

The Uniform Postacute Assessment Tool: Systematically Evaluating the Quality of Measurement Evidence

Mark V. Johnston, PhD, Daniel Graves, PhD, Maureen Greene, MS

ABSTRACT. Johnston MV, Graves D, Greene M. The uniform postacute assessment tool: systematically evaluating the quality of measurement evidence. *Arch Phys Med Rehabil* 2007;88:1505-12.

The U.S. Congress has mandated that the Centers for Medicare & Medicaid Services develop a uniform assessment instrument that characterizes patients' needs for postacute services. What scientific criteria should be used to evaluate the evidence for such a tool? The validity of a measure can be accurately graded only if the constructs measured and their applications are clearly defined. We argue that improving postacute placement is the main purpose of the uniform postacute assessment (recently renamed the Continuity Assessment Record and Evaluation). We argue that placement itself needs to be better defined and measured in terms of transitions in the level and type of treatment and care. Domains that should be measured to provide appropriate rehabilitative placement recommendations include level of skilled medical and nursing care, therapies, routine living support, family support, ability to participate in self-care, and patient preference. Almost no research has been performed to quantify and predict the needed intensity of rehabilitative therapy, a major lacuna in evidence. Criteria and examples are provided for research that will provide minimal, probably adequate, or strong evidence for the validity of systems that recommend care transitions. A long-term program of research and systematic evidence synthesis is needed to support guidelines that improve postacute placement.

Key Words: Evidence-based medicine; Medicare; Occupational therapy; Outcome assessment (health care); Physical therapy (specialty); Rehabilitation; Reliability and validity.

© 2007 by the American Congress of Rehabilitation Medicine and the American Academy of Physical Medicine and Rehabilitation

TO INFORM AND IMPROVE both clinical and policy decisions, postacute care (PAC) needs an evidence basis. This evidence basis should accumulate over time. Criteria and methodologies have been developed for grading the level or strength of evidence for treatments and are relatively well known.¹⁻³ These methodologies have provided the technical

basis for modern systematic reviews and the evidence-based medicine movement, which has enhanced the identification and speeded the accumulation and application of best medical knowledge.^{1,3} It is possible to define criteria and methodologies for grading and synthesizing the strength of evidence for measures and measurement systems as well. The application of such criteria may be expected to speed progress and improve the application of measurement tools, including those used in postacute placement and policymaking.

By defining scientific criteria a priori, before reviewing the actual evidence, we can establish a basis for evaluation that is independent from the findings themselves and thus less subject to bias by transient situational considerations. With well-synthesized evidence, policy debate can at least start from a factual basis. Standardized criteria for grading the level of evidence can enable us to distinguish what is definitely known from what is probably or only possibly known, thus assisting application of measures and guiding priorities for continued research.

Judgments of the scientific evidence for a measurement system should be based on well-established scientific standards of reliability and validity.⁴⁻⁶ Although these principles and standards provide a necessary and useful basis, summary and interpretation are needed to apply them. The past literature on measurement reliability and validity is voluminous, highly detailed (at the level of statistical methods), and often vague (at the level of general principles and criteria). The phrase "reliable and valid" is frequently unqualified; a measure has it or not. When reliability and validity are treated as categories or dichotomies, the same evidence may be used to support or criticize because shades of gray must be labeled as either black or white. Discourse about measures easily becomes a debate of good versus bad or like versus dislike. To increase discrimination, more refined criteria are needed to evaluate the level of evidence for a measure.

A uniform postacute assessment (UPAA) tool is potentially so important that the highest-quality criteria should be applied to its development, validation, improvement, and uses over time. We will use the abbreviation UPAA to emphasize the generic, long-term view taken in this article but will occasionally use the more current name, the Continuity Assessment Record and Evaluation tool.

Objectives

This article is designed to identify research methods and questions that are critical to producing a valid and useful postacute assessment system. Our objectives were as follows:

1. To briefly describe a method of grading the level of evidence for a measure or measurement system based on established principles of reliability and validity of probabilistic measurement.^{4,5}
2. To examine the conceptual basis for UPAA measurement, particularly the primary stated use of a UPAA—improved postacute placement^{7,8}—and its conceptual bases. We sketch domains of patient care that must be distinguished to validate a UPAA, focusing on the needed intensity of continuing rehabilitative therapy.

From the Department of Occupational Therapy, College of Health Sciences (Johnston) and Nursing (Greene), University of Wisconsin-Milwaukee, Milwaukee, WI; and the Institute for Rehabilitation and Research, Baylor College of Medicine, Houston, TX (Graves).

Presented in part to the State-of-the-Science Symposium on Post-Acute Rehabilitation, February 12, 2007, Arlington, VA.

Supported in part by the National Institute on Disability and Rehabilitation Research (grant no. H133G060218).

No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit upon the author(s) or upon any organization with which the author(s) is/are associated.

Reprint requests to Mark V. Johnston, PhD, Occupational Therapy, University of Wisconsin-Milwaukee, PO Box 413, Milwaukee, WI 53201, e-mail: johnsto@uwm.edu.

0003-9993/07/8811-0009\$32.00/0

doi:10.1016/j.apmr.2007.08.117

3. To briefly review the literature on needed intensity of rehabilitative therapies such as physical therapy (PT), occupational therapy (OT), and speech-language pathology (SLP). We cite illustrative articles to provide an overview of the literature but do not attempt a definitive evidence review because extant literature is not routinely indexed in terms of measurement as a concept or subject heading.
 4. To identify a hierarchy of issues on the validity of the UPAA as a guide or tool for postacute placement, concentrating on the intensity of rehabilitative therapies needed by patients. We identify the criteria and study designs needed to validate a UPAA as a measure for type and intensity of therapies, grading these studies by strength of evidence they provide. Validation criteria that provide exploratory, minimal, probably adequate, and strong evidence are described.
- In sum, we asked what is the consequence of a focused application of scientific measurement principles to validation of the UPAA?

ADVANCING MEASUREMENT: TOWARD GRADED QUALITY CRITERIA

An understanding of principles and elements of measurement reliability and validity is a necessary basis for evaluating the quality of measures. Standards for evaluating the reliability and validity of measures have been published for educational and psychological tests⁵ and for interdisciplinary medical rehabilitation.⁴ As an essential basis for grading the quality of a measure, key information on reliability and validity needs to be systematically summarized in an evidence table. Table 1 presents a framework for such an evidence table. The framework is based on years of experience applying measurement principles and incorporates recent experience in grading the quality of outcome measures in spinal cord injury.⁹ It uses terminology found in our older publication on measurement standards in rehabilitation⁴ as well as concepts from item response theory (IRT) and Rasch analysis,¹⁰ which were developed to overcome limitations of classical test theory. Other researchers have also proposed improved methods of summarizing and rating measurement evidence that go beyond classical methods (eg, Andresen⁶).

As shown in the table, essential types of information for evaluating measurement quality include the following.

1. Content validity: although essential, content validity is only a starting point. Validation of a measure by a committee of experts who agree on its content is a common procedure, but it is not enough.
 2. Administrative characteristics: feasibility and cost, for instance, are critical in practice.
 3. Internal consistency, reliability, reproducibility, and biases: measurement error affects virtually all uses of a measure.
 4. Indicators of the internal validity of the scale and sensitivity to change: IRT and Rasch analysis provide strong indicators of the quality of a measure that can and should be used in grading measurement quality.⁶
 5. Criterion-oriented validity: direct evidence is needed that a scale predicts its most important practical criterion or improves the main decision it affects or is supposed to inform.
 6. Information on applicability to particular patient groups.
- Understanding and grading the validity of a measure should be based on a summary of all available factual evidence.

The Need for a Conceptual Foundation

A conceptual basis is needed so that reasonable inferences and limitations to the measurement system can be understood. In past efforts to summarize measurement evidence,⁹ we found that the level of evidence for a measure could be reliably graded only if the construct and its main application were clearly defined. When constructs or applications were vague, judgments of level of evidence became ambiguous; a measure could be valid for one application but of unknown validity for another application, and a scale would be judged as highly valid by a rater with one construct in mind but invalid by another rater who defined function or quality of life in a different way. In sum, the construct being measured and the application need to be defined to grade quality of measurement evidence.

Our overall grading of measurement quality is based on views of the unity of the concept of validity, emphasizing construct validity, which subsumes and integrates various aspects of validity, including content validity, reliability and internal structural characteristics, generalizability, external validity, and consequential validity.¹¹ However, in rating proposed UPAA measures, we particularly emphasize consequential validity (ie, usefulness in practice) because the UPAA encompasses the measurement of multiple constructs. An essential question should be answered: How well does the scale perform in its primary application? Although it is not realistic to expect that a measurement system will be completely validated for all of the uses to which it may be put, a measure should at least be validated for its main application. We will argue that judgment of level of validity is limited by the criteria and methods used in the validation study.

At the same time, indicators of the internal validity of the measure itself—content validity, reliability, and internal structure—are also a needed step not only to understand results found in application but also to understand and facilitate wider uses. Internal characteristics of a measure affect or constrain validity across multiple applications. So both reliability or internal characteristics and criterion-oriented validity need to be considered to judge the overall validity of a measure.

It is possible to synthesize these considerations into an overall grading of measurement validity. We sketch such a scheme designed to provide a simple summary of level of evidence to help readers distinguish measurement systems with stronger versus weaker supportive evidence. Although it is more refined than the traditional “reliable and valid” rubrics, the level grading (number of diamonds) should always be supplemented by explanation of the substantive strengths and limitations of the measure. Conclusions that the measure is (or is not) at 1 level of validity or another should always consider the substantive results as well as the study methodology. We apply these principles of evidence grading to the proposed UPAA later, focusing on issues of methodology and needed criterion domains.

KEY CONCEPTUAL ISSUES

Grading of the evidence for a measure begins by specifying a conceptual basis. We first address the issue of primary uses of the UPAA and then consider the major domains that need to be measured.

What is the Primary Criterion for Validation of a UPAA System?

The primary purposes of the UPAA are to serve as the basis for a payment system for postacute services, to monitor general quality across settings, and to inform “decisions for placement

Download English Version:

<https://daneshyari.com/en/article/3452522>

Download Persian Version:

<https://daneshyari.com/article/3452522>

[Daneshyari.com](https://daneshyari.com)