Philosophy of Medicine
Original Research

On Validators for Psychiatric Categories

Miriam Solomon[1]

1 Department of Philosophy, Temple University, Philadelphia, PA, USA. Email: msolomon@temple.edu

Abstract

The concept of a "validator" as a unit of evidence for the validity of a psychiatric category has been important for more than fifty years. Validator evidence is aggregated by expert committees (for the *Diagnostic and Statistical Manual of Mental Disorders* (DSM), these are referred to as "workgroups"), which use the results to make nosological decisions. Through an examination of the recent history of psychiatric research, this paper argues that it is time to reassess this traditional practice. It concludes with specific suggestions for going forward.

1. Introduction

Until the middle of the twentieth century, diagnostic categories in psychiatry were used without much standardization or precision. When it became clear, at least by the time of Philip Ash's (1949) study, that there was only about 20% agreement among clinicians about individual diagnoses, the need for more standardization became apparent, at least among those who valued psychiatric diagnosis. Sharing knowledge and clinical information as well as collaborating on research require a background of agreed-upon categories that can be used reliably and productively.

Biologically oriented psychiatrists were especially troubled by the rampant equivocation over diagnostic categories. (Psychoanalysts were less concerned because diagnosis has a minimal role in their practices.) As treatments for psychiatric disorders were emerging—drugs, electroconvulsive treatments, talking therapies, and neurosurgeries—biologically oriented psychiatrists wanted to model psychiatry on the rest of medicine, in which diagnosis plays a pivotal role in treatment decisions.

The early versions of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM I and II) were developed for public health and military purposes. When Robert Spitzer took on the task of coordinating the work for DSM-III (around 1978), he recognized an opportunity to create a more rigorous structure that could provide a shared framework for both research and clinical care (Decker 2013). A meeting with Samuel Guze at the National Institute of Mental Health (NIMH) in 1971 had already brought Spitzer into productive engagement with the Washington University group that had been working on this task for some years. Eli Robins, Lee Robins, Samuel Guze, George Winokur, John Feighner, and their colleagues were developing operationally defined diagnostic criteria for major

ISSN 2692-3963 | Date Received 24-08-2021 | Date Accepted 13-09-2022 | Date First Published 02-12-2022
Philosophy of Medicine | DOI 10.5195/pom.2022.74 | Volume 3 | Issue 1 | pp.1–23

psychiatric diseases (Feighner et al. 1972; Robins and Guze 1970) and were already using the language of validity and validation. The Washington University group was influenced by Emil's Kraepelin's careful attention to clinical symptoms in demarcating psychiatric disorders (Engstrom and Kendler 2015) and also by Eli Robins's training with Mandel Cohen (who was himself influenced by Percy Bridgman's operationalism) (Decker 2013).[1] Above all, they wanted to bring some "toughmindedness" to thinking about psychiatric diseases (Guze 1970).

Spitzer collaborated with the Washington University group in working on the NIMH-sponsored Research Diagnostic Criteria (RDC),[2] that ended up being foundational for a much-expanded DSM-III approach to psychiatric classification (Kendler, Muñoz, and Murphy 2010).

This paper traces the concept of a "validator" for psychiatric diagnostic categories from its beginnings in Robins and Guze (1970) and Feighner et al. (1972) to the present usage in making revisions to DSM-5. The original idea was to develop reliable, predictive, and accurate diagnostic categories, using a wide range of evidence about mental disorders. This evidence constitutes the "validators." At the time of DSM-III, it was expected that diagnostic categories would continue to be revised in accordance with new evidence and gradually approach "diagnostic validity" qua correct operational criteria for diagnosing psychiatric syndromes. Ultimately, it was expected that these operational criteria would lead to the identification of distinctive causal processes underlying each kind of psychiatric disorder and, in this way, provide accounts of the nature of specific psychiatric disorders. Fifty years of research has improved the reliability and the diagnostic validity of DSM categories but not produced a pattern of continued convergence toward a single classification of psychiatric disorders. This is partly because the evidence is complex and sometimes contrary, and we do not have standard procedures for aggregating such evidence (Solomon and Kendler 2021). It is also because some recent genetic and neuroscientific findings are difficult to reconcile with DSM categories. Philosophical concerns about truth, representation, underdetermination, and reality have augmented these concerns. By the time DSM-5 was published (2013), there was a widespread "crisis of confidence in the validity of our psychiatric diagnoses" (Phillips 2013, 143). In addition, it has become clear that validity and reliability are not the only goals of psychiatric classification; various kinds of utility and ethical considerations also matter. Nevertheless, aggregation of validators in the usual expert committees is still the primary consideration for revising DSM categories. In this paper, I will argue that it is time to reevaluate this conservative practice.

2. Emergence of the Term "Validator" in the Psychiatric Literature

From the late 1950s through the early 1970s, the Washington University group worked on developing operationalized criteria, in terms of observable behavior and reported symptoms, for fourteen different diagnoses such as "primary affective disorder," "anxiety neurosis," and "alcoholism" (Feighner et al. 1972, 58, 59, 60). Using such criteria was an

[1] Note that the Washington University group was not developing operational definitions of psychiatric disorders; rather, those involved were developing operationally defined diagnostic criteria. The difference is important and will be elaborated in the next section.

[2] Should not be confused with the contemporary Research Domain Criteria (RDoC) project, also at the NIMH, and dating from 2013.

attempt to improve reliability and precision of diagnoses without (at least, obviously) taking sides in disputes between biological and psychoanalytic approaches. The criteria were, in this way, intended to be theory-free.[3] Now known as the Feighner criteria, they came from established knowledge, clinical experience, and expert judgment.[4] They were tested for reliability in clinical practice. Successive iterations refined the categories and distinguished subcategories. For example, Robins and Guze (1970) used this method to separate poor prognosis from good prognosis cases of schizophrenia on the basis of family studies, arguing that cases of "good prognosis schizophrenia" were not really cases of schizophrenia at all.

Ultimately, the goal was to get "diagnostic validity," meaning "a valid classification" (Robins and Guze 1970, 983). ("Validity" and "valid" were not explicitly defined.) The Washington University group specified what they called "five phases"—meaning five parts to their validating method—namely, five different kinds of criteria for establishing diagnostic validity: clinical description, laboratory studies, differentiation from other disorders, follow-up study, and family studies (Robins and Guze 1970). In practice, there were no laboratory studies for diagnosis of psychiatric illness (that is, surprisingly and overwhelmingly, still the case now), so they relied on evidence from the other four categories.

The clearest way to think about the logic of what the Washington University group were doing (this was not always so clear in their writing) is that they were proposing diagnostic categories (the Feighner criteria) and then confirming them, occasionally correcting them, with *further* evidence from the "five phases" ("criteria for establishing diagnostic validity"). The Feighner criteria were not taken as stipulative definitions (they did not inherit Bridgeman's operationalism) but as revisable theses about appropriate diagnostic criteria for psychiatric categories, about which much more was known than simply the diagnostic criteria. It is important not to confuse what we call the "Feighner criteria" (operational definitions of the diagnostic categories) with the "validating criteria" (confirmatory evidence for those categories).

In Robins and Guze (1970), poor prognosis schizophrenia was distinguished from good prognosis schizophrenia by showing that it had different validating evidence as well as different diagnostic criteria. Two family studies were cited in which those with poor prognosis schizophrenia were much more likely to have schizophrenia in first-degree relatives than those with good prognosis schizophrenia, who in turn were more likely than those with poor prognosis schizophrenia to have mood disorder in first-degree relatives. It is worth mentioning that the data from these studies were not "perfect." Poor prognosis schizophrenia patients sometimes did well on follow-up, and good prognosis schizophrenia patients sometimes did badly on follow-up. Poor prognosis schizophrenia patients sometimes had first-degree relatives with mood disorders, and good prognosis schizophrenia patients sometimes had first-degree relatives with schizophrenia. The

[3] These days, when we are careful not to assert that any criteria or observations can be theory-free, we might say, less controversially, that the Feighner criteria were intended to take a neutral position with respect to the psychiatric theories of the time.

[4] Some have criticized the Feighner criteria (and even the DSM criteria) on the grounds that they are based on expert consensus, rather than evidence. This is an unfair criticism, not only because the Feighner criteria were developed twenty years before evidence-based medicine, but also because the expert opinions producing the criteria were informed by a wide range of evidence, some of which was explicitly discussed. Moreover, by the time of DSM-IV (1994), systematic evidence reviews were part of the DSM process. There was, of course, still a role for expert consensus (Kendler and Solomon 2016). There is probably no more (or less) expert consensus involved in DSM construction and revision than in medical guidelines more generally (Solomon 2015).

pattern of inheritance of schizophrenia showed a good deal of noise as well as a definite signal. So, the evidence that poor prognosis schizophrenia and good prognosis schizophrenia are different diseases was somewhat mixed. I mention this lack of perfection in the data to point out that validating evidence for psychiatric categories has *always* been somewhat messy. At the time, Robins and Guze hoped that refinements in the Feighner criteria would eventually lead to much tidier evidence: "The failure to achieve 100 percent success in predicting outcome and the overlap in the results of the family studies indicate that the criteria used for the separation need further refinement" (Robins and Guze 1970, 986).

The Feighner criteria received uptake: first, in the NIMH's RDC of the late 1970s and then, through these, in DSM-III (1980) and the International Classification of Diseases (ICD). They were a starting point for the development of psychiatric categories for DSM-III, which had 265 categories.

The list of types of validating evidence was developed further by Kenneth Kendler (1980), who may have been among the first to call them "validators" (rather than "phases," "validating evidence," or "validating criteria"). Kendler participated in the revisions for DSM-III-R (1987) and subsequent revisions of the DSM. He listed a total of eight (rather than five) possible validators, notably adding "precipitating factors," "response to treatment," and "diagnostic consistency over time" as important validators (Kendler 1980, 700).[5] He commented: "It should not be assumed that all of these potential validating criteria are of equal import," portending the challenge of how to aggregate validator evidence (1980, 700). Kendler's list received wide uptake.

In his 1980 paper, Kendler used the validator framework for aggregating the evidence from the relevant psychiatric literature, arguing specifically that simple delusional disorder is different from both schizophrenia and affective diseases. The relevant validating evidence included differences with precipitating factors, premorbid personality, age of onset, functional deficits, diagnostic consistency over time, and family studies. Interestingly, there was some contrary evidence in the form of family studies (some comorbidity of the two disorders) and response to treatment (both disorders are treatable by antipsychotics), but Kendler concluded that the "bulk of the evidence" supported the conclusion that simple delusional disorder is not a mild version of schizophrenia. Implicitly, Kendler aggregated the evidence, by suggesting that an overall balance of positive evidence is what matters for validity. As with the Robins and Guze (1970) distinction between good prognosis schizophrenia and poor prognosis schizophrenia, the data are somewhat messy.

In this discussion, Kendler is explicit that he is *not* equating the diagnostic criteria with the validating evidence, which was not quite so clear in the earlier papers by Feighner et al. (1972) and Robins and Guze (1970). Kendler makes it clear that after diagnostic criteria are fixed, there can still be *further* evidence for and/or against the validity of the diagnostic disorder, showing that the selected diagnostic criteria were substantive and predictive (rather than arbitrary or stipulative).[6]

[5] Note that "diagnostic consistency over time" embeds the expectation that disorders have some stability. This may have been argued by analogy with physical diseases.

[6] In a recent paper, Kendler distinguishes what he calls an "indexical definition" from what he calls a "constitutive definition," emphasizing that DSM definitions are indexical and not constitutive. This is new language for a similar point (2017, 2054).

Although Kendler (1980) did not explicitly address the question of *how* the validators validate the diagnostic categories, it is implicit in the paper. The idea is that validators tell us something about the disease that we did not already know from the diagnostic criteria: the expected clinical course, family incidence, precipitating factors, response to treatment, and so on. With this, validators confirm that diagnostic categories describe syndromes (rather than arbitrary groups of symptoms), and knowledge that something is a syndrome (rather than an arbitrary grouping) provides predictive information. Validators make predictions about the expected features of a case, predictions that have been useful in clinical care. Arbitrary groupings do not make accurate predictions. Making accurate predictions is often termed "predictive validity" and often taken as prima facie evidence for truth.[7]

Kendler may have been the first to use the term "validator" for psychiatric categories but he did not give an explicit definition of the term. Implicitly, the meaning from Kendler (1980) is "some evidence that a diagnostic category is correct." So, for example, details of the clinical course of a disease can be a validator when the clinical course is widely observed as a syndrome (clinical features occur together more than randomly). The regular clustering of sometimes unrelated symptoms provides prima facie evidence that the syndrome is (or is caused by) a type of disease.[8] Around 1980, it was anticipated that validators would provide substantial positive evidence that well-defined psychiatric categories are correct. Although diagnostic criteria are not accounts of the "nature of" (or causes of, or definitions of) psychiatric diseases (they are diagnostic operational criteria, not operational definitions), valid diagnostic criteria are expected to be *coextensive* with those accounts, identifying the same psychiatric syndromes.

Are valid psychiatric categories the same thing as true claims about the "reality" of psychiatric disorders ("reification" is a term sometimes used for this—see, for example, Hyman 2010)? Not quite. There are at least two kinds of philosophical difficulty with the move from valid psychiatric categories (qua extensionally correct) to "real psychiatric disorders" (full accounts of the nature and causes of psychiatric disorders). One difficulty is underdetermination: there may be more than one categorization that is well supported by the available validators. Bearing in mind that validator evidence tends to be messy (rather than "perfect") increases the amount of underdetermination. Is underdetermination a problem? It is if the expectation is that the DSM process will converge on the single correct set of categories (sometimes the word "truth" is also used). Since this was the expectation through DSM-III and DSM-IV (perhaps not DSM-5), concerns about underdetermination have been appropriate for a good deal of the history of the DSM.

The second set of difficulties is in the assumptions made (explicitly or implicitly) when talking in terms of "real psychiatric disorder," rather than in terms of operationally defined diagnostic categories. These may include assumptions such as causal independence of disorders from one another (or from purportedly "unrelated" diseases), intended level(s) of causal explanation—for example, neural versus psychological versus social (or some combination thereof)—and, in particular, an assumption that "disorder" has a single meaning.

[7] The inference from predictive success to truth is highly fallible and needs much philosophical unpacking, as will be discussed in this paper.
[8] "Provides prima facie evidence" does not mean "proves" or "establishes." It is a fairly weak claim but essential in the early stages of identification of a disease (see, for example, Ankeny 2011).

A good deal of the psychiatric literature conflates the correctness of operationally defined psychiatric categories with more robust ontological claims about the nature of psychiatric disorders. In what follows, I will take care to avoid such conflation while addressing the range of ontological claims.

An additional and related philosophical point is that there is a difference between saying that psychiatric categories are validated (even "well validated") and saying that they are valid. We have a better understanding of the former, which is grounded in the epistemic concept of a "validator," than of the latter, which evokes more ontological concepts, such as truth, reality, and univocality. We make sense of more and less validation in terms of the amount of validating evidence but validity is an idealization that is less clear. I will also keep this in mind. *This paper is more about the epistemology of validation (with the validators) than it is about validity,* and I do not presume any particular concept of validity.[9]

3. Validators and Validity: The Last Forty Years

In 1980, it was reasonable to expect continuous progress in the diagnostic criteria for psychiatric diseases, with occasional revisions made in response to new evidence. Validating evidence for the main diagnostic categories was expected to accrue. Kendler's (1980) list of validators was widely cited and used. In subsequent years and versions of the DSM, scientific evidence for revision of particular DSM categories was presented in the form of evidence from validators. At this point, forty years on and nine years after the publication of DSM-5, the DSM Steering Committee's updated list of validators consists of eleven items, which I review in section 4 below.

In this section, I describe how philosophical concerns about validation and validity became more pressing. By the time of Kendler's "Toward a Scientific Psychiatric Nosology" (1990), there was a realization that the messiness of the validator evidence may not permit determinate resolution of nosological controversies. In that paper, Kendler is concerned about the possibility that different validators may point in different directions, and no longer proposes (as he did in 1980) that "the bulk of the evidence" will be univocal. Instead, he considers a conceptual/philosophical resolution, which is to resolve any under-determination using considerations of value, to be discussed explicitly in the DSM process. He recommends the same for any other "nonempirical" disagreements, such as how fine-grained psychiatric categories should be, or what the appropriate balance is between reliability and validity. He gives an example of a case in which he thinks that the evidence does not resolve classificatory choices; namely, schizotypal personality disorder. Validators from family studies support a narrower diagnostic concept than validators from clinical symptoms.

Exactly what Kendler had in mind by suggesting that the DSM committees deliberate over questions of value is not spelled out.[10] Kendler also does not pursue the implications for validation or validity of this supplement to empirical considerations.[11] Over subsequent

[9] Schaffner (2012, 2020) has explored in more detail the concept of psychiatric validity and its relationship to other concepts of validity, such as those in the psychometric literature.

[10] Kendler may have been influenced by Kuhn (1962) in his 1990 paper. It is probably too early for him to have been influenced by Longino (1990). There are many ways of integrating discussions of value into scientific deliberations.

[11] The sorts of values under consideration here do not generally track truth; see Solomon (2012).

years, Kendler has tried out several different approaches to reconciling underdetermination of diagnostic categories by the validator evidence with the goal of validity (construed as truth/reality of psychiatric categories). As he put it more recently: "While I remain committed for both scientific and personal reasons to the reality of psychiatric disorders, I have struggled to find a more acceptable way to frame those beliefs" (Kendler 2016).

In recent years, Kendler has considered Nancy Cartwright's causal realism (Kendler 2012), Hasok Chang's metaphor of epistemic iteration (Kendler 2013a), Helen Longino's approach to values and objectivity (Kendler 2015), Keith Lehrer's coherentism and pragmatic ideas about truth (Kendler 2016). So far, none of these has settled the issues, although they have each offered intriguing positions. I suggest the relevance of another philosophical view about realism, so-called whig realism (Solomon 2001), below.

Underdetermination is not the only challenge to the validity of DSM psychiatric categories. Evidence *against* the validity of specific DSM categories has also accumulated over the past forty years. It is important to note that this evidence does not "falsify" DSM categories in any simplistic Popperian way. There may be much about them that is correct. But they are also flawed, in ways that we do not yet know how to identify or repair. I will summarize this evidence briefly, as it has been covered in many publications. (For a particularly thorough overview, see Clark et al. 2017.)

Expected "zones of rarity" between syndromic diagnoses have not been found (Kendell and Jablensky 2003), suggesting that that we do not have a comprehensive disease classification. A related point is that the "otherwise specified and unspecified" subcategory,[12] into which cases that do not fit the normal polythetic criteria are assigned, is heavily used (Thiyagu Rajakannan et al. (2016) estimate that about 50% of diagnoses in outpatient care do not fit explicit polythetic criteria). And there is much uncertainty about the cutoffs: about when a cluster of symptoms is severe enough to cross the line from a normal variant to a diagnosis (Clark et al. 2017).

Although many psychiatric disorders are heritable, genetic findings have been less specific than expected. For example, genes associated with schizophrenia are also associated with bipolar disorder and with autism, suggesting that these disorders are not as different as the DSM categorizations suggest (Craddock and Owen 2005, 2010).

Treatments for psychiatric disorders are also surprisingly nonspecific (Clark et al. 2017). With the exception of a few therapies (such as lithium for bipolar disorders and exposure therapy for phobias), both pharmacological and psychosocial interventions have been successfully (as well as unsuccessfully) used in a much wider range of cases than the ones for which they were originally developed. Therapies for major depressive disorder and anxiety, for example, overlap considerably. And atypical antipsychotics, developed to prevent delusional thinking (conceptualized as cognitively deficient), are often used as mood stabilizers. This suggests that we may not be classifying disorders in ways that distinguish their causes and underlying mechanisms.

Other research has also cast doubt on the hope that diagnostic criteria are a route to an account of the nature of (or causes of) psychiatric disorders. Many comorbidities are unexplained (Clark et al. 2017), suggesting that there is a deeper level of understanding of relatedness of psychiatric categories that we do not yet have. The use of polythetic criteria—

[12] This is the DSM-5 subcategory. In previous editions of the DSM, the term used was "not otherwise specified" or "NOS."

in which there may be little overlap in symptoms of different people with the same disorder—suggests that there are underlying causal factors that we have not yet conceptualized. At this point, *reification*—taking the DSM categories to mark distinct pathologies (underlying causal mechanisms)—is often warned against (Hyman 2010). Some of the literature refers to this situation as the "crisis of validity" (for example, Phillips 2013), having in mind a deeper kind of validity than diagnostic validity.

If validity of psychiatric categories means that the current DSM psychiatric categories mark different and distinctive pathologies, DSM categories lack (a good deal of) validity. More modestly, they aim to supply diagnostic criteria and not underlying causes (that is, without reification). Even then, some of these criteria may be incomplete or incorrect. Nevertheless, the DSM psychiatric categories may have partial validity: we hope that there is *something* right about the categories specified. But we are not in a position now to identify which parts of the DSM are correct, and which are incorrect.[13] As several researchers have pointed out (Kendell and Jablensky 2003), it is less controversial to regard DSM categories as having *utility*. In this way, the validators are reinterpreted as evidence for utility rather than truth (validity). Utility can include scientific utility and predictive utility (both supplied by validator evidence) but goes beyond these to include, for example, clinical utility, forensic utility, and administrative utility. I will explore this alternative in section 6.

Despite these concerns about the validity of the DSM, the process for revision still involves aggregating evidence from validators. The most recent statement of the list is from 2019 and is from the Steering Committee that oversees updates to DSM-5. I turn to this next.

4. Current Validators Used by DSM-5 Revisions Steering Committee, with Examples of Each

The asterisk means that the validator is designated as high priority.[14]

(1) Familial aggregation*

For example, family studies (observational studies) of the incidence of schizophrenia show that the disorder is partially inherited. The studies have found that the incidence of schizophrenia is higher when a close genetic relative has the disease. Identical twin studies show that nongenetic factors also play an important role because there is about 50% concurrence in identical twins and about 5% in fraternal twins. The family studies constitute evidence that some genetic factors underlie schizophrenia and that, because of this, schizophrenia is a distinctive disease. They predict the statistical risk of developing schizophrenia.

(2) Sociodemographic and cultural factors

For example, eating disorders have a much higher incidence in adolescent girls/women than in other groups, and much higher incidence in populations where there is food security. This knowledge is based on observational (population) studies. It has predictive value in that it successfully (and usefully) identifies populations at risk. Knowing the pattern

[13] This is a position called "whig realism" (Solomon 2001).

[14] See https://www.psychiatry.org/File%20Library/Psychiatrists/Practice/DSM/DSM5-Proposal-Submissions-General-Guidance.pdf.

of incidence also provides clues to understanding eating disorders, and the clues can be investigated in further studies.

(3) Environmental risk factors

For example, a history of trauma typically precedes post-traumatic stress disorder (PTSD), and a history of sexual abuse often precedes the development of borderline personality disorder. Like the previous validator, this is based on observational studies, has predictive value, and provides clues to understanding the causes of PTSD and borderline personality disorder.

(4) Prior psychiatric history

For example, individuals with a history of dysthymia are more likely to develop major depressive disorder. This is also based on observational studies, has predictive value, and provides clues to understanding depression.

(5) Cognitive, emotional, temperament, and personality correlates (unrelated to the diagnostic criteria)

For example, in autism spectrum disorder (ASD) there is often intellectual disability, and in major depressive disorder there is often the personality trait of neuroticism. These are based on observational studies, have predictive value, and provide clues for further research on the nature of the disorders.

(6) Biological markers*

For at least a hundred years, biologically oriented psychiatrists have hoped to find biological markers for psychiatric diseases, just as biological markers have been found for many physical diseases (for example, poorly differentiated cells for carcinoma, the presence of a virus for HIV-AIDS, and elevated creatinine levels for kidney disease). Pathological findings, blood tests, genetic tests, and brain imaging tests would all count as biological markers, and we have some of these, although in general the biomarkers are not very sensitive or specific. For example, biological markers for Alzheimer's disease do not yet reliably detect the disease (see Sweeney, Sagare, and Zlokovic 2015). Exceptions in which biological markers are more sensitive and specific include narcolepsy (where the MX2 gene is less expressed) and neurocognitive disorder due to Huntington's disease (where there is a single associated gene). Interestingly (and perhaps consequently), both now fall more under neurological than psychiatric diseases.

(7) Patterns of comorbidity

Observational studies have shown that diagnoses often co-occur, more often than would be predicted by chance. For example, there is some comorbidity between attention deficit hyperactivity disorder (ADHD) and ASD in that those diagnosed with one of these are more likely than base rate to be diagnosed with the other. This comorbidity is understood as a predictable regularity, giving important information about possibly shared pathological processes—that is, indicating that the diseases are not as distinct as is suggested by the nosology.

(8) Degree or nature of functional impairment*

This is a new validator for DSM-5 and is intended to identify the categorical and/or dimensional features of some psychiatric illnesses. For example, schizophrenia typically causes serious general impairment. Major depressive disorder, on the other hand, can be mild, moderate, or severe, with only the severe cases causing serious general impairment. This is a validator that gives us important clinical information about disorders, as well as constraints on underlying pathologies.

(9) Diagnostic stability*

We expect disease classifications to be stable because the processes that cause them are ongoing. For example, through clinical experience (observations) we know that schizophrenia rarely transforms into bipolar illness (it does happen on occasion). This is evidence that these two psychiatric diseases are distinct from one another. (Since there is evidence that genetic risks for the two diseases are shared, we also have evidence against the distinctness of the two diseases. Having apparently conflicting evidence is not unusual in any area of ongoing science.)

(10) Course of illness*

Illnesses often have typical courses. For example, schizophrenia has been found to have a generally lifelong course of illness, and bipolar illness is typically relapsing and remitting. This validator gives us prognostic information (has predictive success). And it gives us information about the nature of the disease, which may aid in further research.

(11) Response to treatment*

When a diagnosed illness responds well to a specific treatment, that generally provides evidence that there is a disease process that the treatment interferes with. This is the case with the use of lithium for bipolar illness. However, most therapeutic responsiveness is less reliable and less specific to particular diseases. Selective serotonin reuptake inhibitors (SSRIs), for example, are modestly helpful for some cases of depression, anxiety, social phobia, bulimia, and other disorders but because of this wide effectiveness patient responsiveness to SSRIs does not help to tell us which diagnosis they have. Knowledge about responsiveness to treatment is always of clinical utility.

It has been traditional since Kendler (1980) to classify these into antecedent (1, 2, 3, 4), concurrent (5, 6, 7, 8) and predictive (9, 10, 11) validators. The idea is that antecedent validators can be developed from clinical records or past research, concurrent validators can be obtained in the clinical encounter, and predictive validators obtained through follow-up studies. (In a different sense, all the validators are predictive because they can all be used to make successful predictions.)

I have gone through the list of validators, giving examples of each, to show that they give evidence for a wide variety of claims about psychiatric disorders, not only the basic claim that diagnostic criteria are correctly specified. They also supply evidence about possible causes of psychiatric disorders (such as genetic markers, environmental factors, and cultural factors), give us prognoses, and suggest treatments. Validator evidence is used to support fuller (causal and predictive) accounts of the nature of psychiatric disorders, not only to support the correctness of diagnostic categories.

Implicit in the scheme of validators is the idea that when a disease has a sufficient number or quantity of validators, it is validated. Although the number or range of validators needed for validation is unspecified, the idea is that the more validators, the better validated, and we should choose the most well-validated categories. Some of the validator categories have asterisks, indicating “high priority.” This signals that they are to be more heavily weighted than the lesser priority validator evidence, although by how much is unstated. There has also been no explicit justification of why these validators are “high priority.” These issues will be discussed in the next section.

5. How Validator Evidence Can Be Aggregated

For most psychiatric categories, we have evidence from several validators, sometimes even all validators. For example, the important validators for major depressive disorder include (1) some familial aggregation; (2) some demographic patterns (for example, correlated with poverty, more frequent in women than in men); (3) environmental risk factors (for example, recent loss); (4) prior psychiatric history (of dysthymia or depression); (5) personality correlates (neuroticism); (6) biological markers (with low sensitivity and specificity—for example, MRI images, immune markers); (7) patterns of comorbidity (for example, depression is especially comorbid with anxiety); (8) degree or nature of functional impairment (depression is dimensionally measured in measures of mild, moderate, and severe, with major functional impairment only at the severe end); (9) diagnostic stability; (10) course of illness (relapsing and remitting); and (11) some responsiveness to a variety of treatments, including cognitive behavioral therapy, increase in physical activity, and SSRIs, while noting that responsiveness to these interventions is neither reliable nor specific to depression.

Because major depressive disorder has been the focus of much research, there are many studies within each category of validator, as well as many validators. The studies vary in quality, and the evidence varies in strength. Aggregating different kinds of evidence of varying quality is challenging. One systematic model we have for doing this is from evidence-based medicine, which ranks evidence in terms of quality and aggregates the highest-quality evidence using meta-analysis and systematic evidence review. However, most validator evidence is not the kind of evidence ranked highly in evidence-based medicine, which advocates randomized controlled trials wherever possible. The only validator that *could* have evidence in the form of randomized controlled trials is (11) response to treatment. Most of the validator evidence is, in fact, at the lowest level of the evidence-based medicine hierarchy: it is evidence for hypotheses about psychiatric illnesses, rather than (as is the case for randomized controlled trials) evidence for the effectiveness of an intervention. To say that validator evidence is at the lowest level of the evidence-based medicine hierarchy is not to dismiss it. It is just to point out that the formal techniques of meta-analysis and systematic evidence review are not available or applicable for aggregation of mixed levels and qualities of evidence. In saying this, I am not expressing any view (positive or negative) about the standards of evidence-based medicine. Scientific method in all fields of basic science (as contrasted with applied science) proceeds by finding evidence for and against hypotheses—and has the same aggregation challenges.

In basic scientific contexts, evidence of different kinds is rarely formally aggregated. Scientists assess evidence more informally, often coming to conclusions about "the bulk of the evidence." Recall that Kendler (1980) recommends exactly this, concluding that simple delusional disorder is a different disorder from schizophrenia on the basis of much more evidence for this conclusion than there is against it. This decision process fails when evidence falls about equally on both sides. As discussed above, Kendler (1990) recommended using considerations of value to resolve cases with complex contrary evidence, when there is much evidence on both sides.

One question to ask when aggregating validator evidence is whether some kinds of validator evidence should be weighted more heavily than other kinds of validator evidence—independently of other characteristics of the evidence, such as methodological quality. At this time, validators (1), (6), (8), (9), (10), and (11) have asterisks following the numbers,

with the explanation that asterisks indicate high priority, indicating that the DSM process indeed weights some kinds of validators more heavily than others. While there are independent epistemic reasons for thinking that (11) may have high priority,[15] any other prioritizations developed in "expert consensus" would benefit from more explicit justification, less they seem arbitrary. If there is such justification, it is not recorded in publications.

Miriam Solomon and Kenneth S. Kendler have explored five possibilities for aggregating complex validator evidence: "informal aggregation, weighted informal aggregation (simple evidence hierarchy), formal aggregation, underdetermination, and inclusion of values" (2021, 9). The first three methods—aggregation of validator evidence in more or less formal ways—at best provide an *impartial* result when all are agreed on the method of aggregation. As Jacob Stegenga has argued in the context of aggregating complex and "multimodal" evidence from clinical trials (2011), choice of method of aggregation is partly conventional, and can yield different verdicts depending on the method of aggregation used, leaving conclusions underdetermined. Impartiality is one aspect of objectivity, and thus an explicit impartial method of aggregation is perhaps preferable to informal methods, which are more susceptible to bias, but impartiality does not yield truth or validity, only (at best) general acceptance and consistency. The last two methods discussed by Solomon and Kendler (2021)—accepting underdetermination and/or including broader values in nosological choice—are also unhelpful with addressing concerns about validity, although they are helpful with attaining consensus and a shared language of psychiatric disorders.

6. On Utility and Other Considerations

As mentioned in section 3, some have addressed the difficulties with validation by suggesting that we aim for utility (usefulness) instead of validity (truth). In this section, I explore this strategy.

A psychiatric nosology can be useful in a variety of ways. One of the most important kinds of utility in psychiatry is *clinical* utility, or usefulness in taking care of patients, where this includes activities of diagnosis, prognosis, treatment, and communication between healthcare providers. Michael First (2010; First et al. 2004) has proposed developing formal measures for clinical utility and using these to supplement the validator evidence. There are some steps towards doing this in the most recent guide for submitting proposals for changes to DSM-5 (APA DSM Team 2021). I will say more about this in section 7. Robert Kendell and Assen Jablensky (2003) have gone further and suggested *replacing* the goal of validity (which they think is elusive) with the goal of utility (which is attainable, at least to some degree). Kenneth Schaffner (2012) has also emphasized clinical utility in his pragmatic philosophical framework.

While clinical utility is particularly important, there are also other kinds of utility important in psychiatry. Some other kinds of utility that can be relevant in the DSM process are research utility, predictive utility, educational utility, administrative utility, and even forensic utility. Some kinds of utility overlap with validity considerations—for example, predictive utility and research utility overlap with validity (Kendell and Jablensky would

[15] Validator (11), responsiveness to treatment, is about response to intervention. The other validator evidence is observational. Hacking (1983), Cartwright (1989), and Woodward (2005) have all argued that response to intervention is especially strong evidence.

agree with this)—and some do not (for example, educational utility and administrative utility may be at odds with validity). For clinical utility, there may be partial overlap between validity and utility considerations. Below is a brief survey of relevant kinds of utility.

Research utility includes telling us more about the disorders we are investigating, and especially providing hints about where we might look for pathological mechanisms and other causes.[16] For this kind of utility, validator evidence is relevant, so here there is an overlap between validity and utility. There is also an overlap between each kind of validator evidence and predictive utility because validator evidence supports predictions.

An example of a new diagnosis that is both clinically useful and useful for research is attenuated psychosis syndrome. It is listed in the DSM-5 as a "condition for further study." Clinically, it provides a diagnostic classification for a group of at-risk adolescents who need careful monitoring without unnecessary antipsychotic treatment. Research-wise, it identifies a group about which we want to know more—for example, we would like to find ways to predict later development of full-blown psychotic disorders.[17]

Nosological choices in psychiatry also have an impact on other kinds of utility, such as educational utility and administrative utility. For example, it has proved useful to draw the boundary of ASD widely, so that more children can benefit from the specific educational entitlements given to those with ASD in many Western countries. These educational entitlements are helpful to those not meeting more traditional (narrower) definitions of autism, making a broader category of autism of greater educational utility. On the other hand, drawing the boundary of ASD more narrowly might be of administrative utility, if the goal was to reduce educational costs.

An example of a proviso designed to take educational and administrative utility into account—prioritizing it over scientific utility and validity—is in the DSM-5 diagnosis of ASD. This grandfathers in "individuals with a well-established DSM-IV diagnosis of autistic disorder, Asperger's disorder, or pervasive developmental disorder not otherwise specified" (American Psychiatric Association 2022, 51), thus addressing public concerns that the new diagnostic criteria of ASD would leave out some of those who were benefiting from educational and health entitlements given to those with autism spectrum diagnoses. (Those receiving or being denied new diagnoses of ASD do not have this grandfathering option.)

Generally speaking, it is useful to call conditions that can be successfully treated "disorders" (rather than, for example, "conditions" or "syndromes") because disorder nomenclature is usually a requirement for the provision of treatment by healthcare administrations (whether governmental or private insurance). This usefulness is a combination of clinical utility and administrative utility. The vacillation over the "grief exclusion" (introduced in DSM-IV, eliminated in DSM-5, brought back in another form in DSM-5-TR in 2022) comes in part from the conflict between the wish to regard grief as a normal response to human loss (an ethical or humanistic constraint) and the wish to have treatment for severe grief reimbursed by health insurance companies (which is of clinical and administrative utility).

Allen Frances (2010)—who led the DSM-IV process (probably the strongest process of all the DSMs in terms of conducting systematic literature reviews of validating evidence

[16] Rachel Ankeny's paper, "Using Cases to Establish Novel Diagnoses" (2011) shows the research utility of discerning a syndromic pattern in clinical facts—she uses the example that early epidemiological studies of HIV-AIDS generated hypotheses about the infectious cause of the disease.
[17] More details about this example are in Zachar, First, and Kendler (2020).

before workgroup meetings)—also recommends consideration of practical consequences, mentioning school resources, impact on public health, impact on the legal system, and avoiding a variety of harms to patients. He even says, "To my mind, by far the most important validator is how any decision will help or harm patient care, given the foreseeable circumstances under which it will be used." This is a remarkable use of the term "validator," using it to discuss constraints *other than* the official list of validators; moreover, constraints that have little to do with truth or reality. In this context, the term "validator" means clinical utility (and perhaps other kinds of utility and considerations as well). What Frances is, in effect, saying is that helping patients is the most important goal of our classification system. This demonstrates a commitment to humanistic medicine and is resonant in our culture. But it is quite different from a commitment to validity and/or truth.

Should we follow Frances and expand the meaning of "validator" to include (clinical, and perhaps other kinds of) utility as well as validity? We could do this but then we would have to stop thinking of validators as validating. It would, consequently, be more precise to replace the term "validators" with a broader category of "utilitators." Utilitators would include the traditional validators (which are associated with predictive and research utility) but go beyond them to include all kinds of utility. Including all relevant utilitators complicates the practical task of aggregation of considerations relevant to classification. This is because rather than the two traditional goals of the DSM—reliability and validity—there are multiple kinds of utility, some of which are independent of others. For example, broadening the boundary of ASD may be of educational utility, but may confuse research into basic mechanisms underlying the disorder. Or, for example, personality disorders may be of clinical utility in psychotherapy (less so in prescribing medication) but they have not been administratively useful, since healthcare coverage for them is often denied because they are long-term conditions. Eliminating the category of childhood bipolar disorder in DSM-5 was of clinical utility (helping to avoid excess prescription of atypical antipsychotics) but it has not yet been superseded by a category with research utility.

Utility is a broader concept than validity, perhaps broad enough (when generously construed) to include all the reasons for or against a particular diagnostic category. These reasons can include patient/family acceptability, counteracting stigma, and distinguishing disorders from disapproved or unethical behaviors. Does utility include all such considerations? This may be controversial, as many philosophers argue that ethical considerations, at least, go beyond utility. This could be addressed by (somewhat trivially) allowing a category of "ethical utility," reframing ethical considerations as another kind of utility while, of course, making more complex and intractable the matter of aggregating the different kinds of utility.

The DSM-5 process added a Clinical and Public Health Committee in 2011 to consider "good clinical and public health reasons beyond scientific imperatives" (Yager and Mcintyre 2014). This committee met after recommendations were made by workgroups and vetted by the Scientific Review Committee (a committee created in 2010) (Kendler 2013b). The Clinical and Public Health Committee was charged with considering the clinical and public health implications of workgroup recommendations, focusing especially on questions such as diagnostic boundaries, public policy, communication between clinicians, billing, and reimbursement (Yager and Mcintyre 2014). In this process, reliability and validity of proposed categories was first evaluated by the Scientific Review Committee, and then the results were given to the Clinical and Public Health Committee to review some aspects of

the utility of the proposed categories. At this point, the recommendations were posted for public comment; this was the only stage in the process in which patients and their families (sometimes called "experts by experience" or "stakeholders") could give input.

Tamara Kayali Browne's suggestion (2017) that the DSM-5 revision process should—going forward—include a separate ethics and values review panel could perhaps be addressed by such a Clinical and Public Health Committee, provided that ethics experts were among the committee members. The sorts of questions that she thinks such a panel should consider include: Are conditions such as antisocial personality disorder and pedophilic disorder in fact disorders, or are they ethically problematic predispositions, or perhaps socially abhorrent inclinations? I think Browne is right to argue explicitly for the inclusion of ethical and value concerns in the DSM process. I do, however, challenge the suggestion that this work be done by a committee that is separate from, and procedurally following from, the Scientific Review Committee.

This tour of constraints on the DSM process reveals quite a range of relevant considerations. The addition of some of the constraints, such as educational and administrative utility, is the result of the DSM's (and ICD's) broad acceptance and incorporation into societal institutions. How are all these considerations to be aggregated into an overall decision about diagnostic criteria for a psychiatric category? The situation is even more complex than it was with the challenge of aggregating validator evidence; now, a range of utilities, including ethical considerations, needs to be included. And just as we do not have an algorithm for aggregating validators, we do not have one for aggregating utilitators. It is possible, indeed likely, that there is both underdetermination and individual variability in how these overall assessments are made. This means that it may be difficult to achieve consensus on one alternative, which of course is important for a classificatory system to be successful. Such problems may explain the ongoing difficulties with getting a lasting consensus on conditions such as the "grief exclusion" for depression and the category of premenstrual dysphoric disorder.

I think it is still worthwhile to distinguish between the validators (which offer evidence for the correctness of psychiatric categories) and the utilitators. This is because I do not think that the "crisis of validity" is so dire that we need to give up validity (of some kind) as a goal. Those who disagree may prefer to group the traditional validators and the utilitators together under the general category of utilitators.

The assumption (embedded in both the DSM-5 process and in Browne's paper) that scientific review should be conducted first (via the validators) and other constraints second (the utilitators) is concerning. One danger of this approach is that the top scientific contender(s) that the scientific committee reports to the committee that considers broader issues may not satisfy the utilitators. In order not to beg questions about which constraints (reliability, validity, various kinds of utility) are most important in particular cases, we need all considerations on the table from the beginning of the process. Fortunately, this is now happening in the Steering Committee evaluating proposals for revision to DSM-5.

7. The Current DSM Revision Process

It is an enormous amount of work to aggregate the evidence and other considerations for introducing, changing, or deleting DSM categories. Moreover, as Rachel Cooper (2015) has argued, revising the DSM has become more difficult as it has become more entrenched in

our clinical, research, and other institutions. The bar for making changes has risen over time. These days, it takes either compelling new evidence (in the form of validators) or compelling broader considerations (utilitators) to make a change. Realizing that it is unrealistic to reevaluate every category of disorder on a regular basis, the DSM Steering Committee has decided to consider proposals for changes on a case-by-case basis. The DSM-5 is updated online when a proposal is favorably reviewed by the Steering Committee, the relevant working group, and after a period for online public comment. It is expected to be updated in print (as with the just-published DSM-5-TR) periodically.

The "Guide to Submitting Proposals for Changes to DSM-5" (APA DSM Team 2021; henceforth "the Guide"), published online in February 2021, is a 52-page document that gathers together multiple expectations for proposals to revise DSM categories. There is no longer a Scientific Review Committee or a Clinical and Public Health Committee but instead the Steering Committee (and, when necessary, specific workgroups) takes into account four categories of considerations when evaluating a proposal: validators, reliability, clinical utility, and minimization of "deleterious consequences" (APA DSM Team 2021, 4)

Validity and reliability were always part of the evaluation of proposals for change to DSM categories but the explicit inclusion of clinical utility—with a reference to First et al. (2004)—and minimization of deleterious consequences shows that at least some broader considerations are now an explicit part of the revision process. Are all the broader considerations mentioned in section 6 included? According to the Guide, clinical utility includes "whether proposed changes improve user acceptability, clinicians' ability to apply the diagnostic criteria accurately, clinicians' adherence to practice guidelines, and ultimately clinical outcomes" (APA DSM Team 2021, 8) as well as improvements "in the clinician's ability to select the best treatment or determine prognosis" (9) (which also counts under the validators). Avoidance of deleterious consequences is not given a general definition but an example is given: overdiagnosis.

It is not clear from the Guide what the scope is of the broader considerations now deemed relevant to DSM revisions. For example, it is not clear whether patient/stakeholder acceptability or administrative utility are included, and it is not stated whether avoidance of harms encompasses affirmation of benefits. However, the Guide is remarkable in two ways: it brings broader considerations explicitly into play, and it recommends that all considerations should be on the table at the same time for consideration by one committee, rather than different committees for different considerations, as was described for the process producing DSM-5 in section 6.

A notable additional feature of the Guide is that it insists that the validators should carry more weight than the other constraints (reliability, clinical utility, and avoidance of harms). Proposals to increase reliability and clinical utility should not result in an "undue reduction in validity" (APA DSM Team 2021, 33, 35) and proposals to avoid harms should not result in a "reduction in validity" (37). All proposals need to be accompanied by a list of relevant validators. This privileging of validity is consistent with the continuation of the traditional effort to select the most-validated psychiatric categories. However, it is not explicitly justified in this way or any other way.

8. Some Other Options

The DSM *process* (not the DSM itself) has hardly changed over forty years, despite the "crisis of validity" and the acknowledgment that the validators are not the only constraints on revisions to the DSM. It is still a process that uses group expertise (mostly of psychiatrists, with some psychologists) and privileges evidence for the (extensional) correctness and (sometimes) deeper kinds of validity of psychiatric categories. Even those who continue to regard the DSM categories as having some validity (and I include myself in this group) should have considerable epistemic humility about the current DSM framework. The proliferation of other frameworks for understanding and classifying psychiatric disorders—Research Domain Criteria (RDoC), the Hierarchical Taxonomy of Psychopathology (HiTOP), network models, and predictive processing models (Insel et al. 2010; Kotov et al. 2017; Borsboom 2017; Sterzer et al. 2018)—is a healthy response to the DSM's "crisis of validity." A presumption of the DSM process, that psychiatric categories can be individuated by symptoms and behavior, is now questioned.

I think it is time to consider changing the process that began more than forty years ago, when validity seemed like a much more attainable and univocal goal. But I do not think it is necessary to drop talk of validity and validators altogether. In this section, I offer some options that can be considered individually or collectively.

Privileging validity over other constraints on revising psychiatric categories is less justified when validity is more elusive. This is because other, more attainable goals of the process—such as reliability, clinical utility, and harm avoidance—offer more definite gains. Yet the DSM-5 revision process insists that validity not be significantly compromised to attain these other goals. I think this insistence is worth questioning, as Frances did implicitly when he said that clinical utility was the most important validator (2010). He was wrong about clinical utility being a validator but perhaps not wrong in prioritizing it over the validators. I recommend that, in general, more thought should be given to the balance of validity, reliability, clinical utility, harm avoidance, and other constraints.

This may raise concerns that validity will become even more elusive, and that sacrifices of validity—for example, clinical utility or harm avoidance—will not be productive in the long run, as they are not anchored to "reality." I think these concerns are somewhat justified. One way to address them is to ensure that there is *institutional memory* about the reasons for revision of DSM categories, so that if new validator evidence is produced, there can be proper reconsideration. Historical expertise is already achieved in practice by appointing some people already experienced with the DSM process to the DSM workgroups and Steering Committee. Those with historical expertise inform committee members about past controversies and the decisions made to settle them. They have memory about the role of both validator evidence and broader considerations in coming to decisions about diagnostic criteria for each psychiatric category under discussion. They can help committees avoid wasting time and effort by reminding them about the reasons that decisions were made in the past and identifying any genuinely new data or considerations. The importance of historical expertise on the Steering Committee is something worth explicitly recognizing. That the Steering Committee is already aware of the need for this is shown in the Guide's requirement that proposals "include the historical context" (APA DSM Team 2021, 5). With this expertise on the Steering Committee, concerns that the DSM is straying too far from validity can be addressed.

Such historical expertise may come with conservative attitudes about changing psychiatric categories. While the DSM's conservativeness is key to its success, there is a danger that it will become too "locked-in" (Cooper 2015) and not sufficiently responsive to new findings. I suggest that the DSM committees continuously bring in new members, rotating them through roles as newbies, regulars, leaders, and historical experts. The DSM enterprise has not paid formal attention to group composition and process and I recommend that, going forward, it makes an effort to do so.

Validators and broader constraints can *all* change over time. This is not a surprise for the broader constraints, as clinical utility, for example, is affected by the organizational structure of healthcare, and avoidance of harm affected by improvements in treatments. It may be a surprise that the validators can change over time. But recall that validators can include such things as prevalence, environmental risk factors, and course of illness, which can vary over time and across cultures. This means that updating the DSM at regular intervals is expected.

There are some other ways in which the Steering Committee can update its expertise to address the current DSM situation, which is as much a function of its success (its widespread adoption) as its setbacks in attaining validity.

The expertise of the Steering Committee should be broadened by including experts on public health implications, educational implications, legal considerations, and ethical considerations. Current Steering Committee members cannot be expected to have this expertise at the level necessary for detailed discussions.

John Z. Sadler and Bill Fulford (2004) have argued persuasively that there should be patient (or family, when that is not possible) representation on the Steering Committee because they are experts by experience (patients with inside knowledge of their conditions) and also because they are primary stakeholders (they have most at stake). At this time, the only opportunity for patients and families to participate in the process is during the period for public comment. This opportunity, while valuable, comes with little accountability (there is no public response to public comments), and it inappropriately groups together those members of the public without relevant expertise with those who have expertise from personal experience. As Sadler and Fulford (2004) and others (for example, Stein and Phillips 2013; Tekin 2011) have argued, patients and families deserve more inclusion in the process, on both epistemic and moral grounds. ROAMER (Roadmap for Mental Health Research in Europe) successfully included "affected stakeholders" in its consensus process for setting research priorities (Wykes et al. 2015) and might have resources for guidance.

9. Conclusions

This paper has argued that the concept of a psychiatric validator is a specifically psychiatric concept that was developed for the assessment of a variety of evidence for the validity of particular psychiatric categories. What "validity" means ranges from correct diagnostic criteria to more substantive ontological claims about the causes and nature of psychiatric disorders. When the concept of a validator was first conceived, the validity of psychiatric categories was expected to develop from rough diagnostic criteria to extensionally correct diagnostic criteria to underlying mechanisms, using validator evidence all the way.

The last forty years of experience with revising the DSM has not met these expectations in two ways: the "crisis of validity" has cast doubt on the expectation of progress toward a single correct account and the need for consideration of criteria other than validators has

challenged the idea that aggregation of validators is all that is required. The "crisis of validity" is a result of the DSM not meeting expectations about progress toward either a single account or underlying causes; the need for consideration of broader criteria is the result of the DSM *exceeding* expectations for a shared framework for classifying psychiatric disorders.

DSM-III through DSM-5 used essentially the same process, privileging validator evidence in a consensus process including mostly psychiatrists and some psychologists. A consensus process has the advantage of producing a common diagnostic framework even when there is underdetermination by the evidence and/or contrary evidence for a psychiatric category. What has changed over time is the understanding of how validator evidence should be aggregated and combined with broader constraints in coming to an overall conclusion about a psychiatric category. If the DSM wants to continue its role as a widely accepted framework for classifying psychiatric disorders, it needs to take these broader constraints seriously and include broader expertise in the consensus process. The broader the participation in the process, the broader the potential buy-in to the results.

The combination of the "crisis of validity" and the recognition of the importance of broader constraints should lead to a reevaluation of the DSM process. In particular, it may sometimes be appropriate to weight broader constraints more heavily than validator evidence. And, in view of the widespread adoption of the DSM, it is time to broaden the expertise on the DSM Steering Committee to include policy experts, ethicists, experts by experience, and stakeholders.

Acknowledgments

Thanks to Ken Kendler, Kathryn Tabb, and three anonymous reviewers for *Philosophy of Medicine* for helpful feedback on an earlier version of this paper. I am also grateful to feedback from audiences at the University of Genoa, the Center for the Philosophy of Science at the University of Pittsburgh, the CAMPoS discussion group at Cambridge University, and the Philosophy and Medicine Program at King's College London. This work was completed with the support of a Guggenheim Fellowship and supported leave from Temple University.

Disclosure Statement

No competing interest was reported by the author.

References

American Psychiatric Association. 2022. *Diagnostic and Statistical Manual of Mental Disorders* (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787.

Ankeny, Rachel A. 2011. "Using Cases to Establish Novel Diagnoses: Creating Generic Facts by Making Particular Facts Travel Together." In *How Well Do Facts Travel? The Dissemination of Reliable Knowledge,* edited by Peter Howlett and Mary S. Morgan, 253–272. Cambridge: Cambridge University Press.

APA DSM Team. 2021. "Guide to Submitting Proposals for Changes to DSM-5". https://www.psychiatry.org/File%20Library/Psychiatrists/Practice/DSM/DSM5-Proposal-Submissions-General-Guidance.pdf.

Ash, Philip. 1949. "The Reliability of Psychiatric Diagnosis." *Journal of Abnormal Psychology* 44, no. 2: 272–276.

Borsboom, Denny. 2017. "A Network Theory of Mental Disorders." *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)* 16, no. 1: 5–13. https://doi.org/10.1002/wps.20375.

Browne, Tamara Kayali. 2017. "A Role for Philosophers, Sociologists and Bioethicists in Revising the DSM: A Philosophical Case Conference." *Philosophy, Psychiatry, and Psychology* 24, no. (3): 187–201. http://dx.doi.org/10.1353/ppp.2017.0024.

Cartwright, Nancy. 1989. *Nature's Capacities and Their Measurement*. Oxford: Clarendon Press; New York: Oxford University Press.

Clark, Lee Anna, Bruce Cuthbert, Roberto Lewis-Fernández, William E. Narrow, and Geoffrey M. Reed. 2017. "Three Approaches to Understanding and Classifying Mental Disorder: ICD-11, DSM-5, and the National Institute of Mental Health's Research Domain Criteria (RDoC)." *Psychological Science in the Public Interest* 18, no. 2: 72–145. https://doi.org/10.1177/1529100617727266.

Cooper, Rachel. 2015. "Why Is the Diagnostic and Statistical Manual of Mental Disorders So Hard to Revise? Path-Dependence and 'Lock-in' in Classification." *Studies in History and Philosophy of Biological and Biomedical Sciences* 51: 1–10. http://dx.doi.org/10.1016/j.shpsc.2015.03.001.

Craddock, Nick and Michael J. Owen. 2005. "The Beginning of the End for the Kraepelinian Dichotomy." *British Journal of Psychiatry* 186, no. 5: 364–366. https://doi.org/10.1192/bjp.186.5.364.

———. 2010. "The Kraepelinian Dichotomy—Going, Going … but Still Not Gone." *British Journal of Psychiatry* 196, no. 2: 92–95. https://doi.org/10.1192/bjp.bp.109.073429.

Decker, Hannah S. 2013. *The Making of DSM-III: A Diagnostic Manual's Conquest of American Psychiatry*. New York: Oxford University Press.

Engstrom, Eric J. and Kenneth S. Kendler. 2015. "Emil Kraepelin: Icon and Reality." *American Journal of Psychiatry* 172, no. 12: 1190–1196. https://doi.org/10.1176/appi.ajp.2015.15050665.

Feighner, J.P., E. Robins, S.B. Guze, R.A. Woodruff, G. Winokur, and R. Muñoz. 1972. "Diagnostic Criteria for Use in Psychiatric Research." *Archives of General Psychiatry* 26, no. 1: 57–63. https://doi.org/10.1001/archpsyc.1972.01750190059011.

First, Michael B. 2010. "Clinical Utility in the Revision of the Diagnostic and Statistical Manual of Mental Disorders (DSM)." *Professional Psychology: Research and Practice* 41, no. 6: 465–473. http://dx.doi.org/10.1037/a0021511.

First, Michael B., Harold Alan Pincus, John B. Levine, Janet B.W. Williams, Bedirhan Ustun, and Roger Peele. 2004. "Clinical Utility as a Criterion for Revising Psychiatric Diagnoses." *American Journal of Psychiatry* 161, no. 6: 946–954. https://doi.org/10.1176/appi.ajp.161.6.946.

Frances, Allen. 2010. "Should Practical Consequences Influence DSM5 Decisions? Yes, of Course." *Psychiatric Times,* 28 April. https://www.psychiatrictimes.com/view/should-practical-consequences-influence-dsm5-decisions-yes-course.

Guze, Samuel. B. 1970. "The Need for Toughmindedness in Psychiatric Thinking." *Southern Medical Journal* 63, no. 6: 662–671. https://doi.org/10.1097/00007611-197006000-00012.

Hacking, Ian. 1983. *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge: Cambridge University Press.

Hyman, Steven. 2010. "The Diagnosis of Mental Disorders: The Problem of Reification." *Annual Review of Clinical Psychology* 6, no. 1: 155–179. https://doi.org/10.1146/annurev.clinpsy.3.022806.091532.

Insel, Thomas, Bruce Cuthbert, Marjorie Garvey, Robert Heinssen, Daniel S. Pine, Kevin Quinn, Charles Stanislow, and Philip Wang. 2010. "Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders". *American Journal of Psychiatry* 167, no. 7: 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379.

Kendell, Robert and Assen Jablensky. 2003. "Distinguishing between the Validity and Utility of Psychiatric Diagnoses." *American Journal of Psychiatry* 160, no. 1: 4–12. https://doi.org/10.1176/appi.ajp.160.1.4.

Kendler, Kenneth S. 1980. "The Nosologic Validity of Paranoia (Simple Delusional Disorder): A Review." *Archives of General Psychiatry* 37, no. 6: 699–706. https://doi.org/10.1001/ARCHPSYC.1980.01780190097012.

———. 1990. "Toward a Scientific Psychiatric Nosology." *Archives of General Psychiatry* 47, no. 10: 969–973. https://doi.org/10.1001/archpsyc.1990.01810220085011.

———. 2012. "The Dappled Nature of Causes of Psychiatric Illness: Replacing the Organic-Functional/Hardware-Software Dichotomy with Empirically Based Pluralism." *Molecular Psychiatry* 17, no. 4: 377–388. https://doi.org/10.1038/mp.2011.182.

———. 2013a. "Epistemic Iteration as a Historical Model for Psychiatric Nosology: Promises and Limitations." In *Philosophical Issues in Psychiatry II: Nosology,* edited by Kenneth S. Kendler and Josef Parnas. Online edition, Oxford Academic. https://doi.org/10.1093/med/9780199642205.003.0044.

———. 2013b. "A History of the DSM-5 Scientific Review Committee." *Psychological Medicine* 43, no. 9: 1793–1800. https://doi.org/10.1017/S0033291713001578.

———. 2015. "Introduction to 'Pluralism, Incommensurability, and Scientific Change.'" In *Philosophical Issues in Psychiatry III: The Nature and Sources of Historical Change,* edited by Kenneth S. Kendler and Josef Parnas. Online edition, Oxford Academic. https://doi.org/10.1093/med/9780198725978.003.0001.

———. 2016. "The Nature of Psychiatric Disorders." *World Psychiatry* 15, no. 1: 5–12. https://doi.org/10.1002/wps.20292.

———. 2017. "DSM Disorders and Their Criteria: How Should They Inter-Relate?" *Psychological Medicine* 47, no. 12: 2054–2060. https://doi.org/10.1017/s0033291717000678.

Kendler, Kenneth S., Rodrigo. A. Muñoz, and George Murphy. 2010. "The Development of the Feighner Criteria: A Historical Perspective." *American Journal of Psychiatry* 167, no. 2: 134–142. https://doi.org/10.1176/appi.ajp.2009.09081155.

Kendler, Kenneth S. and Miriam Solomon. 2016. "Expert Consensus v. Evidence-Based Approaches in the Revision of the DSM." *Psychological Medicine* 46, no. 11: 2255–2262. https://doi.org/10.1017/S003329171600074X.

Kotov, Roman, Robert F. Krueger, David Watson, Thomas M. Achenbach, Robert R. Althoff, R. Michael Bagby, Timothy A. Brown, et al. 2017. "The Hierarchical Taxonomy of Psychopathology (HiTOP): A Dimensional Alternative to Traditional Nosologies." *Journal of Abnormal Psychology* 126, no. 4: 454–477. https://doi.org/10.1037/abn0000258.

Kuhn, Thomas S. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Longino, Helen E. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton, NJ: Princeton University Press.

Phillips, James. 2013. "The Conceptual Status of DSM-5 Diagnoses." In *Making the DSM-5: Concepts and Controversies,* edited by Joel Paris and James Phillips, 143–157. New York: Springer Science + Business Media. https://doi.org/10.1007/978-1-4614-6504-1_10.

Rajakannan, Thiyagu, Daniel J. Safer, Mehmet Burcu, and Julie Magno Zito. 2016. "National Trends in Psychiatric Not Otherwise Specified (NOS) Diagnosis and Medication Use among Adults in Outpatient Treatment." *Psychiatric Services* 67, no. 3: 289–295. https://doi.org/10.1176/appi.ps.201500045.

Robins, Eli and Samuel B. Guze. 1970. "Establishment of Diagnostic Validity in Psychiatric Illness: Its Application to Schizophrenia." *American Journal of Psychiatry* 126, no. 7: 983–987. https://doi.org/10.1176/ajp.126.7.983.

Sadler, John Z. and Bill Fulford. 2004. "Should Patients and Their Families Contribute to the DSM-V Process?" *Psychiatric Services* 55, no. 2: 133–138. https://doi.org/10.1176/appi.ps.55.2.133.

Schaffner, Kenneth F. 2012. "A Philosophical Overview of the Problems of Validity for Psychiatric Disorders." In *Philosophical Issues in Psychiatry II: Nosology,* edited by Kenneth S. Kendler and Josef Parnas, 169–189. Oxford: Oxford University Press.

———. 2020. "A Comparison of Two Neurobiological Models of Fear and Anxiety: A 'Construct Validity' Application?" *Perspectives on Psychological Science* 15, no. 5: 1214–1227. https://doi.org/10.1177/1745691620920860.

Solomon, Miriam. 2001. *Social Empiricism*. Cambridge, MA: MIT Press.

———. 2012. "Socially Responsible Science and the Unity of Values. " *Perspectives on Science* 20, no. 3: 331–338. https://doi.org/10.1162/POSC_a_00069.

———. 2015. *Making Medical Knowledge*. Oxford: Oxford University Press.

Solomon, Miriam and Kenneth S. Kendler. 2021. "The Problem of Aggregating Validators for Psychiatric Disorders." *Journal of Nervous and Mental Disease* 209, no. 1:9–12. https://doi.org/10.1097/nmd.0000000000001256.

Stegenga, Jacob. 2011. "Is Meta-Analysis the Platinum Standard of Evidence?" *Studies in History and Philosophy of* Biological *and Biomedical Sciences* 42, no. 4: 497–507. https://doi.org/10.1016/j.shpsc.2011.07.003.

Stein, Dan J. and Katharine A. Phillips. 2013. "Patient Advocacy and DSM-5." *BMC Medicine* 11, no. 1, 133. https://doi.org/10.1186/1741-7015-11-133.

Sterzer, Philipp, Rick A. Adams, Paul Fletcher, Chris Frith, Stephen M. Lawrie, Lars Muckli, Predrag Petrovic, Peter Uhlhaas, Martin Voss, and Philip R. Corlett. 2018. "The Predictive Coding Account of Psychosis." *Biological Psychiatry* 84, no. 9: 634–643. https://doi.org/10.1016/j.biopsych.2018.05.015.

Sweeney, Melanie D., Abhay P. Sagare, and Berislav V. Zlokovic. 2015. "Cerebrospinal Fluid Biomarkers of Neurovascular Dysfunction in Mild Dementia and Alzheimer's Disease." *Journal of Cerebral Blood Flow and Metabolism* 35, no. 7: 1055–1068. https://doi.org/10.1038/jcbfm.2015.76.

Tekin, Şerife. 2011. "Self-Concept through the Diagnostic Looking Glass: Narratives and Mental Disorder." *Philosophical Psychology* 24, no. 3: 357–380. https://doi.org/10.1080/09515089.2011.559622.

Woodward, James. 2005. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Wykes, Til, Josep Maria Haro, Stefano R. Belli, Carla Obradors-Tarragó, Celso Arango, José Luis Ayuso-Mateos, István Bitter, et al. 2015. "Mental Health Research Priorities for Europe." *The Lancet Psychiatry* 2, no. 11: 1036–1042. https://doi.org/10.1016/S2215-0366(15)00332-6.

Yager, Joel and John S. Mcintyre. 2014. "DSM-5 Clinical and Public Health Committee: Challenges and Considerations." *American Journal of Psychiatry* 171, no. 2: 142–144. https://doi.org/10.1176/appi.ajp.2013.13030347.

Zachar, Peter, Michael B. First, and Kenneth S. Kendler. 2020. "The DSM-5 Proposal for Attenuated Psychosis Syndrome: A History." *Psychological Medicine* 50, no. 6: 920–926. https://doi.org/10.1017/s0033291720000653.