James Phillips*

Detecting diagnostic error in psychiatry

Abstract: The question of diagnostic error in psychiatry involves two intertwined issues, diagnosis and error detection. You cannot detect diagnostic error unless you have a reliable, valid method of making diagnoses. Since the diagnostic process is less certain in psychiatry than in general medicine, that will make the detection of error less confidant. Psychiatric diagnostic categories are developed without laboratory tests and other biomarkers. These limitations dramatically weaken the validity of psychiatric diagnoses and render error detection an uncertain undertaking, with no gold standard such as laboratory findings and tissue analysis, as in most of general medicine. With these limitations in mind, I review the methods that are available for error detection in psychiatry.

Keywords: diagnosis; diagnostic error; error; psychiatry.

*Corresponding author: James Phillips, MD, Department of Psychiatry, Yale School of Medicine, 88 Noble Avenue, Milford, CT 06460, USA, Phone: (203) 877 0566, Fax: (203) 877 1404, E-mail: james.phillips@yale.edu

Introduction

The question of diagnostic error in psychiatry involves two intertwined issues, diagnosis and error detection. You cannot detect diagnostic error unless you have a reliable, valid method of making diagnoses. Since the diagnostic process is less certain in psychiatry than in general medicine, that will make the detection of error less confidant.

Diagnosis in psychiatry

Assessing diagnostic error in psychiatry involves two questions: how do we diagnose, and how do we detect error? Starting with the first question, we quickly note a significant difference from general medicine. The latter relies on a host of technological tools that supplement clinical evaluation: laboratory tests, X-ray and other diagnostic imaging, EKG, and biopsies. All of these biomarkers are generally missing in the psychiatric diagnostic process. Although there are exceptions in which definitive biomarkers are present (e.g., Huntington’s Disease), diagnosing in psychiatry is generally like diagnosing migraine and other medical conditions that have no definitive biomarkers and rely exclusively on clinical evaluation.

What are the consequences for psychiatry of diagnosing without biological and other scientific support? One consequence is that the diagnoses are entirely symptom-based. A diagnostic category is developed out of clinically observed, persistent symptom clusters. Such an approach raises obvious concerns over the validity of the categories. It is a bit like trying to diagnose chest pain or shortness of breath without the use of X-ray, EKG, and other instruments.

A second consequence is a proliferation of diagnostic methodologies, of which I will mention two prominent ones. The first, relying entirely on clinical evaluation, is what we call a prototype approach. Out of the symptom clusters we develop a series of prototypes. We diagnose, for example, schizophrenia or bipolar disorder by judging how closely the particular clinical presentation fits the respective prototype. This approach is a mixed dimensional/categorical model. Although we use a dimensional approach in assessing how close the presentation is to, or distant from, the prototype, we do not say, for instance, that the person has a little bit of schizophrenia. Rather, depending on how close the presentation is to the prototype, we say, switching to a categorical approach, that the person does or does not have schizophrenia – or that, at this moment, we cannot make the decision.

A second diagnostic methodology, designed specifically to deal with the ambiguities of the prototype model (as well as the psychoanalytic bias of earlier manuals) and make the diagnostic process more scientific, is the DSM (Diagnostic and Statistical Manual of Mental Disorders) approach, specifically since 1980 with DSM-III. The latter manual introduced the use of operational definitions with diagnostic criteria. To take the example of major depression, we make the diagnosis if the individual exhibits 5 (or more) of 9 listed symptoms (or diagnostic criteria), all for a period of two weeks or longer. (The ICD-10 takes a different approach to these two methodologies, using prototypes for the main manual and placing the diagnostic criteria in a supplemental manual, Diagnostic Criteria for Research.)

Of course the DSM approach has its own ambiguities. In the case of major depression, for instance, the clinician
has to make a decision regarding the strength of each symptom/criterion. Does the person exhibit the symptom strongly or clearly enough to count it as a positive criterion? Further, in structuring the criteria for major depression, the architects of the manual made a major decision regarding how many criteria are necessary to make the diagnosis. If, instead of 5 of the 9 criteria, you decide that 4 of the 9 criteria are sufficient for the diagnosis, you are widening the diagnostic net and possibly creating more diagnostic false-positives; if, on the other hand, you decide that the diagnosis should require 6 of the 9 criteria, you are tightening the diagnostic net and possibly creating more false negatives. There is of course no scientific way to determine what number of criteria is correct for the diagnosis. It is simply a matter of deciding whether you want to prioritize sensitivity (avoid false negatives at risk of false-positives) or specificity (avoid false-positives at risk of false negatives).

The recent DMSs (DSM-III to DSM-5) provide many pages of available scientific (e.g., epidemiologic) data, as well as elaborate descriptive, paradigm-like, presentations of the diagnostic categories. It is an irony of the DSM era that most experienced clinicians do not count diagnostic criteria but rather invoke their own quickly accessed, mental prototypes of the disorders. In that manner the DSM, criteria-based system tends to collapse back into the prototype system described above.

A straightforward example of the confusion generated by these two (both non-biologically based) approaches to psychiatric diagnosis is a recent article in the New York Times [1], “A Glut of Antidepressants,” and the study on which the Times article is based [2]. In the study Ramin Mojtabui determined the prevalence of overdiagnosed depression in community settings by using the DSM-IV criteria in interviewing patients who had carried a diagnosis of depression during the previous 12 months. Mojtabui found that only 38.4% of the participants in his study met DSM-IV criteria for depression, and concluded that “Depression overdiagnosis and overtreatment is common in community settings in the USA” [2]. The Times writer followed suit and declared that “that the condition is being overdiagnosed on a remarkable Scale.” Although both the article and the study acknowledged the use of the DSM-IV manual, neither stated that the finding was overdiagnosis as determined by the DSM standard. We can comfortably assume that the prescribing physicians were using some form of a prototype template for diagnosing depression. Consequently, since the question of diagnosis and overdiagnosis depends on which diagnostic standard is being used, the study tells us relatively little about overdiagnosis of depression in community settings. The researcher might claim that use of the DSM standard is more scientific than the practitioner’s impression, but that is of course the claim in question. When similar disagreements occur in general medicine, the researcher might at least claim more scientific status by invoking the available biomarkers.

Assessing error in psychiatric diagnosis

The assessment of error in psychiatric diagnosis butts up the same limitations that we have seen in the diagnostic process itself. In general medicine diagnostic categories (at least the most definitive categories) involve an array of biological factors. In psychiatry the biological underpinnings are missing, and the diagnostic constructs suffer accordingly. The same distinction is true of the assessment of diagnostic error. Let me illustrate the distinction in the assessment of diagnostic error with an example from general medicine. In a study of diagnostic error in internal medicine [3], Graber, Franklin, and Gordon studied 100 cases of diagnostic error. For confirmation that diagnostic error was committed, they depended on autopsy and laboratory findings. They write: “The error was revealed by tissue in 53 cases (19 autopsy specimens and 34 surgical or biopsy specimens), by definitive tests in 44 cases (24 X-ray studies and 20 laboratory investigations), and from pathognomonic clinical findings or procedure results in the remaining 3 cases” [3]. Thus, in 97 of the 100 cases, they determined error by tissue findings and definitive laboratory tests. The reason for invoking this study is to underline that it is difficult to study error without a gold standard for determining error. And for general medicine the gold standard of error is tissue findings and definitive laboratory tests.

If we now shift our focus to psychiatry, we note immediately that, as was the case in constructing diagnoses, the gold standard of general medicine is missing. As noted above, we diagnose in psychiatry with no biomarkers, no definitive laboratory tests, and no tissue findings. We diagnose in the manner of descriptive psychiatry. If we want then to assess diagnostic error, we assess without the gold standard of the rest of medicine. How do we do it? We have a number of methods, all rather unsatisfactory.

To begin with, as the above-cited study by Mojtabui illustrates, we can use the DSM (or ICD) as our gold standard. With this procedure a diagnosis is in error if it does not
meet the criteria of the DSM category. The problem is that, if this process is taken literally, it does not make sense. It would mean that a person who presented as clinically depressed but did not meet the formal criteria for depression should not be diagnosed as depressed— and that the use of that diagnosis would be a diagnostic error. The DSM can certainly serve as a vade mecum in the evaluation of a depressed individual, but gold standard it is not.

Another standard for determining error is consensus of experts, again, not even close to a gold standard. We can think of such consensus as the error-detection equivalent of prototype methodology in diagnosis. The experts agree on which diagnostic category most closely fits the clinical presentation. Compared with the gold standard of general medicine, the problems with this standard are obvious.

Still another standard for the assessment of diagnostic error in psychiatry is long-term follow-up. If a patient is diagnosed as schizophrenic and later emerges as bipolar, the early diagnosis of schizophrenia is considered an error (an error, in fact, that was frequently made in the US several decades ago). The problem with this standard for assessing error is that it often takes many years to establish the error, in sharp contrast, for instance, to the time for reading a biopsy. This kind of error, of course, stems from the use of clinical presentation to make the initial diagnosis. Given the heterogeneity of presentations and the known instability of psychiatric diagnoses [4], the use of clinical presentation can be an unreliable guide to diagnosis.

Other potential candidates for diagnosis and detection of error are family studies, genetic analysis, and specificity of treatment. Family studies are at least a weak guide; if, for instance, the question is schizophrenia versus bipolar disorder, a family loading of affective disorder tilts the decision toward bipolar disorder.

Genetic studies demonstrate the highly complex, non-Mendelian genetic picture of almost all psychiatric disorders. Although they may hold great promise for diagnosis in the future [5], they are of quite limited use at this time. Finally, the use of treatment specificity has also proved disappointing. The wide range of use of SSRI antidepressants, atypical antipsychotics, and lithium has rendered the concept of treatment specificity almost meaningless.

The discussion of error detection has thus far focused on the quality of standards for detecting error — and the limitations of available standards. We need to add to this discussion that error can also occur with the failure to apply available standards [6], and detection in those cases is detection of the failed application. Examples are neglect to take an adequate personal or family history, or to review records adequately. If evaluator has not obtained the family history of bipolar illness, the personal history of alcoholism, or the social history of criminal behavior, he or she is missing valuable ingredients of a full diagnostic assessment.

In face of the limitations of error detection standards, and barring failure to apply the available standards, the assessment of diagnostic error in psychiatry is often carried out in a number of ways: exclusive use of the DSM or consensus of experts, or a combination of both. The latter seems like that the best that have: a group of knowledgeable psychiatrists working with the DSM or ICD in a judicious way to develop the most likely diagnosis — and doing this in full recognition of the scientifically non-validated nature of the manuals themselves.

Conclusions

I conclude this essay where I began in the Introduction. The lack of strong biomarkers in psychiatry hobbles both the construction of scientifically valid diagnoses and the assessment of diagnostic error. I have suggested the limited, weakly scientific methods currently available for detecting error in psychiatric diagnoses. The future of psychiatric diagnosis and error detection is promising, interesting, and possibly surprising. We may simply firm up the scientific foundations of the current diagnostic categories (i.e., establish validity of the DSM categories), or more dramatically, scientific research may drive a major overhaul of the diagnostic system, ending in categories that are not only different but also more scientifically valid — biomarkers and all.

Conflict of interest statement: The author declares no conflict of interest.

Received October 7, 2013; accepted October 27, 2013
References