

Biological Markers and External Validators in Psychiatry: Progress Report on the Validity of Psychiatric Diagnoses

Ahmed Aboraya
West Virginia University, USA

Wilson Compton III
Washington University in St. Louis, USA

Published online November 7, 2004

Background. In 1970, Robins and Guze proposed five phases to achieve a valid classification of mental disorders: (a) clinical description, (b) laboratory study, (c) exclusion of other disorders, (d) follow-up study, and (e) family study. The objective of this paper is to review the progress made on content, criterion, and construct validity in relation to psychiatry. *Method.* The authors reviewed the literature addressing the concept of validity in psychiatry since 1970. *Results.* Much progress has been achieved with regard to content validity, as demonstrated by numerous publications regarding the diagnostic criteria of mental disorders. With regard to criterion validity, progress has been hampered by the lack of biological markers in psychiatry. Robins and Guze's (1970) phases have been the cornerstone of construct validity in psychiatry. *Conclusions.* The authors propose the acronym, DR. SEEK, which involves Data, Reference Definitions, Standardized Instruments, Clinical Experience, External Validators, and Knowledge. The authors recommend using the DR. SEEK standard to improve the accuracy of psychiatric diagnoses.

Keywords: Validity; Psychiatric diagnosis; Structured interview; Biological markers; External validators.

Introduction

Validity and reliability are two scientific topics vital to the development of modern psychiatry. *Reliability* refers to the extent to which an experiment, test, or any measuring procedure yields the same results on repeated trials (Carmines, 1979). *Validity* is a more confusing term, as it is defined differently in different contexts. In a very general sense, validity entails that scientific propositions describe and explain the empiri-

cal world in a correct way; in a stricter sense, that they are free from random and systematic errors (Swanborn, 1996). When applied to measuring instruments, validity refers to the extent that any measuring instrument measures what it intends to measure (Carmines, 1979).

There are three main types of validity: content, criterion and construct validity. *Content validity* refers to the degree to which an empirical measurement reflects a specific domain of content. An arithmetical operations test, for example, is content-valid if it includes addition, subtraction, multiplication, and division (Carmines, 1979). Typically, the items that represent the domain or disorder are derived from the consensus of experts. For example, experts in psychiatry agree that a patient with schizophrenia has symptoms of delusions, hallucinations, disorganization, and bizarre behavior.

Criterion validity refers to measuring a criterion, which is something that is external to the measuring instrument itself (Nunnally, 1978). Depending on the timing of the measurements, criterion validity can be

Ahmed Aboraya, Department of Behavioral Medicine and Psychiatry, West Virginia University, USA. Wilson Compton III, Department of Psychiatry, Washington University in St. Louis, USA.

Correspondence concerning this article should be addressed to Ahmed Aboraya, PO Box 9137, Department of Behavioral Medicine and Psychiatry, Robert C. Byrd Health Sciences Center of West Virginia University, Morgantown, WV 26506-9137. E-mail: aaboraya@hsc.wvu.edu.

postdictive, concurrent, or predictive. *Postdictive validity* entails correlating a criterion that occurred in the past (e.g., elementary-school grades) with some present measure (e.g., current high-school grades). *Concurrent validity* correlates the measuring instrument with some criterion measured at the same time (e.g., an abnormal glucose tolerance test is a criterion for diagnosing diabetes). *Predictive validity* correlates a measuring instrument with a criterion that will be assessed in the future (e.g., standardized college admission tests and average grades in college four years later).

Construct validity refers to the extent to which a particular measure relates to other measures consistent with theoretically derived hypotheses (Carmines, 1979). Typically, scientists formulate a hypothesis (the construct) that a variety of behaviors will correlate with one another. For example, the construct of the diagnosis of schizophrenia relies on the young age of the patient, the presence of psychosis, the absence of organic causes of psychosis, a positive family history of schizophrenia, and psychological test results that are consistent with schizophrenia. Construct validity is woven into the theoretical fabric of social sciences and psychiatry because of the absence of criterion validity (Carmines, 1979; Tsuang, 2002). Construct validity boils down to the circumstantial evidence for the usefulness of the construct or hypothesis under investigation (Nunnally, 1978).

In 1970, Robins and Guze proposed five phases to achieve a valid classification of mental disorders: (a) clinical description, (b) laboratory study, (c) exclusion of other disorders, (d) follow-up study, and (e) family study (Robins, 1970). Robins and Guze were actually the first investigators to articulate the elements of construct validity in psychiatry. They applied the criteria to schizophrenia and concluded that “good-prognosis schizophrenia” is not a mild schizophrenia, but rather a different illness. It is very important to note that construct validity is the product of clinical research and experience, as well as laboratory, epidemiological, and other research data. Construct validity requires a pattern of consistent findings across studies involving different samples and different settings (Tsuang, 2002).

The objectives of this paper are to review what has been learned since Robins and Guze’s (1970) influential article, as well as to propose a guide for clinicians to achieve more accurate diagnoses in psychiatry.

Method

Computerized literature searches were conducted using Medline and PsycINFO. Searches were conducted for entries dating from January, 1970 to December, 2003 that were published in English. Medline searches were conducted for entries that contained the words “validity” and “psychiatric diagnosis”; this

search yielded 93 citations. Similar searches using PsycINFO with identical search criteria yielded 154 citations. Thereafter, Medline searches were performed for entries containing the words “validity” and “structured interviews,” yielding 107 citations. Similar searches using PsycINFO with the same search criteria yielded 168 citations. Additionally, relevant references attached to published papers were also reviewed, while the authors identified further papers and books through consultations with colleagues and experts in the field. The authors were seeking new knowledge regarding the validity of psychiatric diagnoses.

Results

Progress on Content Validity

As described earlier, content validity is reflected by the consensus of experts in the field as to what constitutes the items of a domain. In the field of psychiatry, the nosology of mental illness has changed dramatically over the past century. During the first half of the twentieth century, little focus was placed on making psychiatric diagnoses. Beginning in the 1940s, clinicians began to diagnose psychiatric disorders. The American Psychiatric Association Committee on Nomenclature and Statistics developed and published in 1952 the first edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-I) (American Psychiatric Association, 1952).

The fourth edition of the DSM, published in 1994, includes the diagnostic criteria of all psychiatric disorders (American Psychiatric Association, 1994). The World Health Organization (WHO) also released several publications on the diagnostic criteria of mental disorders. The International Classification of Diseases—Tenth Edition (ICD-10) is the most recent of these efforts, having been published in 1992 (WHO, 1992). It is important to note that the debate over the validity of DSM and ICD diagnoses is extensive and beyond the scope of this paper (Richters, 1999; Sadler, 2001; Sadler, 1994; Sedler, 1994; Spitzer, 2001; Wilson, 1993). The criteria proposed in the DSM and ICD originated from the best possible consensus of experts in the field. The widespread adoption of the DSM and ICD diagnostic criteria shifted the focus to *procedural validity*, which assesses the extent to which a new diagnostic procedure measures the DSM or ICD criteria (Spitzer, 1980).

Progress on Criterion Validity: Decades of Research on Biological Markers

In criterion validity, there is a criterion against which the measuring instrument or the disease entity is validated. In psychiatry, the search for promising criteria or biological markers has been going on for decades.

Buchsbaum (1983) proposed a definition of a *biological marker* as a measurable indicator of a disease that may or may not be causal. Several biological markers have been studied in psychiatry, including platelet monoamine oxidase (MAO); results of dexamethasone suppression tests; and metabolites of serotonin and noradrenaline in cerebral and spinal fluid. Decades of research on biological markers have resulted in some promising results; however, no single biological marker has been unequivocally identified as a marker for mental disorders (Buchsbaum, 1983; Hoes, 1986; Jablensky, 1984; Muscettola, 1984).

In light of the serious questions regarding the validity of DSM and ICD definitions, and the absence of biological markers for mental disorders, Spitzer proposed the *LEAD* standard (Spitzer, 1983). LEAD is an acronym that stands for the Longitudinal evaluation performed by Expert clinicians who utilize All the Data available. The LEAD standard is an important step toward obtaining the best-estimate diagnosis by requiring expert clinicians to utilize all the available data over time, including information from family members, hospital records, psychological evaluations, and laboratory results. The requirement of LEAD to have expert clinicians make independent assessments, discuss diagnostic disagreements, and reach a consensus diagnosis accounts for the difficulty in implementing the LEAD standard, and its consequent limited use (Antony, 2002; First, 1996; Spitzer, 1983).

Progress on Construct Validity: Recent Advances in Neuroimaging and the Neurosciences

External validators are elements external to the disease definitions; these are not restricted to biological markers. The past decade has witnessed an explosion in brain-imaging techniques that have permitted scientists to study brain structures and function, even at cellular and molecular levels. Structural magnetic-resonance imaging (MRI), functional MRI (fMRI), magnetic resonance spectroscopy (MRS), single-proton emission-computed tomography (SPECT), and positron-emission tomography (PET) are some of these new brain imaging techniques. Andreasen (1995) has used the phrase “new external validators” when describing these neuroimaging techniques and other new branches of neuroscience, and has emphasized their importance in understanding the relationship between individual symptoms and the changes in brain structure, brain function, or both.

Three of Robins and Guze’s (1970) criteria described earlier are relevant to external validators: laboratory studies, follow-up studies, and familial aggregation studies. Although these new techniques have yielded important research findings, these findings cannot yet be translated into clinical practice (George, 1998). However, the new external validators hold

promise in validating psychiatric diagnoses and predicting treatment responses in psychiatry in the near future (Nemeroff, 2001).

Discussion

Definition of a Validity Criterion

The authors define a validity criterion as any knowledge, method (e.g., rating scale or structured interview), or procedure (e.g., blood test, lumbar puncture, or MRI) that can (a) improve the accuracy of the disease or disorder measurement; (b) help to rule out alternative diseases or disorders in the differential diagnosis; or (c) validate a provisional diagnosis of the disease or disorder. Validity is a relative phenomenon, as any knowledge provided by validity criteria helps scientists move closer to understanding the true nature of phenomena and classifications.

As Nunnally (1978) stated, “validity usually is a matter of degree rather than an all-or-none property, and validation is an unending process.” In medicine and psychiatry, clinicians should use all the available validity criteria to obtain the most accurate diagnosis. The more validity criteria are utilized, the more accurate the diagnosis. The clinician who uses his or her clinical skills along with structured interviews will provide a more accurate diagnosis in comparison with the clinician who relies on clinical skills alone. Similarly, the neurologist who uses clinical examinations, lumbar punctures, and MRIs can provide a more accurate diagnosis of multiple sclerosis than can the neurologist who relies on clinical examination alone.

Proposal to Improve the Accuracy of Psychiatric Diagnoses

The most accurate psychiatric diagnoses can be obtained by well-trained and experienced mental-health clinicians who use reliable, comprehensive, and standardized or semi-standardized instruments, while using all available data and external validators, and while relying on reference definitions of psychiatric symptoms. To improve the accuracy of making psychiatric diagnoses, we propose the acronym DR. SEEK, which involves Data, Reference definitions, Standardized instruments, clinical Experience, External validators, and Knowledge. It is important to note that these validity criteria and their specifications are merely proposed guidelines based upon previous research and our own experiences in clinical care and research. The proposed criteria can be further modified and enhanced by other experts in the field.

Data

Mental-health clinicians should obtain and use all data essential for an accurate diagnosis: patients' clinical descriptions, histories, and courses of illness; family information; family histories; psychological testing; and any other pertinent data to the case. The use of all the available data can improve diagnostic validity (Basco, 2000; Kranzler, 1997; Spitzer, 1983).

Reference Guide

Psychiatric symptoms reported by the patient and observed by the clinician constitute the main source of information used by clinicians when diagnosing and treating patients (Aboraya, 2000). In other words, the measurement of psychiatric symptoms continues to be the main source when assessing whether the diagnostic criteria of a disorder are met. Because of the importance of symptoms, mental-health clinicians should clearly define the meaning of symptoms and their level of severity. Clinicians should also define the meaning of their inquiries to their patients. Additionally, it is important to understand the expression of symptoms in different cultures. Structured interviews can help to both define the meaning of terminology and differentiate the levels of symptom severity (Spitzer, 1992; Wing, 1990).

Standardization

The use of standardized or semi-standardized instruments aids the clinician in many ways. First, standardization forces the clinician to cover all areas of the psychopathology under question. Second, standardization provides similarities in the manner in which questions are asked, thereby minimizing variability among clinicians. Standardization is a way of transforming the clinical judgment into explicit rules to prevent the unknown and subjective decisions of individual interviewers (Lewis, 1992). The use of structured interviews can provide a more accurate diagnosis in comparison with routine clinical diagnoses (Basco, 2000; Kranzler, 1997). For instance, upon comparing the validity of LEAD, structured interviews, and clinician diagnoses, Kranzler (1997) concluded that the use of structured interviews can enhance the diagnostic validity of the LEAD standard.

Experience

Mental-health clinicians may be psychiatrists, clinical psychologists, therapists, or other professionals who have actual experience and contact with individuals with mental disorders. However, the most important factor contributing to clinicians' ability to effectively diagnose patients is clinicians' experience. Clinical experience with psychiatric abnormalities, and the

development of skills to elicit them and to ascertain their significance, have been viewed as the reference standards of psychopathology assessments (Brugha, 1999). As reading textbooks of medicine alone does not qualify the reader to diagnose medical disorders, reading textbooks of psychiatry does not qualify the reader to diagnose mental disorders. Clinical experience is thus indispensable when it comes to diagnoses of psychotic, bipolar, or personality disorders, especially in regard to judging the significance of symptoms (Frances, 1998; Spitzer, 1998). Such clinical experience cannot be replaced with a few weeks of training, nor is it exclusive to psychiatrists. Clinicians should have at least two years of experience encompassing various diagnostic entities, such as anxiety disorders, mood disorders, psychotic disorders, and substance-use disorders. A minimum of six months' experience with inpatients is strongly advocated in order to become familiar with severe instances of various psychopathologies. The first author's personal experience of five years of psychiatric research preceding the clinical experience of psychiatry residency forms the basis for this recommendation.

External Validators

External validators, including novel neuroimaging techniques and other branches of neuroscience, advance everyday knowledge and hold the promise of validating psychiatric disorders in the future (Andreasen, 1995; George, 1998; Nemeroff, 2001). Although the current available techniques are not useful as indicators of the presence of psychiatric disorders, they are useful in ruling out alternative disorders in the differential diagnosis. For example, clinicians may use computed tomography (CT) or MRI scans to rule out head traumas, strokes, or multiple sclerosis as the cause of depressive or psychotic symptoms. As another example, thyroid hormone levels are sometimes measured to exclude hypothyroidism or hyperthyroidism as the cause of anxiety or depression.

Knowledge

Obviously, mental-health clinicians should have extensive knowledge about mental health. A minimum of a Master's degree in a mental-health field, such as psychology, with a clinical emphasis, or a medical degree with psychiatry residency training, is recommended.

Conclusions

Much progress has been achieved since Robins and Guze's influential article in 1970. Progress on the concept of content validity was crowned by the worldwide adoption of DSM and ICD criteria in clinical and research applications. The lack of biological markers that validate psychiatric disorders points to the importance

of construct validity in psychiatry. Psychiatric researchers should utilize all the available validity criteria and new external validators to refine and validate the diagnostic entities of the DSM and ICD systems.

Acknowledgements

Dr. Compton is currently at the National Institute on Drug Abuse in Bethesda, Maryland. This paper is based on work completed while Dr. Compton was at the Washington University in St. Louis, Missouri, prior to joining the National Institute on Drug Abuse (NIDA). None of the views expressed in this paper necessarily represent the views of the National Institute on Drug Abuse, the National Institutes of Health (NIH), or the federal government.

References

- Aboraya, A. (2000). The psychiatric interview: Art and science. *Psychiatry Update*, 1, 165-168.
- American Psychiatric Association. (1952). *Diagnostic and statistical manual of mental disorders* (1st ed.). Washington, DC: Author.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Andreasen, N. C. (1995). The validation of psychiatric diagnosis: New models and approaches. *American Journal of Psychiatry*, 152(2), 161-162.
- Antony, M. M., & Barlow, D. H. (2002). Structured and semistructured diagnostic interviews. In D. H. Barlow (Ed.), *Handbook of assessment and treatment planning for psychological disorders* (pp. 3-37). New York: Guilford Publications Inc.
- Basco, M., Bostic, J. Q., Davies, D., Rush, A. J., Witte, B., Hendrickse, W., & Barnett, V. (2000). Methods to improve diagnostic accuracy in a community mental health setting. *American Journal of Psychiatry*, 157(10), 1599-1605.
- Brugha, T., Bebbington, P. E., & Jenkins, R. (1999). A difference that matters: Comparisons of structured and semi-structured psychiatric diagnostic interviews in the general population. *Psychological Medicine*, 29(5), 1013-1020.
- Buchsbaum, M. S., & Haier, R. J. (1983). Psychopathology: Biological approaches. *Annual Review of Psychology*, 34, 401-430.
- Carmine, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. London: SAGE.
- First, M. B., Gibbon, M., Spitzer, R. L., & Williams, J. B. W. (1996). *User's guide for the Structured Clinical Interview for DSM-IV Axis I disorders: Research version (SCID-I)*. New York: Biometrics Research.
- Frances, A. (1998). Problems in defining clinical significance in epidemiological studies. *Archives of General Psychiatry*, 55(2), 119.
- George, M. S. (1998). Advances in brain imaging: An overview of what the primary psychiatrist needs to know. *Primary Psychiatry*, 5(3), 37-45.
- Hoes, M. J. (1986). Biological markers in psychiatry. *Acta Psychiatrica Belgica*, 86(3) 220-241.
- Jablensky, A. (1984). Epidemiological and clinical research as a guide in the search for risk factors and biological markers. *Journal of Psychiatric Research*, 18(4), 541-556.
- Kranzler, H. R., Tennen, H., Babor, T. F., Kadden, R. M., & Rounsaville, B. J. (1997). Validity of the longitudinal, expert, all data procedure for psychiatric diagnosis in patients with psychoactive substance use disorders. *Drug & Alcohol Dependence*, 45, 93-104.
- Lewis, G., Pelosi, A. J., Araya, R., & Dunn, G. (1992). Measuring psychiatric disorder in the community: A standardized assessment for use by lay interviewers. *Psychological Medicine*, 22(2), 465-486.
- Muscettola, G., Di Lauro, A., & Giannini, C. P. (1984). Blood cells as biological trait markers in affective disorders. *Journal of Psychiatric Research*, 18(4), 447-456.
- Nemeroff, C. (2001). The development of new technologies in medicine: What will the impact be in psychiatry? *CNS Spectrums*, 6(1), 21.
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
- Richters, J. E., & Hinshaw, S. P. (1999). The abduction of disorder in psychiatry. *Journal of Abnormal Psychology*, 108(3), 438-445.
- Robins, E., & Guze, S. B. (1970). Establishment of diagnostic validity in psychiatric illness: its application to schizophrenia. *American Journal of Psychiatry*, 126(7), 983-987.
- Sadler, J. Z., Hulgus, J. F., & Agich, G. J. (2001). Hindsight, foresight, and having it both ways: A rejoinder to R. L. Spitzer. *Journal of Nervous and Mental Disease*, 189(8), 493-497.
- Sadler, J. Z., Hulgus, Y. F., & Agich, G. J. (1994). On values in recent American psychiatric classification. *Journal of Medicine and Philosophy*, 19(3), 261-277.
- Sedler, M. J. (1994). Foundations of the new nosology. *Journal of Medicine and Philosophy*, 19(3), 219-238.
- Spitzer, R. L. (1983). Psychiatric diagnosis: are clinicians still necessary? *Comprehensive Psychiatry*, 24(5), 399-411.
- Spitzer, R. L. (1998). Diagnosis and need for treatment are not the same. *Archives of General Psychiatry*, 55(2), 120.
- Spitzer, R. L. (2001). 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". *Journal of Nervous and Mental Disorders*, 189(6), 351-359.
- Spitzer, R. L., & Williams, J. (1980). Classification of mental disorders and DSM-III. In H. Kaplan, A. Freedman, and B. Sadock (Eds.), *Comprehensive textbook of psychiatry*. Baltimore, MA: Williams & Wilkins.
- Spitzer, R. L., Williams, J. B., Gibbon, M., & First, M. B. (1992). The Structured Clinical Interview for DSM-III-R (SCID). I: History, rationale, and description. *Archives of General Psychiatry*, 49(8), 624-629.
- Swanborn, P. G. (1996). A common base for quality control criteria in quantitative and qualitative research. *Quality and Quantity*, 30, 19-35.
- Tsuang, M. T., & Tohen, M. (2002). *Textbook in psychiatric epidemiology*. New York: Wiley-Liss.
- World Health Organization. (1992). *The ICD-10 classification of mental and behavioral disorders: Clinical descriptions and diagnostic guidelines* (10th ed.). Geneva, Switzerland: Author.

Wilson, M. (1993). DSM-III and the transformation of American Psychiatry: A history. *American Journal of Psychiatry*, 150, 399-410.

Wing, J. K., Babor, T., Brugha, T., Burke, J., Cooper, J. E., Giel, R., et al. (1990). SCAN. Schedules for Clinical Assessment in Neuropsychiatry. *Archives of General Psychiatry*, 47(6), 589-593.