Inter-Rater Reliability in Psychiatric Diagnosis

October 05, 2012 | DSM-5 [1], Mood Disorders [2], Schizoaffective [3]
By Jeremy Matuszak, MD [4] and Melissa Piasecki, MD [5]

DSM-5 presents psychiatry with a potential “reset button” for diagnostic reliability.

Inter-rater (or intercoder) reliability is a measure of how often 2 or more people arrive at the same diagnosis given an identical set of data. While diagnostic criteria help establish reliable diagnoses, the methods of gathering and interpreting patient data have a tremendous effect on how likely it is that 2 examiners will come to the same diagnostic conclusion for a given patient. Here we review inter-rater reliability research, identify factors that threaten diagnostic reliability, and offer examples of assessment strategies that may provide increased diagnostic reliability in general clinical settings.

Reliability of psychiatric diagnoses today

In diagnostic assessment, perfect inter-rater reliability would occur when psychiatric practitioners could always arrive at the same diagnosis for a given patient. When evaluating and interpreting inter-rater reliability, the measurement statistic used is kappa; the higher the kappa ranking, the stronger the degree of agreement is between raters. Indeed, no field of medicine is able to achieve perfect agreement—there are degrees of variance among diagnosticians in other specialties.\(^2\) We are not saying that psychiatric diagnoses wholly lack objectivity, nor do we dismiss the “art” of psychiatry. Dependent on the skill of the rater and category of disorder, the inter-rater reliability of psychiatric diagnoses can be quite high and comparable to the kappas obtained for several diagnoses in other medical specialties. For example, Ruskin and colleagues\(^3\) demonstrated a kappa of 0.83 for 4 common psychiatric diagnoses. However, lacking some of the more objective measures enjoyed by other specialties, psychiatry faces additional challenges with inter-rater reliability. Similar to goals of other fields of medicine, a central goal of psychiatric nosology is diagnostic reliability and consistency. Despite this goal, however, diagnostic reliability for some psychiatric disorders remains low. For example, studies have revealed a strong discordance among the diagnoses of psychotic disorders. McGorry and colleagues\(^4\) documented 66% to 76% inter-rater reliability for this class of disorders and a corresponding discordance of 24% to 34%. Likewise, Maj and colleagues\(^5\) demonstrated a low kappa of 0.22 for the diagnosis of schizoaffective disorder.
Despite the ongoing revision of diagnostic criteria with each subsequent iteration of DSM, clinicians report concerns that diagnostic reliability generally remains poor. In a small-survey study in 2007, Aboraya asked how clinicians’ attitudes and beliefs might explain low concordance in psychiatric diagnosis. In contrast to a similar study done in 1962, clinicians in the 2007 study attributed low inter-rater reliability more to individual clinician factors than to problems with nosology. Clinician-related factors that affect psychiatric diagnosis are heterogeneous, and the following is by no means a complete list.

Differences in clinical experience may affect knowledge about psychiatric phenomenology or alter the landscape of the clinical interview. For example, comfort and familiarity with diagnostic criteria, confidence in interviewing, and time management can each affect diagnostic conclusions. A loosening of the diagnostic criteria may occur if the clinician only asks about some of the diagnostic criteria. Availability of data (eg, collateral information) may also affect the ultimate diagnosis. Other important sources of inter-rater divergence are external pressures, such as different reimbursement practices by insurance companies, exposure to popular diagnoses through the media or pharmaceutical marketing, and the cognitive biases and heuristics that affect the practice of all medical specialties.

Clearly, interview styles vary widely and no single approach is the right one. Clinical interviews are often informal with variable structure. The aid of rating scales or structured instruments, therefore, can limit omissions or biases during history taking. While some validated structured instruments (eg, the Structured Clinical Interview for DSM Disorders-Clinician Version [SCID-CV]) require too much time for administration to be practical in everyday clinical work, others (eg, the Diagnostic Interview Schedule for Children [DISC]) may be successfully administered by lay providers or by computer, allowing the clinician to subsequently focus on one or more identified areas of concern. Similarly, systematic screening instruments are available for all major categories of psychiatric illness. In light of these data, would an approach that offers more structure and objectivity increase the reliability of diagnoses? We offer 3 examples of approaches that may improve inter-rater reliability:

- A structured approach to collateral data
- The use of validated instruments to collect patient information
- A cognitive approach that anticipates critical challenges to diagnosis

Collateral data

The general clinician may not routinely seek outside information to test the accuracy and validity of a patient’s self-report, especially in outpatient settings. Yet clinicians should be cautious about obtaining most or all of the diagnostic data directly from the patient interview. Clinicians may empathize with patients even when they find a patient’s self-reported history to be implausible. In contrast, the regular, careful use of collateral information may reduce or limit biases. Opinions and diagnoses based on reviews of relevant outside material, interviews with people who know the patient and, at times, measures to screen for malingering or biased response style may provide a broader, more relevant and reliable history. If important data are missing, such as a drug screen, specifically noting diagnostic uncertainty created by the missing information will help shape the differential diagnoses.

Of course, patients who seek clinical care do not present with a box brimming with past records, and clinicians typically do not have time to sort through volumes of collateral data. Yet, in our experience, collateral data improve the reliability of diagnoses because they offset some of the problems associated with a patient’s recall and report of past events. For example, a patient may describe a past diagnosis of bipolar disorder from a recent hospitalization, but a review of past records might indicate that the patient’s discharge diagnosis was related to substances and that mood problems were characterized as a “rule out bipolar disorder versus substance-induced mood disorder.”

One practical solution is to structure the method for obtaining collateral information. Intake questionnaires are an opportunity to inquire comprehensively about current symptoms, obtain information on past treatment and hospitalizations, and obtain the names of a few people who know the patient well. Release of information forms and laboratory orders for screening tests may be paired with the intake questionnaires. Including past treatment records and laboratory results in the chart allows for a review before the follow-up appointment and critical reflection on the initial impressions from the patient’s self-report. Phone calls to the personal contacts, when consent is given, allow for further clarification, but always keep in mind the contacts’ own agendas or opportunities for secondary gain. If structured into the clinical routine, collateral information will increase the likelihood that a diagnosis is based on a full set of data.

Reflection
The third approach is the mindset that anticipates a challenge to the diagnosis or scrutiny of review by an outside party. Child psychiatrists prepare reports that are scrutinized by parents, social services, juvenile justice professionals, and payers. Forensic psychiatrists prepare opinions with cross-examination in mind. As a diagnostic opinion is formulated, a number of questions arise:

- How good is the evidence that supports this diagnosis? What are the limits of the evidence?
- Is there another legitimate interpretation of the same data?
- What is the disconfirming evidence?
- Could the individual’s psychiatric symptoms be better accounted for by the use of substances?
- Was the individual (attorney or parent) invested in a particular diagnosis and could the self-report be biased?
- What are the biases of the examiner?
- How does payer source affect the ways in which data were obtained and weighed?

Anticipation of these challenges counteracts bias and likely improves the probability of a reliable diagnosis.

**Summary**

DSM-5 presents psychiatry with a potential “reset button” for diagnostic reliability. Our field has an opportunity not only to revisit diagnostic criteria but also to revisit the ways clinicians can apply these criteria in the most valid and reliable fashion. Checklists have improved clinical outcomes in many other fields, but to date, psychiatry has not established a widely accepted or implemented checklist for screening and diagnosis. Perhaps the advent of DSM-5 and the studies of field reliability will lead to a new set of tools for the general psychiatrist’s toolbox. In the meantime, increasing collateral data, using validated screens, and using critical reflection on diagnostic evidence and influences may allow for increased confidence in diagnostic reliability and validity.

**References**


**Source URL:** [http://www.psychiatrictimes.com/dsm-5-0/inter-rater-reliability-psychiatric-diagnosis](http://www.psychiatrictimes.com/dsm-5-0/inter-rater-reliability-psychiatric-diagnosis)

**Links:**