

by AHMED ABORAYA, MD, DrPh; ERIC RANKIN, PhD; CHERYL FRANCE, MD; AHMED EL-MISSIRY, MD, MRCPsych; COLLIN JOHN

All from Department of Behavioral Medicine and Psychiatry, Robert C. Byrd Health Sciences Center of West Virginia University, Morgantown, West Virginia

The Reliability of Psychiatric Diagnosis Revisited:

The Clinician's Guide to Improve the Reliability of Psychiatric Diagnosis

ABSTRACT

Background: The authors reviewed the topic of reliability of psychiatric diagnosis from the turn of the 20th century to present. The objectives of this paper are to explore the reasons of unreliability of psychiatric diagnosis and propose ways to improve the reliability of psychiatric diagnosis. **Method:** The authors reviewed the literature on the concept of reliability of psychiatric diagnosis with emphasis on the impact of interviewing skills, use of diagnostic criteria, and structured interviews on the reliability of psychiatric diagnosis.

Results: Causes of diagnostic unreliability are attributed to the patient, the clinician and psychiatric nomenclature. The reliability of psychiatric diagnosis can be enhanced by using diagnostic criteria, defining psychiatric symptoms and structuring the interviews. **Conclusions:** The authors propose the acronym 'DR.SED,' which stands for diagnostic criteria, reference definitions, structuring the interview, clinical experience, and data. The authors recommend that clinicians use the DR.SED paradigm to improve the reliability of psychiatric diagnoses.

Key Words: reliability, psychiatric diagnosis, rating scales and structured interview.



ADDRESS CORRESPONDENCE TO:

Ahmed Aboraya, MD, DrPH, Department of Behavioral Medicine and Psychiatry, Robert C. Byrd Health Sciences Center of West Virginia University, PO Box 9137, Morgantown, WV 26506-9137
Phone: (304) 269-1210; Fax: (304) 269-5849; E-mail: aaboraya@hsc.wvu.edu

INTRODUCTION

Validity and reliability are two important topics vital to the development of modern psychiatry. Validity, in a very general sense, refers to examining the approximate truth or falsity of scientific propositions and is the topic of another paper.¹ Reliability refers to the extent to which an experiment, test, or any measuring procedure yields the same results on repeated trials and is the topic of this paper.² A valid measurement or a system is reliable by definition, but not necessarily the vice versa. Hence there is no guarantee that a reliable system is also valid. Such a system could be reliably incorrect (e.g., a scale that is always off by 10 pounds) and would therefore be invalid.

The reliability of psychiatric diagnoses has posed a serious challenge to psychiatrists, psychologists, and mental health professionals for decades. Historically, during the first half of the 20th century, clinicians were not heavily interested in making psychiatric diagnosis and did not have adequate tools to do so had they had the interest. With the beginning of the second half of the 20th century, psychiatric nomenclature started to expand gradually due to the effort of the World Health Organization (WHO) and the American Psychiatric Association (APA). WHO published the sixth revision of the International Classification of Diseases (ICD-6) in 1948, which included a mental disorders section.³ Several publications of the International Classification of Diseases (ICD) followed and the latest is the 10th edition published in 1993.⁴ In the United States, the American Psychiatric Association Committee on Nomenclature and Statistics developed and published in 1952 the first edition of the *Diagnostic and Statistical Manual: Mental Disorders* (DSM-I).⁵ Several

publications followed and the latest is the fourth edition of the DSM, Textbook Revision, published in 2000.⁶

As psychiatry moved toward the medical model, more emphasis was placed on using the psychiatric nomenclature and making psychiatric diagnoses. The use of psychiatric classification and nomenclature increases communication among clinicians about clinical features, etiology, course of illness, and treatment.⁷ As the diagnostic criteria of mental disorders have become more specific and detailed, many structured interviews were developed to measure the symptoms that comprise psychiatric disorders. The Schedules for Clinical Assessment in Neuropsychiatry (SCAN), Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I), Diagnostic Interview for Genetic Studies (DIGS), and Mini-International Neuropsychiatric Interview (MINI) are some of these structured interviews.⁸⁻¹¹

Although the diagnostic criteria of psychiatric disorders were developed along with many structured interviews, the unreliability of psychiatric diagnosis remains a serious problem.^{12,13} The goals of this article are to review the trends of reliability of psychiatric diagnosis from the turn of the 20th century to the present, explore the reasons for unreliability of psychiatric diagnosis, and propose a method to improve the reliability of psychiatric diagnosis in the clinical setting.

METHODS

Computerized literature searches were conducted using Medline and PsychInfo for entries containing the words "reliability" and "psychiatric diagnosis" that were published in English from January, 1900, to January, 2005. A medline search yielded 80

citations, and PsychInfo yielded 138 citations. Additionally, relevant references attached to published papers were also reviewed while the authors identified more papers and books through consultations with colleagues and experts in the field. The authors explored the reasons for unreliability of psychiatric diagnosis and looked for ways to improve the reliability of psychiatric diagnosis in clinical practice.

RESULTS

Trends in psychiatric nomenclature and the reliability of psychiatric diagnosis. Dohrenwend described three generations of psychiatric epidemiology studies since the turn of the 20th century.¹⁴ The following is a historical review of the status of psychiatric nomenclature and its reliability in each generation.

First generation, from the turn of the 20th century to World War II. During this generation, clinicians were not heavily interested in making psychiatric diagnoses due to the dominance of psychoanalysis. Traditional psychoanalytic thought considered psychiatric diagnosis as largely irrelevant for making psychotherapy treatment decisions.¹⁵ In addition, American psychiatry was influenced by Adolf Meyer, a prominent psychiatrist and advocate of social psychiatry who trained several generations of psychiatrists at Johns Hopkins in Baltimore between 1910 and 1941.¹⁶ For Adolf Meyer and other social psychiatrists, life history of the individual was the most significant element in the etiology of mental illness.¹⁶ Social psychiatrists feared that psychiatric nomenclature would lead to ignoring or minimizing the importance of environmental and social factors on the etiology of mental illness.

Due to the influence of psychoanalysis and social psychiatry, progress toward psychiatric nomenclature in this era was minimal. During this generation, a clinician seeking guidance on the criteria for a given diagnosis was dependent on textbooks and individual articles in which typical cases were described and the author's own conception of the illness was explicated.¹⁷ Researchers of this generation relied on key informants and agency records to identify cases of mental illness in the community.¹⁴ Consequently, the reliability of psychiatric diagnosis was rarely studied. To

generation relied on direct interviews with subjects utilizing the psychiatric nomenclature developed by the WHO and APA.¹⁴ In addition, a review of the general trends in psychiatric research between 1953 and 1983 by Reich, et al., found increasingly sophisticated research design.¹⁹ As a result, several investigators studied the reliability of psychiatric diagnoses with disappointing results. Sandifer, et al., conducted a study that involved 91 cases that were diagnosed by 10 experienced psychiatrists. The overall likelihood of a second opinion agreeing with the first was 57

had definite preferences in the selecting diagnostic categories.^{24,25}

Even with the publications of the early diagnostic manuals, the descriptions of various psychiatric disorders were so brief and general that their impact on psychiatric education, research, and clinical practice was minimal.¹⁵ Robert Spitzer spearheaded the effort of American psychiatrists, who were dissatisfied with the 1975 ICD-9, to develop and publish the DSM-III in 1980.²⁶ The DSM-III publication represented a benchmark in the history of psychiatric nomenclature because it included the long-awaited,

ALTHOUGH THE DIAGNOSTIC CRITERIA of psychiatric disorders were developed along with many structured interviews, the unreliability of psychiatric diagnosis remains a serious problem.

the best of our knowledge and consulted experts in the field, the only reliability study published in this generation was the 1938 study of Masserman and Carmichael. They followed 100 patients post-discharge from a mental hospital and found that in 40 percent of the cases, the diagnosis required "major revision" one year after discharge.¹⁸

Second generation, from World War II to the publication of DSM-III in 1980. During this generation, psychiatric nomenclature and diagnosis expanded gradually with the publications of ICD 6, 7, 8, and 9 and DSM-I and -II. As opposed to the first generation, studies of this

percent.²⁰ Another study by Schmidt, et al., involved 426 state hospital patients who were diagnosed independently by a pair of psychiatrists. Agreement with respect to specific diagnostic subtypes occurred only in one half of the cases.²¹ Beck, et al., examined 153 outpatients and found 54-percent agreement for specific diagnoses.²² Kreitman, et al., found 63-percent agreement on 90 outpatients examined independently at approximately three-day intervals.²³ A review of six studies between 1956 and 1972 by Spitzer and Fleiss using Kappa statistic showed that the reliability of psychiatric diagnoses is still a major problem.⁷ Another observation was that psychiatrists

detailed, explicit, specific criteria of many psychiatric disorders. The DSM-III was shown to be used and preferred over the ICD-9 by psychiatrists and mental health professionals worldwide.²⁷ The publication of the DSM-III in 1980 also marked the beginning of the present generation of psychiatric epidemiology studies.

Third generation, from the publication of DSM-III in 1980 to the present. With the publication of the DSM-III and subsequently the DSM-III-R and DSM-IV as well as the ICD-10, studies on the reliability of psychiatric diagnosis have expanded greatly. Searches from Medline between January, 1980, and January, 2005, for entries that

TABLE 1. The reliability of psychiatric diagnosis in three generations

FIRST GENERATION—from the turn of the 20th century to 1945—Psychiatrists used textbooks and articles describing typical cases as guide on criteria of mental disorders. One paper was published in 1938 by Masserman on reliability of psychiatric diagnosis.

SECOND GENERATION—from 1945 to 1980—ICD 6, 7, 8, and 9 and DSM-I and -II were published. Several reliability papers were published with disappointing results.

THIRD GENERATION—from 1980 to present—ICD 10 and DSM-III, III-R, and -IV were published. More reliability papers were published and the reliability of psychiatric diagnosis has improved.

contained the words “reliability” and “psychiatric diagnosis” yielded 62 citations published in English. Similar searches from Medline between January, 1945, and January, 1980, for the same entries yielded 18 citations published in English. Searches from PsychInfo between January, 1980, and January, 2005, for entries that contained the words “reliability” and “psychiatric diagnosis” yielded 105 citations published in English. Similar searches from PsychInfo between January, 1945, and January, 1980, for the same entries yielded 33 citations published in English.

The design of reliability studies in this generation has continued to improve and has become more elaborate and sophisticated. Reliability studies have expanded to include studies of adolescent and elderly populations.²⁸⁻³⁴ Psychiatric diagnoses made remotely by telecommunication has even been studied and found to be reliable.³⁵ Overall, the reliability of psychiatric diagnoses of this generation has improved at the level of research studies due to the use of stringent design, diagnostic criteria, and structured interviews.³⁶⁻³⁸ Table 1 summarizes the reliability of psychiatric diagnoses in the three generations.

However, the reliability of psychiatric diagnosis among practicing clinicians is still poor.¹³ A review of the chart of inpatient

admissions of a single patient will reveal multiple different diagnoses for the same patient.

Causes of unreliability of psychiatric diagnosis. Reasons for diagnostic unreliability in psychiatry have been studied. Ward, et al., conducted a study to pinpoint the reasons for diagnostic disagreement among psychiatrists. One experienced psychiatrist interviewed the patient first and a second psychiatrist interviewed the patient after a resting period of few minutes. After the second interview, both psychiatrists met, discussed their diagnosis and established reasons for disagreement. The three main reasons for diagnostic disagreement were: inconstancy of the patient (5%), inconstancy of the clinician (32.5%), and inadequacy of the nomenclature (62.5%).³⁹

Patient factors—Patient’s psychological state. Some patients are in a state that enables them to provide useful and reliable information to the clinician. Some patients may forget important information due to anxiety, poor concentration, or poor memory. Other patients with disorganized thoughts are unable to provide any useful information due to psychosis. Others may omit information due to shame, denial, fear of legal consequences, to obtain or avoid particular treatments, or for a host of other

reasons. Patients with personality disorders may make an effort to manipulate the clinician. Such factors related to the patient are usually not under the control of the clinician, whose role is to elicit the information from the patient, assess the information, and make a rational judgment in the form of psychiatric diagnosis.

Patient factors—The use of proxy information. For patients who are unable or unwilling to provide reliable information, the clinician must, when possible, resort to proxy information, which can be either incomplete or distorted. The individual providing the proxy information may have a vested interest in minimizing or exaggerating elements of the history. Perhaps the most extreme example of a problem with reliance on proxy information would be Munchausen Syndrome by Proxy (MSP). MSP is a label for a pattern of behavior in which caretakers deliberately exaggerate and/or fabricate and/or induce physical and/or psychological-behavioral-mental health problems in others. In such instances it would be anticipated that information by the proxy would only obfuscate a reliable diagnosis. Another example is the inclusion of family input in making a diagnosis of early dementia, which is often essential. In that case, clinical reports by concerned family can vary widely depending upon whether they have a primary or secondary role in the patient care.⁴⁰ Whether the family member resides with the patient, has contact during day and evening times, or is limited to patient reports (e.g., telephone conversations) rather than direct observation will all influence the quality of their information and its reliability. Therefore, when soliciting proxy information, it is important that the clinician be able to assess its quality in order to determine its relative utility.

Patient factors—Atypical

presentations of psychiatric disorders. Since most clinicians use the DSM and ICD criteria, typical presentation can be defined as the psychiatric disorder that meets the criteria specified in the DSM or the ICD manuals. For example, a patient with a major depressive episode should have a two-week period of depressed mood or anhedonia and four of the following symptoms: weight loss, insomnia or hypersomnia, fatigue, feeling of worthlessness or guilt, poor concentration, or suicidal thoughts. Some patients have typical presentations of a major depressive episode. Many others with serious depressive symptoms that require clinical treatment do not fit the criteria described in the textbooks. In other instances, the physiologic

atypical presentations pose a challenge even to experienced clinicians and cause diagnostic unreliability. It is important to remember that the majority of patients do not fit the classic psychiatric diagnoses. For most patients, forcing the diagnostician to choose among the categories requires an arbitrary decision that may contribute to the unreliability of psychiatric diagnoses.⁴² That explains why removing cases with an atypical presentation increases the reliability of psychiatric diagnosis.⁴³

Clinician Factors—The clinician interview. Clinicians learn the skills of interviewing during their training. As a clinician practices interviewing, each develops his or her own

symptoms that have brought the patient to clinical attention to the detriment of other symptoms, which are present, but less acute. For example, the same patient might present with mainly depressive symptoms at one time and with nightmares and flashbacks of past abuse at another time. If only focusing on the present prominent symptoms, the clinician may diagnose major depression at one visit and posttraumatic stress disorder at the next. However, without further exploration, it is unclear if the patient has both diagnoses or has only one with some overlapping symptoms of a second. The workload and time constraints of clinicians create a disincentive to perform a comprehensive assessment and

THE DMS-III PUBLICATION represented a benchmark in the history of psychiatric nomenclature because it included the long-awaited, detailed, explicit, specific criteria of many psychiatric disorders.

changes associated with normal age often influence the presentation of psychiatric disorders in later life. For example, current DSM-IV diagnosis for alcohol dependency requires the presence of tolerance and withdrawal. However, it has been noted for elders, dependency upon alcohol may not result in either clinical symptom due to a decline in their ability to clear the substance.⁴¹ This, in turn, is likely to cause diagnostic unreliability in this age group. Such patients with

style of interviewing. Clinicians with good interviewing skills establish a therapeutic rapport with the patient and make the patient feel comfortable enough to provide the information.

Most clinicians use an open-form type of interview in routine psychiatric evaluation. The reliability of psychiatric diagnoses using this type of routine clinical psychiatric evaluation has been found to be low.⁷ There are three main reasons for this. First, clinicians typically focus on the most pressing presenting

assign multiple diagnoses.⁴⁴ Second, clinicians are pressured by institutional requirements and financial incentives to make diagnoses that reimburse at a higher rate.^{44,45} Patients, their families, or the clinician's desire to be helpful to the patient may lead him or her to make a diagnosis that will facilitate receiving governmental assistance or insurance coverage. Third, use of an unstructured interview increases the risk of overlooking important areas of inquiry.⁴⁶ Some even question

...THE USE OF DIAGNOSTIC CRITERIA for psychiatric disorders has been shown to increase the reliability of psychiatric diagnosis.

whether free-form interview is adequate for good clinical practice.⁴⁷

Clinician Factors—Clinician training, experience, and school of thoughts. Clinician background and training may influence interpretation of symptoms. One clinician with developmental training might explain the hallucinatory experience of the patient as part of posttraumatic experience of past abuse. Another clinician with a biomedical orientation might explain the same hallucinations as part of a schizophrenic process. This explains the tendency of some clinicians to overuse or underuse a particular diagnosis. The reliance on the patient's subjective symptoms, the clinician's interpretation of the symptoms, and the absence of objective measure (such as blood test) implant the seeds of diagnostic unreliability of psychiatric disorders.

Clinician factors—Reliance on observation. Use of clinical data derived from direct observation is a core component in any mental status examination. Nevertheless, there remains the potential that the standard methods employed by clinicians to solicit information may in fact bias patient data. One clinical area where data collection methods may unduly influence clinical findings is in the diagnostic assessment of dementia. Current diagnostic criteria require

documentation of a drop in functional status as well as decline in two or more cognitive domains.⁴⁸ Collateral data has often been used to satisfy this requirement; however, it too may be biased by various caregiver variables.⁴⁹ Consequently, performance-based measures of functional abilities have been developed to allow the clinician to directly observe the patient's ADL abilities. Reliance on these instruments of direct observation, however, introduces their own potential sources of systematic bias. Most notably is the increased likelihood of false negatives due to the analogue nature created by testing patient's daily living skills in a controlled setting.⁵⁰ Therefore, reliance on direct observation to the exclusion of other data sources (e.g., self-report, collateral) may, at least in some cases, further contribute to poorer diagnostic reliability.

Clinician factors—Psychiatric nomenclature. As mentioned previously, Ward, et al., found that inadequacy of the psychiatric nomenclature is the main reason for diagnostic unreliability (62.5%).³⁹ The dissatisfaction with the DSM-I and -II and ICD-9 sparked major efforts by prominent psychiatrists and researchers and eventually led to the publication of the DSM-III in 1980.

The initial work that led to the development of DSM-III goes back to the diagnostic criteria

developed at the Washington University School of Medicine in St. Louis. The criteria developed by the St. Louis group (often referred to as the "Feighner criteria") included the diagnostic criteria for 15 psychiatric conditions.⁵¹ Spitzer and others subsequently developed the Research Diagnostic Criteria (RDC), which is an expansion and modification of the Feighner criteria and included descriptions of 25 diagnostic categories.⁵² The Research Diagnostic Criteria (RDC) was developed for research purposes. The American Psychiatric Association (APA) Task Force on Nomenclature and Statistics recommended using diagnostic criteria and categories that can serve both research and clinical purposes. To meet this goal, the Research Diagnostic Criteria (RDC) was expanded and modified and resulted in the publication of DSM-III in 1980.²⁶ Due to its use of specific, clear, and detailed criteria for mental disorders, the DSM-III was accepted, preferred, and used worldwide over the ICD-9.²⁷

The development of the DSM-III and its subsequent versions has been a major accomplishment in the history of psychiatric nomenclature. Clinicians use the DSM criteria in clinical practice as an effective way to communicate the clinical picture, the course of illness, and efficacy of treatment. Psychiatric nomenclature is also used in teaching, research, and

legal and billing purposes. It is rare to see a clinician who does not use psychiatric nomenclature of DSM or ICD. The use of diagnostic criteria removed an important area of diagnostic unreliability.³⁷

However, the widespread use of diagnostic criteria does not mean by any measure that the status of psychiatric nomenclature is perfect or void of problems. In fact, the critics of current psychiatric nomenclature are extensive. First, the debate over the validity of DSM and ICD diagnoses is extensive and beyond the scope of this paper.^{15,53-57} The use of DSM in billing purposes has led some clinicians to choose diagnoses that are reimbursable by the insurance companies.^{44,45} Similarly, the diagnostic criteria of some disorders are still vague and difficult to operationalise. Finally, it is important to note that the diagnostic criteria do not eliminate the need for clinical experience.

PROPOSAL TO IMPROVE THE RELIABILITY OF PSYCHIATRIC DIAGNOSES

The authors propose the acronym 'DR.SED,' which stands for diagnostic criteria, reference definitions, structuring the interview, clinical experience, and data. The authors propose that using the DR.SED paradigm increases the reliability of psychiatric diagnosis.

Diagnostic criteria. The

absence of clear diagnostic criteria for psychiatric disorders has been known to be an important source of diagnostic unreliability for decades.³⁹ As discussed earlier in detail, the use of diagnostic criteria for psychiatric disorders has been shown to increase the reliability of psychiatric diagnosis.^{43,58}

Reference definition.

Reference definition refers to the definition of psychiatric symptoms and their levels of severity. A basic rule of reliability is to clearly define a symptom without ambiguity. Typically, many rating scales and structured interviews define the symptoms measured. The Schedules for Clinical Assessment in Neuropsychiatry (SCAN) system goes well beyond the existing structured interviews by publishing two separate books: the SCAN manual and the SCAN glossary, both of which are WHO documents.^{59,60} The SCAN manual contains all the questions covering different areas of psychopathology. For each question in the SCAN manual, the SCAN glossary contains a detailed definition of the symptom and reasons for excluding other possible symptoms. Even if clinicians do not agree on the definition of symptoms of the SCAN glossary, they can apply the SCAN definitions when using the SCAN interview. If clinicians agree on the presence or absence of symptoms, they are more likely to agree on the diagnosis. If

clinicians cannot agree on the presence or absence of symptoms, due to multiple definitions of a symptom, the diagnostic criteria are of little use.⁵⁸

Structuring the interview.

The skills required for interviewing patients are acquired over years of practice. As discussed previously, most clinicians use open-form interview in routine psychiatric evaluation and the reliability of this type of evaluation is low.⁷ Doubts about the adequacy of unstructured interviews have led to the development of more systematic instrument, such as rating scales and structured interviews.⁵⁸ For most clinicians, barriers exist that prevent them from using rating scales and structured interviews. Often clinicians have not been skilled in using structured instruments or feel that they are too time consuming and interfere with building therapeutic rapport with patients. Some rating scales such as the Positive and Negative Syndrome Scale (PANSS), Brief Psychiatric Rating Scale (BPRS), Hamilton Rating Scale for Depression (Ham-D), and others have become somewhat more widely used in clinical practice.⁶¹⁻⁶³ But, rating scales produce clusters of symptoms that do not necessarily take the entire clinical picture into account and have not been shown to have the same predictive utility as diagnoses.⁶⁴

The first author developed the Schedules for Clinicians'

THE SCIP TAKES A COMPARABLE amount of time as a psychiatrist normally needs for an open-form interview and does not require training.

APPENDIX A. SCIP screening questions

Codes: 0=absent, 1=present, 8=unsure, 9=missing data, unless otherwise specified in the question

Questions apply to the present episode, typically the past month, unless otherwise specified by the interviewer.

HAVE YOU:

1. Felt very anxious and afraid out of proportion to the situation (with or without physical symptoms) for more than one month?
2. Had panic attacks, when you suddenly felt anxious and frightened and developed physical symptoms, such as fast heart beat, shaking, or sweating?
3. Been afraid of going out of the house alone, traveling alone, being alone, being in crowds?
4. Been afraid and anxious doing things in front of people, such as eating in public, speaking in public?
5. Had unpleasant and unwanted thoughts or images coming into your mind over and over even if you try to get rid of them? Examples: Contamination or aggressive, sexual, or religious thoughts.
6. Had the urge to do things over and over and could not resist doing them (such as washing your hands even if they are clean, checking doors, counting up to certain numbers, reciting phrases)?
7. Witnessed or experienced a traumatic event that involved actual or threatened death or serious injury to you or someone else (e.g., physical or sexual abuse, terrorist attack, natural disaster, war)? Did you feel intense fear and helplessness?
8. Re-experience the traumatic event in the last month in a distressing way (flashback, nightmare)?
9. Had physical symptoms or physical illness for which doctors did all necessary work up and could not find medical explanation?
10. Had pain and your doctor did all necessary work up and could not really explain?
11. Worried about gaining weight to the point that you self-induced vomiting, or used diet pills, laxatives, or heavy exercise?
12. Eaten a large amount of food within an hour or so, that is binge eating?
13. Felt or described your mood as sad, downcast, gloomy, low in spirits, or depressed?
14. Been unable to enjoy things like walking, working at your hobbies, or socializing with friends as usual?
15. Had thoughts about harming yourself or even made an attempt at suicide (Include whether thought was due to depression or not)?
16. Felt very happy, elated without reason, or very irritable without reason?
17. Had mood swings: periods of depression and elation or irritability?
18. Felt that people are spying on you, follow you around, talk about you? Felt that there is a plot or conspiracy against you?
19. Felt that people are trying to harm you or poison your food?
20. Had experiences of hearing voices or noises that other people cannot hear?
21. Had experiences of seeing things (images, flashes, shadows, objects, people, whole scene) that other people cannot see?
22. Been violent in the past (with or without the influence of alcohol or drugs)?
23. I would like to ask you questions on alcohol use over the past year:
 - A. On days when you drank, did you drink >5 alcohol drinks per day (sometimes)?
 - B. Did you have any problems resulting from drinking alcohol?
24. I would like to ask you questions on illicit drug use (e.g. marijuana) over the past year:
 - A. Did you use the illicit drug >10 times per month?
 - B. Did you have any problems resulting from using the illicit drug?

Interviews in Psychiatry (SCIP) to facilitate the use of more systematic assessment by psychiatrists during their routine clinical practice.⁶⁵ The SCIP interview simulates a routine psychiatric interview in that it has two main phases: an initial open phase followed by a structured phase. During the initial phase, the patient is allowed to talk freely and explains why he or she is seeking professional help. The psychiatrist listens to the patient with a sense of warmth and empathy and builds a therapeutic rapport with the patient.⁶⁶ As the patient provides the history, the psychiatrist asks screening questions that cover different areas of psychopathology (e.g., anxiety, depression, psychosis, alcohol, and drug problems). Please refer to Appendix A for all 24 SCIP screening questions. At the conclusion of the initial phase, the psychiatrist should be able to hypothesize about possible diagnoses. The psychiatrist then starts the structured phase by utilizing disease-specific modules consisting of closed-ended questions. The structured phase of the SCIP has seven modules, covering the majority of Axis I disorders: Module A for anxiety and panic disorders, Module B for mood disorders, Module C for psychotic disorders, Module D for alcohol and psychoactive drugs disorders, Module E for somatoform disorders, Module F for eating disorders, and Module G for adjustment disorders. Each of the specific modules are available upon request from the first author. The SCIP takes a comparable amount of time as a psychiatrist normally needs for an open-form interview and does not require training. Preliminary analysis of SCIP data shows good inter-rater reliability.

Experience. Clinical experience can increase the reliability of psychiatric diagnosis in two ways. First, as the clinician

gains experience in different clinical settings and with different patient populations, the clinician is more likely to recognize and accurately diagnose various psychiatric disorders.

Consequently, the reliability of psychiatric diagnosis will increase. Second, the experience in using certain structured interviews increases the reliability of symptoms measurement and psychiatric diagnoses. Clinicians with similar training and experience tend to agree more on the diagnosis.⁶⁷

Data. Comprehensive data collection on the patient's clinical picture, history of episodes, obtaining the appropriate proxy information, and reviewing the old records helps to increase the

the reliability of psychiatric diagnosis is an important step toward validating the diagnostic categories of psychiatric disorders.

REFERENCES

1. Cook T, Campbell D. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally, 1979.
2. Carmines EG, Zeller RA. *Reliability and Validity Assessment*. London: SAGE, 1979.
3. WHO. *Manual of the International Classification of Diseases, Injuries and Causes of Death, Sixth Edition*. Geneva: 1948.
4. WHO. *The ICD-10 Classification of Mental and Behavioral Disorders: Diagnostic Criteria for Research, Tenth Edition*. Geneva: World Health Organization (WHO), 1993.
5. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, First Edition*. Washington, DC: American Psychiatric Press, Inc., 1952.
6. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revised*. Washington, DC: American Psychiatric Press, Inc., 2000.
7. Spitzer RL, Fleiss JL. A re-analysis of the reliability of psychiatric diagnosis. *Br J Psychiatry* 1974;125(0):341-7.
8. Sheehan D, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* 1998;59(20):22-33.
9. Spitzer RL, Williams JB, Gibbon M, First MB. The Structured Clinical Interview for DSM-III-R (SCID). I: History, rationale, and description. *Arch Gen Psychiatry* 1992;49(8):624-9.
10. Wing JK, Babor T, Brugha T, Burke J, et al. SCAN. Schedules for clinical assessment in neuropsychiatry. *Arch Gen Psychiatry* 1990;47(6):589-93.
11. Nurnberger JI Jr., Blehar MC, Kaufmann CA, et al. Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative. *Arch Gen Psychiatry* 1994;51(11):849-59.
12. Goodman AB, Rahav M, Popper M, et al. The reliability of psychiatric diagnosis in Israel's Psychiatric Case Register. *Acta Psychiatr Scand* 1984;69(5):391-7.
13. Kitamura T, Shima S, Sakio E, Kato M. Psychiatric diagnosis in Japan. 2. Reliability of conventional diagnosis and discrepancies with research diagnostic criteria diagnosis. *Psychopathology* 1989;22(5):250-9.
14. Dohrenwend BP, Dohrenwend BS. Perspectives on the past and future of psychiatric epidemiology. The 1981 Rema Lapouse Lecture. *Am J Public Health* 1982;72(11):1271-9.
15. Spitzer RL. Values and assumptions in the development of DSM-III and DSM-III-R: An insider's perspective and a belated response to Sadler, Hulgus, and Agich's "On values in recent American psychiatric classification." *J Nerv Ment Dis* 2001;189(6):351-9.
16. Grob G. The origins of American psychiatric epidemiology. *Am J Public Health* 1985;75:229-36.
17. Spitzer RL, Endicott J, Robins E. Clinical criteria for psychiatric diagnosis and DSM-III. *Am J Psychiatry* 1975;132(11):1187-92.
18. Masserman J, Carmichael H. Diagnosis and prognosis in psychiatry: With a follow-up study of the results of short-term general hospital therapy of psychiatric cases. *J Ment Sci* 1938;84:893-946.
19. Reich J, Black DW, Jarjous D. Architecture of research in psychiatry, 1953 to 1983. *Arch Gen Psychiatry* 1987;44(4):311-13.
20. Sandifer MG Jr., Pettus C, Quade D. A study of psychiatric diagnosis. *J Nerv Ment Dis* 1964;139:350-6.
21. Schmidt HO, Fonda CP. The reliability of psychiatric diagnosis: A new look. *J Abnorm Psychol* 1956;52(2):262-7.
22. Beck AT, Ward CH, Mendelson M, Mock JE, Erbaugh JK. Reliability of psychiatric diagnosis. 2. A study of consistency of clinical judgments and ratings. *Am J Psychiatry* 1962;119:351-7.
23. Kreitman N, Sainsbury P, Morrissey J, et al. The reliability of psychiatric assessment: An analysis. *J Ment Sci* 1961;107:887-908.
24. Mehlman B. The reliability of psychiatric diagnoses. *J Abnorm Psychol* 1952;47(2 Suppl):577-8.
25. Pasaminick B, Dinitz S, Lefton M. Psychiatric orientation and its relation to diagnosis and treatment in a mental hospital. *Am J Psychiatry*

IMPROVING THE RELIABILITY of psychiatric diagnosis is an important step toward validating the diagnostic categories of psychiatric disorders.

validity of psychiatric diagnosis.⁶⁸⁻⁷⁰ As mentioned earlier, a valid system is reliable. Consequently, comprehensive data gathering on the patient leads to reliable diagnosis.

CONCLUSIONS

The unreliability of psychiatric diagnosis has been and still is a major problem in psychiatry, especially at the clinician level. Clinicians need to make more efforts to improve the reliability of psychiatric diagnosis by using the 'DR.SED' paradigm (diagnostic criteria, reference definitions, structuring the interview, clinical experience, and data). Improving

- 1959;116:127–32.
26. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Third Edition*. Washington, DC: American Psychiatric Press, Inc., 1980.
 27. Helzer J, Canino G. Comparative analysis of alcoholism in 10 cultural regions. In: Helzer J, Canino G (eds). *Alcoholism in North America, Europe, and Asia*. New York: Oxford University Press, 1992.
 28. Gjaerum B, Bjornerem H. Psychosocial impairment is significant in young referred children with and without psychiatric diagnoses and cognitive delays—applicability and reliability of diagnoses in face of co-morbidity. *Eur Child Adolesc Psychiatry* 2003;12(5):239–48.
 29. Hickin N, Slate JR, Saarnio DA. Consistency in diagnoses for a sample of adolescents at a private psychiatric hospital. *Adolescence* 1996;31(123):553–9.
 30. Rey JM, Plapp JM, Stewart GW. Reliability of psychiatric diagnosis in referred adolescents. *J Child Psychol Psychiatry* 1989;30(6):879–88.
 31. Skovgaard AM, Isager T, Jorgensen OS. The reliability of child psychiatric diagnosis. A comparison among Danish child psychiatrists of traditional diagnoses and a multiaxial diagnostic system. *Acta Psychiatr Scand* 1988;77(4):469–76.
 32. Strober M, Green J, Carlson G. Reliability of psychiatric diagnosis in hospitalized adolescents. Interrater agreement using DSM-III. *Arch Gen Psychiatry* 1981;38(2):141–5.
 33. Tsang HY, Chong MY, Cheng AT. Development of the Chinese version of the geriatric mental state schedule. *Int Psychogeriatr* 2002;14(2):219–26.
 34. Weller EB, Weller RA, Fristad MA, et al. Children's Interview for Psychiatric Syndromes (ChIPS). *J Am Acad Child Adolesc Psychiatry* 2000;39(1):76–84.
 35. Ruskin PE, Reed S, Kumar R, et al. Reliability and acceptability of psychiatric diagnosis via telecommunication and audiovisual technology. *Psychiatr Serv* 1998;49(8):1086–8.
 36. Andreasen NC, Grove WM, Shapiro RW, et al. Reliability of lifetime diagnosis. A multicenter collaborative perspective. *Arch Gen Psychiatry* 1981;38(4):400–5.
 37. Grove WM, Andreasen NC, McDonald-Scott P, et al. Reliability studies of psychiatric diagnosis. Theory and practice. *Arch Gen Psychiatry* 1981;38(4):408–13.
 38. Corty E, Lehman AF, Myers CP. Influence of psychoactive substance use on the reliability of psychiatric diagnosis. *J Consult Clin Psychol* 1993;61(1):165–70.
 39. Ward CH, Beck AT, Mendelson M, et al. The psychiatric nomenclature. Reasons for diagnostic disagreement. *Arch Gen Psychiatry* 1962;7:198–205.
 40. Rankin E. Detection and assessment of caregiving adequacy in families with an Alzheimer's patient. *WV J Psychologic Res Prac* 1994;3:105–15.
 41. Oslin DW. Late-life alcoholism: Issues relevant to the geriatric psychiatrist. *Am J Geriatr Psychiatry* 2004;12(6):571–83.
 42. Strauss JS, Gabriel KR, Kokes RF, et al. Do psychiatric patients fit their diagnoses? Patterns of symptomatology as described with the biplot. *J Nerv Ment Dis* 1979;167(2):105–13.
 43. Helzer JE, Clayton PJ, Pambakian R, et al. Reliability of psychiatric diagnosis. II. The test/retest reliability of diagnostic classification. *Arch Gen Psychiatry* 1977; 4(2):136–41.
 44. Jensen AL, Weisz JR. Assessing match and mismatch between practitioner-generated and standardized interview-generated diagnoses for clinic-referred children and adolescents. *J Consult Clin Psychol* 2002;70(1):158–68.
 45. Pomerantz. Deliberate misdiagnoses of behavioral disorders: How widespread? *Behavioral Health Trends* 1996;20–30.
 46. Weitzel WD, Morgan DW, Cuyden TE, Robinson JA. Toward a more efficient mental status examination. Free-form or operationally defined. *Arch Gen Psychiatry* 1973;28(2):215–18.
 47. Eastwood MR, Ross HE. The reliability of the psychiatric anamnesis. *Br J Psychiatry* 1974;124(0):357–8.
 48. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*. Washington, DC: American Psychiatric Press Inc., 1994.
 49. Patterson MB, Mack JL, Neundorfer MM, et al. Assessment of functional ability in Alzheimer disease: A review and a preliminary report on the Cleveland Scale for Activities of Daily Living. *Alzheimer Dis Assoc Disord* 1992;6(3):145–63.
 50. Rankin E, Keefover R. Clinical cutoffs in screening functional performance for dementia. *J Clin Geropsychol* 1998;4(1):31–43.
 51. Feighner JP, Robins E, Guze SB, et al. Diagnostic criteria for use in psychiatric research. *Arch Gen Psychiatry* 1972;26(1):57–63.
 52. Spitzer RL, Endicott J, Robins E. Research diagnostic criteria: rationale and reliability. *Arch Gen Psychiatry* 1978;35(6):773–82.
 53. Richters JE, Hinshaw SP. The abduction of disorder in psychiatry. *J Abnorm Psychol* 1999;108(3):438–45.
 54. Sadler JZ, Hulgus YF, Agich GJ. On values in recent American psychiatric classification. *J Med Philos* 1994;19(3):261–77.
 55. Sadler JZ, Hulgus JF, Agich GJ. Hindsight, foresight, and having it both ways: A rejoinder to R. L. Spitzer. *J Nerv Ment Dis* 2001;189(8):493–7.
 56. Sedler MJ. Foundations of the new nosology. *J Med Philos* 1994;19(3):219–38.
 57. Wilson M. DSM-III and the transformation of American Psychiatry: A history. *Am J Psychiatry* 1993;150:399–410.
 58. Helzer JE, Robins LN, Taibleson M, et al. Reliability of psychiatric diagnosis. I. A methodological review. *Arch Gen Psychiatry* 1977; 34(2):129–33.
 59. World Health Organization. *Schedules for Clinical Assessment in Neuropsychiatry Manual, Second Edition*. Geneva: World Health Organization, 1994.
 60. World Health Organization. *Schedules for Clinical Assessment in Neuropsychiatry Glossary, Second Edition*. Geneva: World Health Organization, 1994.
 61. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960;23:56–62.
 62. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull* 1987;13(2):261–76.
 63. Overall JE, Gorham DR. The Brief Psychiatric Rating Scale (BPRS): Recent developments in ascertainment and scaling. *Psychopharmacol Bull* 1988;24:97–9.
 64. Wing JK, Birley JL, Cooper JE, et al. Reliability of a procedure for measuring and classifying "present psychiatric state." *Br J Psychiatry* 1967;113(498):499–515.
 65. Aboraya A, Tien A. Schedules for Clinicians' Interviews in Psychiatry (SCIP): Work in Progress. eCOMMUNITY. *Int J Mental Health Addict* 2004;November .
 66. Carlat DJ. *The Psychiatric Interview*. Philadelphia, PA: Lippincott Williams & Wilkins, 1999.
 67. Sandifer MG, Hordern A, Timbury GC, Green LM. Psychiatric diagnosis: A comparative study in North Carolina, London and Glasgow. *Br J Psychiatry* 1968;114(506):1–9.
 68. Basco M, Bostic JQ, Davies D, et al. Methods to improve diagnostic accuracy in a community mental health setting. *Am J Psychiatry* 2000;157(10):1599–605.
 69. Kranzler HR, Tennen H, Babor TF, et al. Validity of the longitudinal, expert, all data procedure for psychiatric diagnosis in patients with psychoactive substance use disorders. *Drug Alcohol Depend* 1997;45:93–104.
 70. Spitzer RL. Psychiatric diagnosis: Are clinicians still necessary? *Compr Psychiatry* 1983;24(5):399–411. ●