Conceptual and Practical Issues in Measurement Validity

MERMAID Series, December 10, 2010

Galen E. Switzer, PhD

Clinical and Translational Science Institute
VA Center for Health Equity Research and Promotion
University of Pittsburgh
1. Measurement Overview
2. Reliability and Validity
3. Survey Research
Measurement Overview
What is Measurement?

- observation of people, clinical events, biological/physiological processes, lab observations
- conversion of observations to quantitative data
- assignment of numbers to different qualitative levels of the variable
  - cancer staging 1-4
  - blood pressure reading
  - Likert scaling
A Few Key Terms

- **construct, concept, conceptual variable** – a theoretical/unobserved variable

- **indicator, item, question, measured variable** – an observed/measured variable

- **measure, instrument, scale** – generally a group of measured indicators assessing a single construct
Measurement Process

conceptualize

↓

operationalize

↓

select measures

↓

gather data
Create a Conceptual Model

• review existing models
  - models of health behavior when measuring vaccination attitudes/behavior

• identify key conceptual variables

• depict relationships among variables
  - draw the model
  - consistent with study aims
Categories of Variables

- **demographic**
  - gender
  - ethnicity
  - age
  - occupation

- **clinical**
  - disease status
  - blood pressure
  - graft versus host disease
  - HLA types

- **psychosocial**
  - depression
  - satisfaction with care
  - health beliefs
  - religiosity

- **social**
  - social support
  - family cohesion
  - access to health care
  - patient/physician communication
Conceptual Model Example

Culturally-based Characteristics
  Race

Psychosocial Characteristics

Donation-related Characteristics

CT-stage Decision

Donor background characteristics to be controlled
- gender
- age
- marital status
- education
- occupation
- SES
- religion
- blood donor
- # children
Measurement Process

conceptualize

operationalize

select measures

gather data
Operationalize

Operationalization

Construct/Concept

Measured Variable

Risk of Breast Cancer

Genetic Marker
Family History
Previous Cancer
Operationalize

Operationalization

Construct/Concept

Measured Variable

Medical Knowledge

MCAT
USMLE – 1,2,3
Board Certification Experience
Operationalize

- identify key conceptual variables
  - conceptual model or literature

- review available measures

- select measures based on
  - degree of specificity
  - measurement properties
  - appropriateness for study group
Measurement Process

conceptualize

operationalize

select measures

gather data
Measure Selection
Reliability and Validity

- **Reliability** is the extent to which a test consistently measures something. If a test is reliable it produces the same result as long as there is no change in the true score.

- **Validity** of a measurement tool (e.g. test in medical education) is the degree to which the test measures what it claims to measure.
Validity

The extent to which a measure:

• measures what it was intended to measure
• is related to other variables in predicted ways

Constraints on validity:

• it is rare for a measure to be completely validated
  (e.g., HLA typing – serology, DNA, blood, saliva)
• re-validate the measure when used for a different purpose
  (e.g., QoL measures for children, Distress measures for Hispanics)

**Reliability is necessary but not sufficient for validity**
Building a Case for Validity

• most measures are never perfectly validated
• there is no clear point at which a measure is considered valid
• additional evidence only adds or diminishes support for validity
A Sampling of Measures Currently Seeking Validity Evidence

• NIH Peer review process
• ProMIS laparoscopic simulator
• Biomarkers for Gastric Cancer Screening
• Physical activity monitors
• Fatigue Severity Scale
• PALOC-s: A Post-Acute Level of Consciousness scale
• Mini International Neuropsychiatric Interview
• Tumour diameter assessed by magnetic resonance imaging
• Electronic surveillance systems in hospitals
• Detection of Pain-Related Malingering
Types of Validity

1. face
2. content
3. criterion-related
   • concurrent
   • predictive
   • postdictive
4. construct
   • factor
   • convergent
   • discriminant

Related Procedures

1. Known-groups
2. Multitrait – Multimethod
3. Factor Analysis
Face Validity

- extent to which the measure appears to assess the intended construct
- subjectively evaluated
- useful for public acceptance of a measure
Face Validity of SCL-90R

items included in somatization dimension:

- headaches
- faintness or dizziness
- pains in heart or chest
- pains in lower back
- nausea or upset stomach
- soreness of your muscles
- trouble getting your breath
- hot or cold spells
- numbness or tingling in parts of your body
- heavy feeling in your arms or legs
Content Validity

• Refers to how well the content of the measure represents the domain being measured.

• Content validity is assessed through judgment of experts in the domain content.

• Experts evaluate whether each indicator/item reflects the concept/construct indicators taken together cover the full range of the construct.
Content Validity

- Content validity asks – do selected indicators represent the full range of possible indicators for the domain.
Evaluating Content Validity

• define the concept/construct
  • review the literature
  • consult experts

• define the universe of indicators
  (e.g., S. Weisbord, Dialysis Symptom Inventory)

• experts determine if indicators represent construct
  • calculate an average indicator % agreement
    > 70% is acceptable
  • experts identify missing indicators
  • experts include patients
Content Validity and Apache II

1. Age
2. Temperature
3. Respiratory rate
4. Organ insufficiency immunocompromised
5. Mean arterial pressure
6. Heart rate
7. Oxygenation
8. Serum potassium
9. Serum creatinin
10. Serum sodium
11. White blood count
12. Arterial pH
13. Glasgow coma score
Criterion Validity

- assesses measure’s ability to predict an outcome
  
  (e.g., health status using SF-36 → hospitalization)
  (e.g., Apache II score → death)

- the measure of interest is the **predictor**

- the “outcome” is the **criterion**

- **correlate** predictor and criterion scores

- higher correlations indicate higher criterion validity of predictor variable
Predictive Criterion Validity

- Medical Potential
- MCAT
- USMLE scores
Construct Validity

Construct validity encompasses multiple types of validity evidence and answers the fundamental question:

“Is the test measuring what it is intended to measure?”
Construct Validity

- Construct validity requires gathering several types of validity evidence; e.g., content validity, criterion-related validity
- Construct validity cannot be established through a single validity study
- Support for construct validity must be built from a series of studies
- Ideally, a series of validity studies are done for a measure which provides validity evidence for that measure.
Some Questions to Ask

• If a college student has a high score on a measure of self-rated math ability, how confident can we be that she really has high math ability?

• If an medical student has a low score on an internal medicine test, how confident can we be that he really lacks sufficient knowledge of the content area?

• If a cardiac stress test shows that a patient has heart disease, how sure can the doctor be that the patient really does have heart disease?
Measurement issues in Questionnaire Design
Survey Research: Measure Selection & Design

- existing vs new measure
- item terminology/structure
- response alternatives
- item sequence
Existing vs New Measure

- do standard measures exist
  (health related quality of life measures)

- is comparison to other samples important
  (Nottingham Health Profile in Transplantation)

- is content appropriate for study group
  (Friedman – gay and lesbian youth, adult measures)

- are measurement properties adequate
  (e.g. reliability and validity)

- what method of administration is required
  (telephone interview, in-person interview, paper & pencil, online)

- is respondent burden minimized
  (consider motivation – bone marrow donors, compromised health status)
Item Terminology/Structure: Things to Avoid

• difficult or unfamiliar words
  instead of .......... use
  candid ............. honest
  priority ............ most important
  assistance .......... help
  virtually ........... nearly

• maximum 8\textsuperscript{th} grade reading level
  (4\textsuperscript{th} grade recommended)

• use reading level check in Word or other program
  tools .......... spelling/grammar .......... readability
How many hours a day do you spend watching TV?

<table>
<thead>
<tr>
<th>Set 1</th>
<th>Set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>up to ½ hour</td>
<td>up to 2 ½ hours</td>
</tr>
<tr>
<td>½ to 1 hour</td>
<td>2 ½ to 3 hours</td>
</tr>
<tr>
<td>1 to 1½ hours</td>
<td>3 to 3½ hours</td>
</tr>
<tr>
<td>1½ to 2 hours</td>
<td>3½ to 4 hours</td>
</tr>
<tr>
<td>2 to 2½ hours</td>
<td>4 to 4½ hours</td>
</tr>
<tr>
<td>&gt; 2 ½ hours</td>
<td>&gt; 4 ½ hours</td>
</tr>
</tbody>
</table>

“more than 2½”

set 1 = 16%  
set 2 = 38%
Things to Avoid

- questions that include qualifying phrases or clauses

“Have you experienced any of the following symptoms in the past year, not including the past month?”

“What types of medication have you used in the past week? (Do not include nonprescription medications.)”
Item Terminology/Structure: Things to Avoid

- items that include multiple ideas (double-barreled)

“With their precarious health, it really isn’t fair to expect transplant recipients to return to work.”

“Managed care cuts costs and provides the best patient care.”

“How much difficulty did you have getting your medications or seeing a doctor during the past 6 months?”
Item Terminology/Structure: Things to Avoid

- questions with ambiguous or vague words

  phrase: “over the last few years”

  meant: no more than 2 yrs 12%
         up to 7 yrs 31%
         up to 10 yrs 32%
         > 10 yrs 19%

  “What kind of headache remedy do you usually use?”

  “Do you attend religious services regularly?”

  “How many times in the past year have you seen or talked with someone about your health?”
Response Alternatives

• open vs closed item
  • Open = richer info, but more difficult to complete and interpret
  • Closed = limited info, but easier to complete and interpret

• text labels vs numbers (Likert)
  recommend both
  1 strongly agree
  2 agree
  3 neutral
  4 disagree
  5 strongly disagree

• number of response alternatives
  (≤ 7 to avoid false specificity)

• odd vs even number of responses
  (forced choice versus neutral category)
Example of a Less Than Ideal Item
(from online survey of health habits)

1. When you are exercising in your usual fashion, how would you perceive your level of exertion?
   - Nothing
   - Very weak
   - Weak
   - Moderate
   - Moderate to strong
   - Strong
   - Strong to very strong
   - Very strong
   - Very strong to very, very strong
   - Very, very strong
   - Very, very strong to maximal
   - Maximal
Item Sequence

- **Goals**
  - to ease the respondent’s task
  - to avoid biases in responses to later items

- **overall sequence**
  - establish trust (begin with topical item)
  - demographics last
  - less important items later

- **sequence within a topic area**
  - general to specific

- **transition statements**
  - introduce each major section
Questionnaire
Appearance

- item formatting
  - visually appealing
  - varied
  - easy to read
  - low perceived respondent burden
Recommendations for Questionnaire Construction

• choose existing, well-standardized measures

• word questions carefully to avoid misunderstandings and ambiguities – pretest

• eliminate double-barreled questions – pretest

• be aware that respondents will make judgments based on response categories

• consider using some open-ended questions

• carefully format questions

• control the question sequence
Locating Measures

review the literature

• published articles & chapters
• handbooks & directories
  – *The Mental Measurements Yearbook* (Spies & Plake, 2005)
  – *Measuring Health* (McDowell, 2006)
• computerized databases
  – *Health and Psychosocial Instruments (HaPI)*