

# How “Objective” Are Psychiatric Diagnoses? (Guess Again)

by **RONALD PIES, MD**

**AUTHOR AFFILIATIONS:** Dr. Pies is Professor of Psychiatry, SUNY Upstate Medical University, Syracuse, New York; and Clinical Professor of Psychiatry, Tufts USM, Boston, Massachusetts.

**ADDRESS CORRESPONDENCE TO:** Dr. Ronald Pies, Box 332, Bedford MA 01730;  
E-mail: ronpies@massmed.org

**KEY WORDS:** kappa, interrater reliability, objectivity, diagnosis



## ABSTRACT

Critics of psychiatry often argue that psychiatric diagnosis lacks “objectivity,” particularly when compared with diagnosis in other medical specialties. However, when one examines interrater reliability—an important component of objectivity—the kappa values for several major psychiatric disorders are generally on a par with those in other medical specialties. Nonetheless, in psychiatry as in all of general medicine, there is an irreducible element of the subjective. That is part of the “art” of medical and psychiatric practice.

## INTRODUCTION

You know the scoop on psychiatry and “objective” data. You have heard the same critique many times from the media, certain movie stars, and even some of our medical colleagues: Psychiatrists have no “objective” tests or criteria for disease; therefore, they cannot provide objective diagnoses. Consider this unsettling state of affairs:

*“...there is frequently no one best doctor and no one best treatment,” said Dr. John H. Glick of...the University of Pennsylvania. When patients consult him for second opinions or to transfer their care to his center, Dr. Glick estimated that he and his colleagues concur completely with the original doctor in about 30 percent of cases. But in another 30 to 40 percent of cases, they recommend major changes in the treatment plan, like a totally different...[medication] regimen...Sometimes his team makes a completely different diagnosis.”<sup>1</sup>*

Sound all too familiar? Well, what you might not know (since I omitted a few key words) is that Dr. Glick was describing his own experience in the field of cancer treatment—that good, solid, “objective” science of oncology. (Dr. Glick is with the Abramson Cancer Center in Philadelphia).

Indeed, the notion that when an oncologist or a pathologist looks at a group of cells under the microscope, he or she is making a purely objective assessment turns out to be a gross oversimplification if, by *objective*, we mean *totally free from any subjective judgment, bias, or tendency based on personal experience*. Of course, this is a procrustean definition of objective. If applied uniformly throughout the medical disciplines, it is doubtful any specialty would survive as an objective science. On the other hand, if we employ a more philosophically grounded definition of objective, we find that psychiatric diagnosis is often as objective as that in most other medical specialties.

### SO WHAT IS OBJECTIVITY?

As I have pointed out elsewhere,<sup>2</sup> the philosopher Amartya Sen described two essential features of objectivity: *observation dependence* and *impersonality*.<sup>3</sup> The first term

implies that empirical observation is a central feature of any objective science. As physicians, we look, listen, poke, prod, and measure as part of our daily work. As psychiatrists, we carry out detailed mental status exams; perform limited neurological exams; interview family members; and order a variety of ancillary studies. The second term, *impersonality*, implies that in order for an observation to be objective, the observer's conclusions should be more or less reproducible by other observers, within the natural limits of human perception. I believe that psychiatry meets both of Sen's criteria for objectivity,<sup>2</sup> and that with respect to interrater reliability we actually do about as well as most other medical specialties.

Interrater (or interobserver) reliability is usually expressed in terms of a function called kappa. Essentially, Cohen's kappa was designed to estimate the degree of consensus between two judges after

correcting for the amount of agreement that could be expected by chance alone.<sup>4</sup> Statisticians and researchers have pointed out several limitations associated with kappa,<sup>5</sup> but most biomedical and behavioral research continues to use this as a measure of interobserver agreement. Basically, the higher the kappa, the more reliable the observation is considered. Kappa values from 0.41 to 0.60 are usually considered moderate, and values above 0.60 are substantial.<sup>5</sup> A value of zero on kappa indicates that two observers did not agree with each other any more than would be predicted by chance alone.

With this prologue in mind, how does psychiatry fare when compared with interrater reliability in some other medical specialties? One might assume, for example, that two pathologists looking at the same tissue specimen under the microscope would have a fairly high kappa, say greater than 0.60. On the other hand, everybody knows that if you put two psychiatrists in a room, they walk out with three opinions, right? Well, maybe not.

In one study of interobserver variability, pathologists were asked to examine histological specimens of malignant mesothelioma using several different techniques.<sup>6</sup> It turned out that "...most indexes of agreement between pathologists ranged from poor (needle biopsy) to moderate (necropsy/surgery)." In fact, agreement regarding material obtained by needle biopsy had a median kappa of only 0.21, not much better than chance. For tissue obtained at necropsy or surgery, the agreement was only moderately good, with a median kappa of 0.57. Similarly, in a study of cytologic interpretation of epithelial cells<sup>7</sup> (i.e., were the cells mildly atypical, markedly atypical, or malignant?), the kappa for interrater agreement was only 0.46. This is arguably not a stellar showing, considering that cytology is often held up as an exemplar of objective science.

How about a medical specialty more closely related to psychiatry? How, for example, do our colleagues

**TABLE 1.** Interrater reliability (kappas) for four psychiatric diagnoses and diagnoses from other medical specialties

PSYCHIATRIC DIAGNOSES <sup>10</sup>	KAPPAS FOR IN-PERSON INTERVIEWS
Major depression	0.73
Alcohol dependence	0.86
Bipolar disorder	0.76
Panic disorder	1
Average kappa = 0.83	
MEDICAL/NEUROLOGICAL DIAGNOSES	KAPPAS
Ischemic stroke <sup>8</sup>	0.53 [average]
Colorectal adenocarcinoma <sup>9</sup>	0.78
Renal stenosis <sup>11</sup>	0.43
Knee osteoarthritis <sup>12</sup>	0.1
Breast cancer (mammographic density measurements) <sup>13</sup>	0.89
Average kappa = 0.55	

NOTE: values above 0.60 indicate substantial interrater reliability

There is still a great deal of “art” in our field, and this is nothing for which we should apologize. To be sure, some of the “data” psychiatrists evaluate differ from those scrutinized by our colleagues in pathology or internal medicine...But in the end, all physicians must make difficult judgments based on imperfect knowledge.

in neurology do, when it comes to interrater agreement? In one study of ischemic stroke, 160 cases were reviewed by three pairs of board-certified neurologists “with a special interest in stroke.”<sup>8</sup> The kappa was only “fair to good” in most categories of ischemic stroke, with a high of 0.70 for oral contraceptive-related stroke and a low of 0.28 for lacunar infarcts; the average kappa for all categories was 0.53—at best, a moderate level of agreement.

These selective results, to be sure, represent only a few studies that I chose as demonstration cases. One can certainly find higher kappas for some medical or neurological disorders. For example, pathological diagnosis of colorectal cancer appears to have very robust inter-rater reliability, with a kappa of 0.78.<sup>9</sup> But as a rough generalization, I believe that interrater reliability is no higher in many nonpsychiatric medical specialties than in psychiatry.

#### **FOUR COMMON PSYCHIATRIC DISORDERS**

Consider the intriguing study by Ruskin and colleagues.<sup>10</sup> Two trained interviewers each interviewed the same 30 psychiatric inpatients using the Structured Clinical Interview for DSM-III-R. Fifteen subjects had two in-person interviews, while 15 subjects had one in-person and one telecommunication (videoconferencing) interview. Interrater reliability was calculated for the four most common diagnoses: major depression, bipolar disorder, panic disorder, and alcohol dependence. For each diagnosis, interrater reliability was identical or nearly so for the patients who had standard, in-person interviews and those who had an in-person and a

remote interview. But more important for our purposes, the kappas for the regular interviews were impressively high. Indeed, the average kappa for these four psychiatric diagnoses (0.83) exceeds that reported (0.78) for the diagnosis of colorectal cancer.<sup>9</sup> As shown in Table 1, the kappa for these major psychiatric disorders compares favorably with kappas in several other medical specialties.<sup>8-13</sup>

To be sure, kappa is not the alpha and omega of psychiatric diagnosis. Nor is the foregoing discussion intended as a ringing endorsement of the DSM system of categorical classification. Many critiques have pointed out deep-seated problems with the DSM’s diagnostic “pigeonholing,” and many reforms, such as a diagnostic system based on dimensions of pathology, have been proposed.<sup>14</sup> My point is simply that flawed as it is psychiatric diagnosis often yields at least as much interrater reliability as seen in other medical specialties. That said, there are some psychiatric diagnoses, such as schizoaffective disorder, that appear to have quite low kappas<sup>15</sup>—a fact that will not surprise many psychiatrists.

#### **PSYCHIATRIC DIAGNOSIS IN PERSPECTIVE**

Recently, this writer had the experience of being diagnosed with laryngoesophageal reflux (LER), proffered as an explanation for some chronic throat discomfort. The ear-nose-throat (ENT) doctor—a highly-regarded specialist—offered this diagnosis with great confidence and prescribed high doses of a proton-pump inhibitor. The treatment helped. But upon some further investigation, I learned that LER is a rather controversial entity. One

gastroenterologist (I sense there is some tension between the GI and the ENT specialists) opined that many cases of so-called LER can be explained on the basis of “...voice abuse, smoking, repetitive throat clearing, asthma...and postnasal drip.”<sup>16</sup> The beginnings, perhaps, of the Myth of LER?

It is also instructive to ask how much physico-anatomical evidence supports the “reality” of several frequently diagnosed medical conditions, such as fibromyalgia, chronic fatigue, and restless legs syndromes, compared with the neuropathological evidence supporting the existence of, say, schizophrenia and bipolar disorder.<sup>17,18</sup> After all, if psychiatry, among all the medical specialties, is to be singled out for criticism (if not abuse), it may be relevant to ask if this antipathy represents, well, “fair” treatment. Though a review of these disorders is beyond the scope of this commentary, my literature search reveals that in none of these three—fibromyalgia, chronic fatigue, or restless legs syndrome—is there any consistently identified lesion, cellular pathology, or pathophysiologic explanation, notwithstanding a profusion of criteria and hypotheses.<sup>19-21</sup> This fact does not render these disorders myths. They have legitimate, if provisional, places in medical nosology because these syndromes are associated with suffering and incapacity in our patients. And it is suffering and incapacity, as I have argued elsewhere,<sup>22</sup> that compose the essence of the disease concept. (Psychiatrists may be ruefully amused at this observation by my neurologist colleague, John Winkelman, MD, on restless legs



syndrome: “Critics do not regard RLS as a disorder at all, but, rather, as the fabrication of an omnivorous pharmaceutical industry.”<sup>21</sup>

## CONCLUSION

I believe there is as much objective science in psychiatry as there is in most other medical specialties, which is to say an impressive but not overwhelming amount. This acknowledgment does not disparage psychiatry or the other medical specialties. It simply affirms what most good physicians know from long, hard experience: There is still a great deal of “art” in our field, and this is nothing for which we should apologize. To be sure, some of the “data” psychiatrists evaluate differ from those scrutinized by our colleagues in pathology or internal medicine; thought processes, after all, are not leukocytes. But in the end, all physicians must make difficult judgments based on imperfect knowledge. Objectivity in medicine, vital though it is, is merely a means to an end: the relief of human suffering and the promotion of wellbeing.

## REFERENCES

1. Grady D. Cancer patients lost in a maze of uneven care. *New York Times* July 29, 2007.
2. Pies R. Psychiatry clearly meets the ‘objectivity’ test. *Psychiatr News* 2005;40(19):17.
3. Sen A. Objectivity and position. Accessed at: [www.globalhealth.harvard.edu/hcps/wpweb/90\\_01.pdf](http://www.globalhealth.harvard.edu/hcps/wpweb/90_01.pdf). Access date: September 28, 2007.
4. Cohen J. A coefficient for agreement for nominal scales. *Educ Psychol Meas* 1960;20:37–46.
5. Stemler SE. A comparison of consensus, consistency, and measurement: Approaches to estimating interrater reliability. Accessed at: <http://pareonline.net/getvn.asp?v=9&n=4>. Access date: September 28, 2007.
6. Andrión A, Magnani C, Betta PG, et al. Malignant mesothelioma of the pleura: interobserver variability. *J Clin Pathol* 1995;48:856–60.
7. Visvanathan K, Santor D, Ali SZ et al. The importance of cytologic intrarater and interrater reproducibility: The case of ductal lavage. *Cancer Epidemiol Biomarkers Prev* 2006;15:2553–6.
8. Johnson CJ, Kittner SJ, McCarter RJ, et al. Interrater reliability of an etiologic classification of ischemic stroke. *Stroke* 1995;26:46–51.
9. Vobecky J, Leduc CP, Devroede G et al. The reliability of routine pathologic diagnosis of colorectal adenocarcinoma. *Cancer* 2006;64:1261–5.
10. Ruskin PE, Reed S, Kumar R et al. Reliability and acceptability of psychiatric diagnosis via telecommunication and audiovisual technology. *Psychiatr Serv* 1998;49:1086–8.
11. Schreij G, de Haan MW, Oei TK, et al. Interpretation of renal angiography by radiologists. *J Hypertens* 1999;17(12 Pt 1):1737–41.
12. Weidow J, Cederlund CG, Ranstam J, et al. Ahlbäck grading of osteoarthritis of the knee: Poor reproducibility and validity based on visual inspection of the joint. *Acta Orthop* 2006;77:262–6.
13. Gao J, Warren R, Warren-Forward H, et al. Reproducibility of visual assessment on mammographic density. *Breast Cancer Res Treat* 2007 Jul 7; [Epub ahead of print].
14. van Praag HM. “Make Believes” in *Psychiatry or the Perils of Progress*. London: Brunner-Mazel, 1992.
15. Maj M, Pirozzi R, Formicola AM, et al. Reliability and validity of the DSM-IV diagnostic category of schizoaffective disorder: Preliminary data. *J Affect Disord* 2000;57:95–8.
16. Johnson DA. What is the prevalence of GERD: Signs in the laryngopharyngeal area during routine upper gastrointestinal endoscopy? *Medscape Gastroenterology* (posted 7/12/07). Accessed at: [www.medscape.com/viewarticle/557597](http://www.medscape.com/viewarticle/557597). Access date: September 28, 2007.
17. Vita A, De Peri L, Silenzi C, et al. Brain morphology in first-episode schizophrenia: a meta-analysis of quantitative magnetic resonance imaging studies. *Schizophr Res* 2006;15:75–88. Epub 2005 Dec 27.
18. Yildiz-Yesiloglu A, Ankerst DP. Neurochemical alterations of the brain in bipolar disorder and their implications for pathophysiology: A systematic review of the *in-vivo* proton magnetic resonance spectroscopy findings. *Prog Neuropsychopharmacol Biol Psychiatry* 2006;30:969–95. Epub 2006 May 4.
19. Mease P. Fibromyalgia syndrome: Review of clinical presentation, pathogenesis, outcome measures, and treatment. *J Rheumatol Suppl* 2005;75:6–21.
20. Wyller VB. The chronic fatigue syndrome: An update. *Acta Neurol Scand Suppl* 2007;187:7–14.
21. Winkelman JW. Periodic limb movements in sleep: Endophenotype for restless legs syndrome? *N Engl J Med* 2007;357:703–6.
22. Pies R. Moving beyond the “myth” of mental illness. In: Schaler JA (ed). *Szasz Under Fire*. Chicago, IL: Open Court, 2004:327–53. ●