-disordered distress and deviance, psychiatric research – and psychiatric epidemiology in particular – confronted a novel conceptual-validity challenge (Wakefield, 2010*b*; Wakefield & First, 2013*a*, *b*). Unlike other forms of research using samples preselected for manifest pathology, psychiatric epidemiology generally confronts community distress unfiltered, relying entirely on DSM's diagnostic criteria to discriminate disorder from normality. If DSM has false positives problems, then psychiatric epidemiology has them squared.

The false positives problem has had three periods of unusual salience in American psychiatry's recent history. It first arose with a vengeance during the 1960s and 1970s as a critique of the legitimacy of psychiatry, expressed in diverse critiques that came to be labelled the 'anti-psychiatry movement.' The critiques portrayed psychiatric diagnosis as unable to distinguish the normal from the disordered and as a bogus application of medical terminology justifying the use of medical power to control socially disapproved or deviant behaviour. The DSM-III's definition of mental disorder and operationalised symptom criteria were in part a response to these critiques, and a largely successful one.

The second period occurred when the field of psychiatric epidemiology applied the new criteria to community epidemiological surveys in a quest for the true prevalence of mental disorder. The prevalence rates emerging from community studies were much higher than expected, leading to concern about false positives. As Darrel Regier stated at the time, 'Based on the high prevalence rates identified in both the ECA and the NCS, it is reasonable to hypothesise that some syndromes in the community represent transient homeostatic responses to internal or external stimuli that do not represent true psychopathologic disorders' (Regier *et al.* 1998, p. 114). Attempts to address the problem by adding clinical significance requirements to symptom criteria (Wakefield, 1996; Frances, 1998; Narrow *et al.* 2002) failed to yield a satisfying solution (Spitzer & Wakefield, 1999; Wakefield & Spitzer, 2002; Zimmerman *et al.* 2004; Wakefield *et al.* 2010).

The false-positives problem was apparent without DSM-5. As epidemiological surveys improved methodologically (e.g., from cross-sectional to longitudinal), prevalence rates increased substantially (Moffitt *et al.* 2010), to a degree challenging credibility. Moreover, when one out of five boys nationally is diagnosed with attention-deficit/hyperactivity disorder (ADHD) and most are given medication (Center for Disease Control and Prevention, 2013; Visser *et al.* 2014), and almost a quarter of all women in their 30 s and 40 s are taking antidepressants, it bears serious consideration whether diagnosis has been untethered from any medical reality.

DSM-5 offered the perfect opportunity to tackle the false positives problem. This was probably the most realistic and important overarching goal DSM-5 could have adopted. Given the growing visibility and social importance of psychiatric diagnosis, there is growing public interest in what is classified as disordered, so this issue cannot be safely hidden away in the obscurities of scholarly discourse. DSM-5's tone deaf approach to this issue brought it repeatedly and justifiably to public and media attention, triggering the current third period of focus on the false positives issue.

DSM-5 false positives progress

The news from DSM-5 is not all bad. Many false positive problems are obvious and, once noticed, are dealt with by adding contextual qualifiers to diagnostic criteria – as, for example, in requiring for selective mutism diagnosis that students are able to speak the language of instruction, a requirement added in DSM-IV. Many changes to each DSM edition consist of such commonsense conceptually driven revisions aimed at reducing false positives, and DSM-5 is no exception. Such changes involve no new empirical evidence, but rather reflect previous revisions' lack of systematic conceptual reviews. Yet, in a pivotal error, the DSM-5 Task Force rejected proposals to include a conceptual review committee in the DSM-5 revision process to avoid such errors (Kendler *et al.* 2008).

Three DSM-5 commonsense anti-false-positive changes are, first, the addition to insomnia disorder (formerly 'primary insomnia') of the criterion, 'The sleep difficulty occurs despite adequate opportunity for sleep.' DSM-IV had no such provision so, for example, one could be diagnosed with insomnia disorder if a disruptive shift-work schedule or a neighbour's loud late-night television prevented sleep. Second, DSM-5 oppositional defiant disorder newly excludes diagnosis if the defiant behaviour is directed only at a sibling. As many parents know from firsthand experience, the behaviour described in the criteria can be part of normal if frustrating sibling relations. Finally, in a welcome change that corrects an egregious long-standing oversight, DSM-5 adds to sexual dysfunction criteria the exclusion that 'The sexual dysfunction is not better explained…as a consequence of severe relationship distress (e.g., partner violence) or other significant stressors' (p. 433), so an individual who does not sexually respond to an abusive partner is no longer diagnosable as sexually dysfunctional.

There are also some evidence-based DSM-5 changes to prevent false positives. The best example is bipolar disorder. DSM-IV manic episode required 'abnormally and persistently elevated, expansive, or irritable mood.' However, a clinical validation study of community epidemiological data showed that a remarkable 71% of bipolar I DSM diagnoses were false positives, due mostly to citing the 'irritable' option when in fact there was a contextual cause of irritability (Kessler *et al*. 1997). A reanalysis of how to distinguish true cases from false positives led DSM-5 to add a requirement for 'abnormally and persistently increased activity or energy' to manic episode criteria. Requiring such activity or energy not only prevents irritability false positives but also helps reduce the frequent misdiagnosis of the emotional extremes of borderline personality disorder as bipolar disorder (Zimmerman *et al*. 2008). Given the recent extraordinary rise in bipolar diagnoses in adults and children, this change is probably one of DSM-5's most important.

Granted some victories over false positives by various DSM-5 workgroups, overall the false-positive situation was greatly worsened. I now turn to some examples.

Substance use disorder

DSM-IV had two categories of substance use disorder, dependence and abuse. Whereas dependence had good support as a valid indicator of addiction, extensive evidence indicated that abuse was not a valid disorder category, and research disconfirmed the claim that abuse is prodromal or mild dependence. Attempts to eliminate the category going back to DSM-III-R were rejected for practical reasons. To its credit, DSM-5 finally eliminated this category. However, instead of moving abuse symptoms to the V Codes, DSM-5 merged three abuse symptoms with the seven dependence symptoms and a new 'craving' criterion to form an enlarged dependence/addiction category with eleven possible symptoms, renamed 'substance use disorder' (SUD) (Hasin *et al*. 2013).

With a larger pool of eleven rather than seven symptoms, and weaker abuse symptoms now part of the mix, one might think the SUD threshold would be increased above DSM-IV's three-symptoms-out-of-seven requirement. Instead, the workgroup lowered the threshold to two symptoms out of eleven, based on an a priori decision not to lower the overall prevalence of substance use disorders from the DSM-IV level of dependence plus abuse – despite having judged abuse invalid. The two-symptom SUD threshold maintained the former prevalence.

Asked whether the new criteria will pathologise mild conditions, the workgroup Chair, Charles O'Brien, explained, 'We can treat them earlier. And we can stop them from getting to the point where they are going to need really expensive stuff like liver transplants' (Urbina, 2012). However, the individuals newly classified with SUD are unlikely to be the ones who eventually need liver transplants, nor is there evidence that the few who do end up with liver transplants would be identified and their later difficulties prevented, nor is there any cost-benefit assessment of diagnosing and treating enormous numbers of non-disordered individuals to attempt to catch a few who may be disordered. And, the numbers are very large; DSM-5 lifetime substance use disorder prevalence is 30–40% or more according to major epidemiological surveys, and only a minority satisfy a more rigorous harmful-dysfunction type criteria set (Wakefield & Schmitz, 2014*d*, *e*; 2015).

DSM-5's SUD revision allowed continued diagnosis of those with two abuse symptoms (e.g., drives under the influence and argues with spouse about it). Moreover, diagnosis newly applies to those with two dependence symptoms, dubbed 'diagnostic orphans' as if they had been incorrectly abandoned. However, studies show that both these groups resemble former abuse cases more than dependence cases in terms of prognosis, and in terms of addiction are probably are mostly false positives.

Elimination of the major depression bereavement exclusion

DSM-5's elimination of the bereavement exclusion (BE) from the diagnostic criteria for major depressive disorder (MDD) was the most controversial diagnostic revision since DSM-III's depathologisation of

homosexuality. To opponents, it seemed to fly in the face of common sense and to pathologise a normal if painful human experience. As well, it further expanded the MDD category, already bloated beyond plausibility with over half the population diagnosable at some point in life (Moffit *et al*. 2010; Rohde *et al*. 2013).

The BE was evidence-based (Maj, 2012; Wakefield, 2013*a*). Grief normally includes depressive symptoms that could be mistaken for disorder, with up to 50% of grievers reaching the 5-symptom DSM-5 MDD diagnostic threshold (Clayton *et al*. 1968), yet sometimes grief triggers true MDD (Parkes, 1964). The BE was designed to guide the clinician in distinguishing normal grief from grief that has transformed into MDD. Longitudinal studies of non-clinical samples indicate that some MDD symptoms (e.g., sadness, insomnia, lowered appetite, difficulty concentrating, lowered interest or pleasure, moderate role impairment) regularly accompany normal loss reactions and are best considered general-distress indicators. Because DSM defines MDD as having any five out of nine symptoms lasting more than 2 weeks, normal grief may spuriously qualify for MDD if it includes five of the general-distress symptoms. To avoid false positive diagnoses of the bereaved, the BE eliminated bereaved individuals from MDD diagnosis if they had only these general distress symptoms. However, bereaved individuals were still diagnosed with MDD if they experienced any one or more of six 'complicated' symptoms that went beyond the normal manifestations of grief, including psychomotor retardation, suicidal ideation, sense of worthlessness, marked impairment, psychotic ideation or lengthy prolongation of grief. The elimination of the BE meant that bereaved individuals who manifest five general-distress depressive symptoms for 2 weeks after a loss are now classifiable as having MDD.

Proponents of eliminating the BE were concerned about missing MDD cases, and cited evidence that supposedly supported their 'similarity thesis' that excluded cases are just like other standard MDD, for example in having elevated suicide rates and responding to medication. These arguments turned out to be spurious; the cited evidence did not bear on the question, and the existing research supported the BE's validity (Wakefield & First, 2012). The medication evidence, for example, consisted of one small-N uncontrolled study in which the reduction of symptoms in recently bereaved medication users was comparable to that which occurs in unmedicated recently bereaved individuals. Similarly, the claim that excluded cases have elevated suicide attempt rates was examined in four epidemiological data sets and disconfirmed (Wakefield & Schmitz, 2014*a*, *c*). Further epidemiological studies of both concurrent and predictive validity revealed that the negative

outcomes characteristic of standard MDD, including recurrence, suicide attempt and anxiety disorders, occurred no more often in BE-excluded cases than in the general population with no history of MDD, unlike standard MDD which had high rates of these negative outcomes (Mojtabai, 2011; Gilman *et al*. 2012; Wakefield & Schmitz, 2012, 2013*b*). The evidence spoke clearly, but DSM-5 refused to listen.

The controversy was so heated that, as a supposed concession, DSM-5 included a vague note suggesting that clinical judgment is needed in distinguishing normal grief from depressive disorder. The note helpfully abandons DSM-IV's absurd 2-month limitation on normal grief (Wakefield *et al*. 2011*a*, *b*), and acknowledges that stressors other than bereavement can also trigger normal depressive-like reactions, as epidemiological research strongly indicates (Wakefield *et al*. 2007; Wakefield & Schmitz, 2013*a*, *c*, 2014*a*, *b*, *c*). However, unlike the BE, the note includes no inclusion/exclusion symptom guidelines. It is thus invisible to researchers and, by opening every depression diagnosis to clinical judgment, if taken seriously would challenge the usefulness of DSM's criteria in epidemiological research (Maj, 2013).

Intermittent explosive disorder

Intermittent explosive disorder (IED) is conceptualised as a pathological failure of control over aggressive impulses, so that angry reactions are disproportionate to the situation. Despite the social importance of out-of-control rage, IED has been relatively neglected in psychiatric nosology and research in contrast to disorders of other emotions such as depression, anxiety, and elation – perhaps, some might argue, because of the lack of dedicated medications. Consequently, DSM-5's substantial changes to IED have largely fallen under the radar of critics. The revised criteria promise to bring IED more to psychiatry's attention by radically expanding the definition.

The challenge in defining pathological aggressiveness is that anger can be naturally intense. DSM-IV attempted to validly distinguish anger disorders from non-disordered anger by their extreme outcomes, requiring 'several discrete episodes of failure to resist aggressive impulses that result in serious assaultive acts or destruction of property' (p. 663). However, aggression of this severity is apparently not uncommon; McLaughlin *et al*. (2012) report that 'nearly two-thirds of adolescents (63.3%) reported lifetime anger attacks that involved destroying property, threatening violence, or engaging in violence' (p. 1133), with 39.3% of all attacks involving actual violence. To prevent invalidly overdiagnosing IED when there is assaultive or property-damaging behaviour, DSM-5

tightened DSM-IV's requirement regarding assaultive and property-damaging behaviour of 'several discrete episodes,' which could have occurred over multiple years, to a requirement of at least 'three behavioral outbursts...occurring within a 12-month period' (p. 433), a change that decreases the prevalence of qualified episodes by about a third. This narrowing was based on evidence that the narrower definition improves validity (Coccaro, 2012; McLaughlin *et al.* 2012).

However, DSM-5 then abandoned caution and, based on broader criteria developed by a leading IED researcher (Coccaro *et al.* 1998; Coccaro, 2011), added a vast new domain of less severe angry acts to those that can qualify for diagnosis. An individual can now be diagnosed with IED based on 'verbal aggression (e.g., temper tantrums, tirades, verbal arguments or fights)' or 'physical aggression toward property, animals, or other individuals' even when it 'does not result in damage or destruction of property and does not result in physical injury to animals or other individuals' – if the behaviours occur on average twice a week for at least 3 months (DSM-5, p. 466).

Regarding verbal arguments and fights, it is unclear in what sense such limited aggressive actions are truly out of control. Moreover, in conflicted couples, it is common to think that one's partner's anger is out of proportion to one's offending behaviour. From jealous rage to explosive anger after a lengthy accumulation of minor slights, proportionality is often not maintained in normal intense anger. Controlled intense aggression often takes the form of displacement, where physical aggression is directed towards an object without causing damage, such as kicking or throwing something without breaking it. As to prevalence, Coccaro *et al.*'s (2004) 6.32% ECA estimate for this broadened definition will likely prove to be much too low; depression had about the same prevalence in the ECA, and in the most recent longitudinal studies is closer to 50%. Very speculatively, the narrow DSM-IV criteria have a 7.3% lifetime prevalence in the NCS-R (Kessler *et al.* 2006), and the broader criteria seem to yield roughly 50% greater prevalence than the narrow criteria (Coccaro *et al.* 2004), and longitudinal studies seem to yield about twice the prevalence of cross-sectional studies (Moffitt *et al.* 2010), so lifetime prevalence for the new criteria could approach a quarter of the population. Most problematically, DSM-5's broadening of IED was done without epidemiological data illuminating the effect of the change on prevalence or false positives, especially in normal people who are in stressful or conflictual anger-triggering relationships or circumstances.

Attention-deficit/hyperactivity disorder

DSM-5's false positives problem also consisted of 'acts of omission' in which DSM-5 failed to address manifest threshold issues. For example, the evidence is overwhelming that ADHD is highly overdiagnosed. Of children in a given school grade, the youngest children have much higher rates of ADHD diagnosis (Elder, 2010; Evans *et al.* 2010; Zoëga *et al.* 2012), suggesting that normal variations in developmental rate are being mistaken for disorder. ADHD kids have higher rates of normal genetic variants that produce novelty seeking behaviour and less tolerance for boredom, found at higher rates in nomadic populations (Ding et al. 2002; Eisenberg *et al.* 2008). Brain development studies reveal slower development of inhibitory control in ADHD kids but no abnormal brain growth (Shaw *et al.* 2007; Sripada *et al.* 2014). However, instead of trying to refine the diagnostic criteria to address a massive false-positives problem, the DSM-5 instead altered the ADHD criteria to facilitate expanding diagnosis to adults, which risks perpetuating the same high false positive rate among adults as well by encompassing normal variation within disorder.

Conclusion

DSM-5 was a missed opportunity to increase the conceptual validity of psychiatric diagnosis by aggressively addressing false-positive issues. In squandering this opportunity, DSM-5 neglected the legacy of DSM-III and placed the hard-won integrity of psychiatry as a medical discipline at risk. It also, I suggest, ushered in what will be a third period of concern about false positives and the conceptual validity of psychiatry's scientific foundations.

Financial support

Conflict of interest

None.

References

American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn: *DSM-IV*. American Psychiatric Association: Washington, DC.

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of mental Disorders*, 5th edn: *DSM-5*. American Psychiatric Association: Arlington, VA.

Center for Disease Control and Prevention. (2013). Key findings: trends in the parent-report of health care provider diagnosis and medication treatment for ADHD: United States, 2003–2011. Retrieved 27 November 2013 from

http://www.cdc.gov/ncbddd/adhd/features/key-findings-adhd72013.html.

Clayton P, Desmarais L, Winokur G (1968). A study of normal bereavement. *American Journal of Psychiatry* **125**, 168–178.

Coccaro E, Schmidt C, Samuels J, Nestadt G (2004). Lifetime and 1-month prevalence rates of intermittent explosive disorder in a community sample. *Journal of Clinical Psychiatry* **65**, 820–824.

Coccaro EF (2011). Intermittent explosive disorder: development of integrated research criteria for Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. *Comprehensive Psychiatry* **5**, 119–125.

Coccaro EF (2012). Intermittent explosive disorder as a disorder of impulsive aggression for DSM-5. *American Journal of Psychiatry* **169**, 577–588.

Coccaro EF, Kavoussi RJ, Berman ME, Lish JD (1998). Intermittent explosive disorder–revised: development, reliability, and validity of research criteria. *Comprehensive Psychiatry* **39**, 368–376.

Ding YC, Chi HC, Grady DL, Morishima A, Kidd JR, Kidd KK, Flodman P, Spence MA, Schuck S, Swanson JM, Zhang YP, Moyzis RK (2002). Evidence of positive selection acting at the human dopamine receptor D4 gene locus. *Proceedings of the National Academy of Science USA* **99**, 309–314.

Eisenberg DTA, Campbell B, Gray PB, Sorenson MD (2008). Dopamine receptor genetic polymorphisms and body composition in undernourished pastoralists: an exploration of nutrition indices among nomadic and recently settled Ariaal men of northern Kenya. *BMC Evolutionary Biology* **8**. http://www.biomedcentral.com/1471-2148/8/173, 1–12.

Elder TE (2010). The importance of relative standards in ADHD diagnoses: evidence based on exact birth dates. *Journal of Health Economics* **29**, 641–656.

Evans WN, Morrill MS, Parente ST (2010). Measuring inappropriate medical diagnosis and treatment in survey data: the case of ADHD among school-age children. *Journal of Health Economics* **29**, 657–673.

First MB, Wakefield JC (2010). Defining 'mental disorder' in DSM-V. *Psychological Medicine* **40**, 1779–1782.

Frances A (1998). Problems in defining clinical significance in epidemiological studies. *Archives of General Psychiatry* **55**, 119.

Frances A (2012). DSM-5 is a guide, not a bible: simply ignore its 10 worst changes. Retrieved 25 December 2012 from http://www.huffingtonpost.com/allen-frances/dsm-5_b_2227626.html.

Gilman SE, Breslau J, Trinh NH, Fava M, Murphy JM, Smoller JW (2012). Bereavement and the diagnosis of major depressive episode in the National Epidemiologic Survey on Alcohol and Related Conditions. *Journal of Clinical Psychiatry* **73**, 208–215.

Greenberg G (2010). Inside the battle to define mental illness. *Wired* **19** available at http://www.wired.com/2010/12/ff_dsmv/all/ (accessed 3 February 2015), 1–5.

Hasin DS, O'Brien CP, Auriacombe M, Borges G, Bucholz K, Budney A, Compton WM, Crowley T, Ling W, Petry NM, Schuckit M, Grant BF (2013). DSM-5 criteria for substance use disorders: recommendations and rationale. *American Journal of Psychiatry* **170**, 834–851.

Kendler KS (2013). A history of the DSM-5 scientific review committee. *Psychological Medicine* **43**, 1793–1800.

Kendler KS, Appelbaum PS, Bell CC, Fulford KW, Ghaemi SN, Schaffner KF, Waterman GS, First MB, Sadler JZ (2008). Issues for DSM-V: DSM-V should include a conceptual issues work group. *American Journal of Psychiatry* **165**, 174–175.

Kessler RC, Rubinow D, Holmes C, Aberson J, Zhao S (1997). The epidemiology of DSM-III-R bipolar I disorder in a general population survey. *Psychological Medicine* **27**, 1079–1089.

Kessler RC, Coccaro EF, Fava M, Jaeger S, Jin R, Walters E (2006). The prevalence and correlates of DSM-IV intermittent explosive disorder in the National Comorbidity Survey Replication. *Archives of General Psychiatry* **63**, 669–678.

Maj M (2012). Bereavement-related depression in the DSM-5 and ICD-11. *World Psychiatry* **11**, 1–2.

Maj M (2013). "Clinical judgment" and the DSM-5 diagnosis of major depression. *World Psychiatry* **12**, 89–91.

McLaughlin KA, Green JG, Hwang I, Sampson NA, Zaslavsky AM, Kessler RC (2012). Intermittent explosive disorder in the National Comorbidity Survey Replication Adolescent Supplement. *Archives of General Psychiatry* **69**, 1131–1139.

Moffitt TE, Caspi A, Taylor A, Kokaua J, Milne B, Polanczyk G, Poulton R (2010). How common are common mental disorders? Evidence that lifetime prevalence rates are doubled by prospective versus retrospective ascertainment. *Psychological Medicine* **40**, 899–909.

Mojtabai R (2011). Bereavement-related depressive episodes: characteristics, 3-year course, and implications for DSM-5. *Archives of General Psychiatry* **68**, 920–928.

Narrow WE, Rae DS, Robins LN, Regier DA (2002). Revised prevalence estimates of mental disorders in the United States: using a clinical significance criterion to reconcile 2 survey's estimates. *Archives of General Psychiatry* **59**, 115–123.

Parkes C (1964). Recent bereavement as a cause of mental illness. *British Journal of Psychiatry* **110**, 198–204.

Regier DA, Kaelber CT, Rae DS, Farmer ME, Knauper B, Kessler RC, Norquist GS (1998). Limitations of diagnostic criteria and assessment instruments for mental disorders: implications for research and policy. *Archives of General Psychiatry* **55**, 109–115.

Regier DA, Narrow WE, Rae DS (2004). For DSM-V, it's the "disorder threshold," stupid. *Archives of General Psychiatry* **61**, 1051.

Regier DA, Narrow WE, Kuhl EA, Kupfer DJ (2009). The conceptual development of DSM-5. *American Journal of Psychiatry* **166**, 645–650.

Regier DA, Narrow WE, Kuhl EA, Kupfer DJ (2011). Introduction. In *The Conceptual Evolution of DSM-5* (ed. DA Regier, WE Narrow, EA Kuhl and DJ Kupfer), pp. xxi–xxix. American Psychiatric Publishing: Arlington, VA.

Rohde P, Lewinsohn PM, Klein DN, Seeley JR, Gau JM (2013). Key characteristics of major depressive disorder occurring in childhood, adolescence, emerging adulthood, and adulthood. *Clinical Psychological Science* **1**, 41–53.

Schatzberg AF, Scully JH, Kupfer DJ, Regier DA (2009). Setting the record straight: a response to Frances commentary on DSM-V. *Psychiatric Times* 26. Retrieved from http://www.psychiatrictimes.com/dsm/content/article/10168/1425806.

Shaw P, Eckstrand K, Sharp W, Blumenthal J, Lerch JP, Greenstein D, Clasen L, Evans A, Giedd J, Rapoport JL 2007. Attention-deficit/hyperactivity disorder is characterized by a delay in cortical maturation. *Proceedings of the National Academy of Science USA,* 104, 19649–19654.

Spitzer RL, Wakefield JC (1999). DSM-IV diagnostic criterion for clinical significance: does it help solve the false positives problem? *American Journal of Psychiatry* 156, 1856–1864.

Sripada CS, Kessler D, Angstadt M (2014). Lag in maturation of the brain's intrinsic functional architecture in attention-deficit/hyperactivity disorder. *Proceedings of the National Academy of Science USA,* September 2014. DOI: 10.1073/pnas.1407787111.

Urbina I (2012). Addiction diagnoses may rise under guideline changes. *New York Times,* May 12, 2012, A11.

Visser SN, Danielson ML, Bitsko RH, Holbrook JR, Kogan MD, Ghandour RM, Perou R, Blumberg SJ (2014). Trends in the parent-report of health care provider-diagnosed and medicated attention-deficit/hyperactivity disorder: United States, 2003–2011. *Journal of the American Academy of Child and Adolescent Psychiatry* 53, 34–46.

Wakefield JC (1992a). The concept of mental disorder: on the boundary between biological facts and social values. *American Psychologist* 47, 373–388.

Wakefield JC (1992b). Disorder as harmful dysfunction: a conceptual critique of DSM-III-R's definition of mental disorder. *Psychological Review* 99, 232–247.

Wakefield JC (1993). Limits of operationalization: a critique of Spitzer and Endicott's (1978) proposed operational criteria for mental disorder. *Journal of Abnormal Psychology* 102, 160–172.

Wakefield JC (1996). DSM-IV: are we making diagnostic progress? *Contemporary Psychology* 41, 646–652.

Wakefield JC (1997a). Diagnosing DSM-IV, Part 1: DSM-IV and the concept of mental disorder. *Behavior Research and Therapy* 35, 633–650.

Wakefield JC (1997b). Diagnosing DSM-IV, part 2: Eysenck (1986) and the essentialist fallacy. *Behaviour Research and Therapy* 35, 651–666.

Wakefield JC (1999a). Evolutionary versus prototype analyses of the concept of disorder. *Journal of Abnormal Psychology* 108, 374–399.

Wakefield JC (1999b). Disorder as a black box essentialist concept. *Journal of Abnormal Psychology* 108, 465–472.

Wakefield JC (2006). The concept of mental disorder: diagnostic implications of the harmful dysfunction analysis. *World Psychiatry* 6, 149–156.

Wakefield JC (2008). The perils of dimensionalization: distinguishing personality traits from personality disorders. *Psychiatric Clinics of North America USA* 31, 379–393.

Wakefield JC (2009). Disability and diagnosis: should role impairment be eliminated from DSM/ICD diagnostic criteria? *World Psychiatry* 8, 87–88.

Wakefield JC (2010a). False positives in psychiatric diagnosis: implications for human freedom. *Theoretical Medicine and Bioethics* 31, 5–17.

Wakefield JC (2010b). Misdiagnosing normality: psychiatry's failure to address the problem of false positive diagnoses of mental disorder in a changing professional environment. *Journal of Mental Health* 19, 337–351.

Wakefield JC (2013a). The DSM-5 debate over the bereavement exclusion: psychiatric diagnosis and the future of empirically supported practice. *Clinical Psychology Review* 33, 825–845.

Wakefield JC (2013b). DSM-5 and clinical social work: mental disorder and psychological justice as goals of clinical intervention. *Clinical Social Work Journal* 41, 131–138.

Wakefield JC (2013c). DSM-5 and the general definition of personality disorder. *Clinical Social Work Journal* 41, 168–183.

Wakefield JC (2014). Wittgenstein's nightmare: why the RDoC grid needs a conceptual dimension. *World Psychiatry* 13, 38–40.

Wakefield JC (2015). Psychological justice: DSM-5, false positive diagnosis, and fair equality of opportunity. *Public Affairs Quarterly* 29, 33–76.

Wakefield JC, First MB (2012). Does the empirical evidence support the proposal to eliminate the major depression "bereavement exclusion" in DSM-5? *World Psychiatry* 11, 3–10.

Wakefield JC, First MB (2013a). Clarifying the boundary between normality and disorder: a fundamental conceptual challenge for psychiatry. *Canadian Journal of Psychiatry* 58, 603–605.

Wakefield JC, First MB (2013b). Diagnostic validity and the definition of mental disorder: a program for conceptually advancing psychiatry. *Canadian Journal of Psychiatry* 58, 653–655.

Wakefield JC, Schmitz MF (2010). The measurement of mental disorder. In *A Handbook for the Study of Mental Health: Social Contexts, Theories, and Systems* (ed. T Teresa Scheid and T Brown), pp. 20–45. Cambridge University Press: New York.

Wakefield JC, Schmitz MF (2011). The challenge of measurement of mental disorder in community surveys. In *The Sage Handbook of Mental Health and Illness* (ed. D Pilgrim, A Rogers and B Pescosolido), pp. 26–48. Sage: New York.

Wakefield JC, Schmitz MF (2012). Recurrence of bereavement-related depression: evidence for the validity of the DSM-IV bereavement exclusion from the Epidemiologic Catchment Area Study. *Journal of Nervous and Mental Disease* 200, 480–485.

Wakefield JC, Schmitz MF (2013a). Can the DSM's major depression bereavement exclusion be validly extended to other stressors?: evidence from the NCS. *Acta Psychiatrica Scandinavica* 128, 294–305.

Wakefield JC, Schmitz MF (2013b). Normal vs. disordered bereavement-related depression: are the differences real or tautological? *Acta Psychiatrica Scandinavica* 127, 159–168.

Wakefield JC, Schmitz MF (2013c). When does depression become a disorder? Using recurrence rates to evaluate the validity of proposed changes in major depression diagnostic thresholds. *World Psychiatry* 12, 44–52.

Wakefield JC, Schmitz MF (2014*a*). Predictive validation of single-episode uncomplicated depression as a benign subtype of unipolar major depression. *Acta Psychiatrica Scandinavica* **129**, 445–457.

Wakefield JC, Schmitz MF (2014*b*). Uncomplicated depression is normal sadness, not depressive disorder: further evidence from the NESARC. *World Psychiatry* **13**, 317–319.

Wakefield JC, Schmitz MF (2014*c*). Uncomplicated depression, suicide attempt, and the DSM-5 bereavement-exclusion debate: an empirical evaluation. *Research on Social Work Practice* **24**, 37–49.

Wakefield JC, Schmitz MF (2014*d*). Corrigendum: how many people have alcohol use disorders?: using the harmful dysfunction analysis to reconcile prevalence estimates in two community surveys. *Frontiers in Psychiatry* **5**, doi: 10.3389/fpsyt.2014.00144, 1–3.

Wakefield JC, Schmitz MF (2014*e*). How many people have alcohol use disorders? using the harmful dysfunction analysis to reconcile prevalence estimates in two community surveys. *Frontiers in Psychiatry* **5**, doi: 10.3389/fpsyt.2014.00010, 1–22.

Wakefield JC, Schmitz MF (2015). The harmful dysfunction model of alcohol use disorder: revised criteria to improve the validity of diagnosis and prevalence estimates. *Addiction* Published online 26 January 2015, DOI: 10.1111/add.12859, 1–33.

Wakefield JC, Spitzer RL (2002). Lowered estimates – but of what? *Archives of General Psychiatry* **59**, 129–130.

Wakefield JC, Schmitz MF, First MB, Horwitz AV (2007). Should the bereavement exclusion for major depression be extended to other losses? Evidence from the National Comorbidity Survey. *Archives of General Psychiatry* **64**, 433–440.

Wakefield JC, Schmitz MF, Baer JC (2010). Does the DSMIV clinical significance criterion for major depression reduce false positives? Evidence from the NCS-R. *American Journal of Psychiatry* **167**, 298–304.

Wakefield JC, Schmitz MF, Baer JC (2011*a*). Did narrowing the major depression bereavement exclusion from DSM-III-R to DSM-IV increase validity? Evidence from the NCS. *Journal of Nervous and Mental Disease* **199**, 66–73.

Wakefield JC, Schmitz MF, Baer JC (2011*b*). Relation between duration and severity in bereavement-related depression. *Acta Psychiatrica Scandinavica* **124**, 487–494.

Zimmerman M., Chelminski I, Young D (2004). On the threshold of disorder: a study of the impact of the DSM-IV clinical significance criterion on diagnosing depressive and anxiety disorders in clinical practice. *Journal of Clinical Psychiatry* **65**, 1400–1405.

Zimmerman M, Ruggero CJ, Chelminski I, Young D (2008). Is bipolar disorder overdiagnosed? *Journal of Clinical Psychiatry* **69**, 935–940.

Zoëga H, Valdimarsdóttir UA, Hernández-Díaz S (2012). Age, academic performance, and stimulant prescribing for ADHD: a nationwide cohort study. *Pediatrics* **130**, 1012–1018.