

PediatricsⁱⁿReview[®]

Research and Statistics : Reliability and Validity in Pediatric Practice

Nikeea Copeland-Linder

Pediatrics in Review 2009;30;278

DOI: 10.1542/pir.30-7-278

The online version of this article, along with updated information and services, is located on the World Wide Web at:

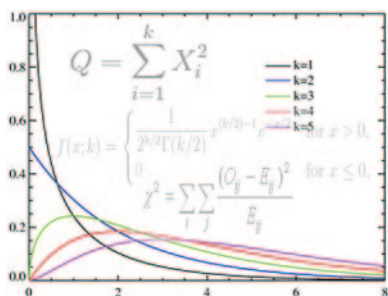
<http://pedsinreview.aappublications.org/content/30/7/278>

Pediatrics in Review is the official journal of the American Academy of Pediatrics. A monthly publication, it has been published continuously since 1979. Pediatrics in Review is owned, published, and trademarked by the American Academy of Pediatrics, 141 Northwest Point Boulevard, Elk Grove Village, Illinois, 60007. Copyright © 2009 by the American Academy of Pediatrics. All rights reserved. Print ISSN: 0191-9601.

American Academy of Pediatrics

DEDICATED TO THE HEALTH OF ALL CHILDREN™





RESEARCH & STATISTICS

Author Disclosure

Drs Copeland-Linder and Cheng have disclosed no financial relationships relevant to this article. This commentary does not contain a discussion of an unapproved/investigative use of a commercial product/device.

Reliability and Validity in Pediatric Practice

Nikeea Copeland-Linder, PhD, MPH*

Introduction

Welcome to a new series in *Pediatrics in Review*! The Pediatrics Review and Education Program (PREP The Curriculum®) includes *Pediatrics in Review* (PIR) and PREP Self-Assessment and is based on the American Board of Pediatrics' (ABP) 3,700 content specifications. These content specifications are the focus of the Board's Program for Maintenance of Certification. PREP The Curriculum® commits to covering all of the 3,700 content specifications over a 5-year period, helping the pediatrician prepare for the cognitive examination based on these knowledge statements.

In late 2008, the ABP added 26 new content specifications on research and statistics. To cover these, we introduce a new series. Clinician researchers from the Johns Hopkins University Division of General Pediatrics and Adolescent Medicine will be writing brief case-based or practice-oriented articles to address these topics. We hope that these articles will be useful for readers in evaluating the scientific literature and adopting evidence-based practice. We welcome your feedback on this new endeavor!—Tina L. Cheng, MD, MPH, Associate Editor

Case Studies

- A parent in your practice who has a 2½-year-old child asks whether she should use the newest tympanic thermometer on the market. You wonder if this thermometer will measure core body temperature

and are interested in the reliability and validity of the thermometer.

- Your colleagues suggest that your practice start using a maternal screening questionnaire assessing maternal depression in new mothers. You wonder if the six-question survey is reliable and valid in identifying maternal depression.

Having a clear understanding of reliability and validity in research is paramount to interpreting and determining the impact and generalizability of an instrument or test, whether it be a thermometer or a questionnaire. **Reliability** refers to the consistency or repeatability of scores. If you took the temperature again or gave the screening questionnaire again, would you get the same results? Do you get the same results when different staff members take the temperature or administer the questionnaire? **Validity** in research addresses whether an instrument or test actually measures what it is intended to measure. Does the tympanic thermometer measure core body temperature? Does the screening questionnaire really assess depression? This article addresses various aspects of reliability and validity to consider when evaluating an instrument or test.

Reliability

Test-retest reliability assesses whether an instrument or test yields the same results each time it is used with the same study sample under the same study conditions. One way to determine whether an instrument or test is reliable or consistent is to administer it with the same subject or sample more than once. For example, if the tympanic thermometer is used in the

*Division of General Pediatrics and Adolescent Medicine, Johns Hopkins University, Baltimore, Md.

right ear now, will it give the same reading a few seconds later? The two readings are correlated to determine stability. The instrument would be considered to have high test-retest reliability if the scores are similar on each administration. Test-retest reliability is useful when measuring a stable construct in which there is no substantial change in the phenomenon being measured between the two time points. Obviously, measuring test-retest reliability using the thermometer more than once within seconds makes more sense than testing over a long period of time when body temperature may change.

Internal consistency reliability is a measure of the consistency of the items within a test. Some instruments or tests are comprised of several items that are combined to create a composite score. The six-item maternal screening questionnaire is an example. The six items should be related because they are trying to measure the same construct (maternal depression). To assess the consistency of the items that make up the measure, split-half reliability may be calculated by splitting the items in half and correlating the items on each half of the test for each subject. Two additional methods of determining internal consistency are the Cronbach alpha and the Kuder-Richardson 20. The Cronbach alpha is used for multiple-ordered data, and scores range from 0 to 1, with higher scores indicating higher inter-item correlations. The Kuder-Richardson 20 is used for items that are dichotomous (eg, yes or no, true or false). In general, an alpha of 0.80 or greater is an indicator of high internal consistency.

Inter-rater reliability is the degree to which two raters independently score an observation similarly. Threats to reliability include time lapses between administration, with higher

time lapses contributing to lower correlations between scores, and having multiple raters or administrators of the test or instrument. The multiple rater issue can be addressed by giving clear instruction on how to use the instrument and assessing inter-rater reliability. In the tympanic thermometer example, would two different parents get the same temperature reading in the same child? For the screening questionnaire, training and education regarding implementation can improve the consistency between raters. Before deciding to use an instrument or test, researchers and practitioners should consult the instrument manual for published information on reliability and review the research literature for studies that report the reliability in different populations.

Validity

Content validity refers to the extent to which aspects of items that make up an instrument or test are representative of a particular construct. Two types of content validity include face validity and sampling validity. Face validity is a judgment about whether elements of an instrument make intuitive sense. Sampling validity refers to whether the instrument incorporates all of the aspects under study. For example, does the “screener” for depression include symptoms that represent various aspects of depression, including loss of energy, eating and sleeping disturbances, and difficulties with concentration?

Criterion validity is the degree to which the measurement correlates with an external criterion or another instrument or test that is considered valid. There are many different types of criterion validity:

- Convergent validity is the degree to which independent measures of the same construct are highly cor-

related. For example, administration of the depression screener may yield findings that are similar to a clinician’s assessment of a patient’s depression obtained by a clinical interview. In assessing the thermometer, does the tympanic thermometer and a rectal thermometer give the same reading?

- Predictive validity is the ability of an instrument or test to predict some future criterion. For example, do patients who are found to be at risk for developing depression according to the screening instrument develop a depressive episode later?
- Discriminant validity requires that an instrument or test show little or no correlation with measures from which it differs. For example, a depression screener would not be expected to have a high positive correlation to a measure of optimism.

No single type of evidence is sufficient for evaluating validity or reliability; instead, several types of validity or reliability should be evaluated in assessing a particular instrument or test. In fact, instruments or tests are not labeled “valid” or “reliable” across the board, but are so designated for a specific purpose or specific population. It is important to consider the different types of validity and reliability and the evidence that exists in judging an instrument or test in practice.

Suggested Reading

- Gliner JA, Morgan G, Harmon RJ. Measurement reliability. *J Am Acad Child Adolesc Psychiatry*. 2001;40:486–488
- Morgan G, Gliner J, Harmon RJ. Measurement validity. *J Am Acad Child Adolesc Psychiatry*. 2001;40:729–731
- Hudley S, Cummings SR, Browner WS, et al. *Designing Clinical Research*. Philadelphia, Pa: Lippincott Williams & Wilkins; 2001

Research and Statistics : Reliability and Validity in Pediatric Practice

Nikeea Copeland-Linder
Pediatrics in Review 2009;30;278
DOI: 10.1542/pir.30-7-278

Updated Information & Services

including high resolution figures, can be found at:
<http://pedsinreview.aappublications.org/content/30/7/278>

References

This article cites 2 articles, 0 of which you can access for free at:

<http://pedsinreview.aappublications.org/content/30/7/278#BIBL>

Subspecialty Collections

This article, along with others on similar topics, appears in the following collection(s):

Research and Statistics

http://pedsinreview.aappublications.org/cgi/collection/research_statistics

Permissions & Licensing

Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at:

</site/misc/Permissions.xhtml>

Reprints

Information about ordering reprints can be found online:

</site/misc/reprints.xhtml>

American Academy of Pediatrics

DEDICATED TO THE HEALTH OF ALL CHILDREN™

