Helio G. Rocha Neto[1,2]
https://orcid.org/0000-0002-7342-0634

Maria Tavares Cavalcanti[2,3]
https://orcid.org/0000-0003-1872-4210

Diogo Telles Correia[4]
https://orcid.org/0000-0003-4217-2823

Diagnostic unreliability between research and clinical practice in psychiatry still matters: a call for discussion about medical history taking and diagnostic interview basic principles

DOI: 10.1590/0047-2085000000419

Reliability and Validity are closely related concepts in philosophy and medicine. Validity concerns the existence of a specific concept or object in a shared reality, while reliability relates to the agreement among different observers regarding the existence of a concept or object[1]. Both validity and reliability are fundamental to the issue of mental disorders and psychiatry's goal of being a science-based medical specialty[1].

Mental disorders encompass biological, subjective, and social aspects of human life[2]. Despite anti-psychiatry movement critics, many of these disorders exist as independent constructs and are therefore valid. However, the low reliability among clinicians indicates limited validity of mental disorders. To address this, psychiatry introduced the "operational revolution," which involves describing mental disorders through operational categories and using Structured Diagnostic Interviews (SDIs) as a guide for diagnosis[3].

Operational categories undergo continuous review by the DSM and ICD, but it's unclear how they are used in daily clinical practice[4,5]. On the other hand, SDIs are rarely used in clinical practice, leading to an unspoken problem in evidence-based psychiatry. Research relies on subjects diagnosed using operational criteria obtained through SDIs, while clinical practice relies on individual diagnostic prototypes obtained through Non-Standard Diagnostic Interviews (NSDIs) that lack standardization[6].

Surprisingly, there are very few studies measuring the reliability between SDIs and NSDIs, and almost none focusing on NSDI reliability since the development of SDIs in the late seventies and early eighties[7]. This scarcity suggests that NSDI unreliability is now taken for granted or that reliability issues are considered irrelevant. The latter hypothesis is reinforced by the DSM-5 work group's goal of achieving kappa reliability of around 0.4 for diagnostic items[8], a value only slightly better than random agreement[9], and worse than NSDI reliability studies in the pre-operational revolution era[10].

The problem of conducting research with a definition of mental disorders and a diagnostic instrument that differ from clinical practice and whose reliability is unknown becomes evident. Given that the kappa agreement between SDIs and NSDIs for bipolar disorder is 0.4[7], the likelihood of a subject receiving the same diagnosis in both assessments is slightly above 15%[9]. This means that almost 85% of all patients undergoing treatment for bipolar disorder in an outpatient setting, after being diagnosed with NSDIs, would not be selected as research subjects. Consequently, they would receive treatment that is not evidence-based if they rely solely on clinical trials.

Received in: Jun/12/2023. **Approved in:** Jun/20/2023

1 Lisbon University, Medical Faculty, Lisbon, Portugal.
2 Federal University of Rio de Janeiro, Institute of Psychiatry, Postgraduate Program in Psychiatry and Mental Health, Rio de Janeiro, RJ, Brazil.
3 Federal University of Rio de Janeiro, Health Sciences Center, Rio de Janeiro, RJ, Brazil.
4 University of Lisbon, Faculty of Medicine, University Clinic of Psychiatry and Medical Psychology, Lisbon, Portugal.
Address for correspondence: Helio G. Rocha Neto. Av. Prof. Egas Moniz MB, 1649-028, Lisboa, Portugal. E-mail: helio.neto@edu.ulisboa.pt

On the other hand, SDIs only identify a subset of mental disorders diagnosed by clinicians[7]. The effects of this restriction on subjects' representation in medical development and research on organic disturbances in mental disorders are unknown. However, completely dismissing SDIs and operational criteria is akin to throwing out the baby with the bathwater. Clear diagnostic definitions and standardized assessments are crucial in mitigating common diagnostic biases that impact clinical assessments, such as missing information, anchoring, confirmation, and diagnostic availability biases[11,12]. If clear diagnostic definitions and standardized assessments are essential, they must be improved rather than discarded.

Operational criteria alone may be insufficient for a comprehensive description of mental disorders[4,13]. However, the previous model based on a simple narrative description was also inadequate. Prototypes naturally form the basis of clinical diagnostic reasoning[13,14], but diagnostic prototypes can and should incorporate operational operators as part of their descriptors. A valuable suggestion is to use prototype adequacy ranges, where clinicians can compare their observations with an ideal prototype that serves as a scaffold for diagnosis[13]. This approach is compatible with the dimensional approach in the latest classification system.

Diagnostic interviews are akin to diagnostic tests and require standardization. However, SDIs were directly built from operational criteria, following an up-down strategy (starting from the diagnosis and verifying its signs and symptoms), which is the opposite of the down-up strategy taught in clinical textbooks (collecting signs and symptoms first and then attempting to classify the disorder). Medical history taking, as a diagnostic technology, has been poorly studied, lacking a MeSH thesaurus or a valid global standard[6]. Nonetheless, understanding its components and refining its structure for research purposes might be easier to translate into clinical practice than using diagnostic criteria converted into questionnaires.

Currently, most reliability studies in psychiatry today are related to the validation of new diagnostic instruments or their comparison with SDIs, as well as the scales used to measure symptom intensity[7]. Many of these instruments are not meant for clinical practice, and their usage by clinicians remains unclear. The reason why reliability studies between research and clinical methods have been neglected is unclear, and the assumption that they are unnecessary is inaccurate. We are entering a new era of technological support for diagnosis and the review of diagnostic systems[6], stemming from a "brain decade" during which very few, if any, groundbreaking discoveries were made in psychiatry using SDIs and operational criteria as the diagnostic gold standard. It is perhaps time to recalibrate research and clinical diagnostic instruments, acknowledges their true limitations, and avoid falling into the trap of the sunk cost bias: the more we invest in a failed project, the more challenging it becomes to abandon it.

REFERENCES

1. Telles Correia D. Different perspectives of validity in psychiatry. J Eval Clin Pract. 2017;23(5):988-93.
2. Telles Correia D, Stoyanov D, Rocha Neto HG. How to define today a medical disorder? Biological and psychosocial disadvantages as the paramount criteria. J Eval Clin Pract. 2022;28(6):1195-204.
3. Helzer JE, Clayton PJ, Pambakian R, Reich T, Woodruff R, Reveley MA. Reliability of Psychiatric Diagnosis: II. The Test/Retest Reliability of Diagnostic Classification. Arch Gen Psychiatry. 1977;34(2):136-41.
4. First MB, Westen D. Classification for clinical practice: how to make ICD and DSM better able to serve clinicians. Int Rev Psychiatry. 2007 Oct;19(5):473-81.
5. Rocha Neto HG, Sinem TB, Koiller LM, Pereira AM, de Souza Gomes BM, Veloso Filho CL, et al. Intra-rater Kappa Accuracy of Prototype and ICD-10 Operational Criteria-Based Diagnoses for Mental Disorders: A Brief Report of a Cross-Sectional Study in an Outpatient Setting. Front Psychiatry. 2022;13:793743.
6. Rocha Neto HG, Cavalcanti MT, Correia DT. Structured Solutions for Medical History Taking: A Historical Review. Int J Psychiatry. 2022;7(2):144-52.
7. Rocha Neto H, Moreira ALR, Hosken L, Langfus JA, Cavalcanti MT, Youngstrom EA, et al. Inter-Rater Reliability between Structured and Non-Structured Interviews Is Fair in Schizophrenia and Bipolar Disorders – A Systematic Review and Meta-Analysis. Diagnostics (Basel). 2023;13(3):526.
8. Regier DA, Narrow WE, Clarke DE, Kraemer HC, Kuramoto SJ, Kuhl EA, et al. DSM-5 Field Trials in the United States and Canada, Part II: Test-Retest Reliability of Selected Categorical Diagnoses. Am J Psychiatry. 2013;170(1):59-70.
9. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012;22(3):276-82.
10. Spitzer RL, Cohen J, Fleiss JL, Endicott J. Quantification of Agreement in Psychiatric Diagnosis A new approach. Arch Gen Psychiatry. 1967;17(1):83-7.
11. Croskerry P, Singhal G, Mamede S. Cognitive debiasing 1: Origins of bias and theory of debiasing. BMJ Qual Saf. 2013;22 Suppl 2(Suppl 2):ii58-ii64.
12. Croskerry P, Singhal G, Mamede S. Cognitive debiasing 2: impediments to and strategies for change. BMJ Qual Saf. 2013;22(Suppl 2):ii65-72.
13. Westen D. Prototype diagnosis of psychiatric syndromes. World Psychiatry. 2012;11(1):16-21.
14. Parnas J. Differential diagnosis and current polythetic classification. World Psychiatry. 2015;14(3):284-7.