What is Measurement?

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

- **construct, concept, conceptual variable** – a theoretical/unobserved variable – e.g., satisfaction

- **indicator, item, measured variable** – a variable that assigns numbers to a concept – e.g., satisfaction item

- **instrument, scale** – generally a group of measured indicators assessing a single construct – e.g., Consumer Assessment of Healthcare Providers and Systems (CAHPS)
Measurement Process

1. identify a concept
2. operationalize
3. select measures
4. gather data
Examples of Variable Categories

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

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

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

**social**
- social support
- family cohesion
- access to health care
- patient/physician communication
Variables are often organized into Conceptual Models

- Visually depict relationships among variables
- Useful for rapid visual understanding of information
- Guide study hypotheses
- Guide study analyses
- Are critical for most grant proposals
Stress Response Model

Hans Selye, Stress Response Model
Measurement Process

- identify a concept
- operationalize
- select measures
- gather data
**Operationalize**

- identify key conceptual variables
  - conceptual model or literature
- review available measures
- select measures based on
  - measurement properties
  - appropriateness for study group
  - other study-specific needs
Operationalize

Construct/ Concept

Operationalization

Risk of Breast Cancer

Operationalization

Measured Variable

Genetic Marker
Family History
Previous Cancer
Operationalize

Construct/Concept

Heart Attack

Operationalization

Measured Variable

EKG
CK-MB
Troponin
Other cardiac biomarkers
Indicators of a Concept
(Domain Sampling Theory)

• individual indicator is one of multiple possible indicators

• individual indicators may be imperfect

• any combined measure should be a reflection of all indicators
Domain Sampling Theory

Health Status

- Daily Activities
- Blood Pressure
- Glucose Level
- Liver Enzymes
- Lung Diffusion Capacity
Operationally Defining Academic Career Success

Achievement
- Undergrad GPA
- MCAT
- Research Experience

Academic Success
- Papers
- Grants
- Awards
Measurement Process

conceptualize

operationalize

select measures

gather data
Considerations in Measure Selection

- Reliability (consistency)
- Validity
What is Reliability?

- Reliability is about consistency

- If a test is reliable it produces the same result as long as there is no change in the true score

- “The extent to which the measurement is inherently reproducible”

- The minimization of random error
Reliability

Two inter-related concepts:

• **consistency**
  consistency across
  - **raters** (x-ray)
  - **time** (blood pressure)
  - **versions/instruments** (depression screen, bp cuff)
  - **indicators** (patient satisfaction)

• **minimization of random error**
Reliability

Consistency across:

- raters = inter/intra-rater
- time = test-retest
- versions = alternate form
- indicators = internal consistency
What is Adequate Reliability?

• depends on the goal of the measurement

• higher reliability needed for
  - diagnostic tests used with individual patients
    e.g., diagnosis of HIV versus prevalence rates
  - studies with small N
    larger N provides more reliability – outliers have less effect

• guidelines for various situations
  Psychometric Theory, Nunnally & Bernstein (1994)
What is Validity?

• Validity is the extent to which conclusions from a measure are accurate – are you measuring what you think you are measuring?

• Validity affects the degree of confidence we have in interpreting the scores from an instrument.

• Validity is situation specific. Interpretations of scores from a measure are not necessarily valid in a new group or context.
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
Validity

Validity in measurement

• face/content
• criterion
  - predictive
  - concurrent
• construct
  - convergent/discriminant
  - factor
• generalizability
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

• How well do the indicators measure the full domain or construct being measured?

  Example: Does a robotic surgical skills simulator test the full range of tasks/functions that surgeons will actually perform?

• Content validity is assessed by expert opinion

• Experts evaluate
  - whether each indicator/item reflects the domain
  - whether the indicators together cover the full range of the domain/construct of interest.

Illustration of experts rating items for Content Validity

Please use the following scale to rate the relevance of the eight symptoms to the experience of patients undergoing dialysis.

1 = highly relevant  
2 = somewhat relevant  
3 = not relevant

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Rater 1</th>
<th>Rater 2</th>
<th>Rater 3</th>
<th>Rater 4</th>
<th>Rater 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dry skin</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Fatigue</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Itching</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Bone pain</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Dry mouth</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Worrying</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Nausea</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Headache</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Criterion Validity

- assesses measure’s ability to predict an outcome
  (e.g., health status using SF-36 $\rightarrow$ hospitalization)
  (e.g., Apache II score $\rightarrow$ 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.
Measurement issues in Questionnaire Design
Survey Research: Measure Selection & Design

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

- do adequate measures already exist
  (health-related quality of life measures)
- is the content appropriate for the study group
  (pain scales for the cognitively impaired)
- is the method of administration appropriate
  (telephone interview, in-person interview, paper & pencil, web-based)
- is respondent burden minimized
  (consider motivation and ability – e.g., 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 8th grade reading level
  (4th grade recommended)

• use reading level check in Word or other program
  tools .......... spelling/grammar .......... readability

In MS Word
1. Click the File tab, and then click Options.
2. Click Proofing.
3. Select Show readability statistics.
**Item Terminology/Structure: Things to Avoid**

- unproven or biased response alternatives

How many hours a day do you spend watching TV?

<table>
<thead>
<tr>
<th>First Administration</th>
<th>Second Administration</th>
</tr>
</thead>
<tbody>
<tr>
<td>up to ( \frac{1}{2} ) hour</td>
<td>up to 2 ( \frac{1}{2} ) hours</td>
</tr>
<tr>
<td>( \frac{1}{2} ) to 1 hour</td>
<td>2 ( \frac{1}{2} ) to 3 hours</td>
</tr>
<tr>
<td>1 to 1( \frac{1}{2} ) hours</td>
<td>3 to 3( \frac{1}{2} ) hours</td>
</tr>
<tr>
<td>1( \frac{1}{2} ) to 2 hours</td>
<td>3( \frac{1}{2} ) to 4 hours</td>
</tr>
<tr>
<td>2 to 2( \frac{1}{2} ) hours</td>
<td>4 to 4( \frac{1}{2} ) hours</td>
</tr>
<tr>
<td>&gt; 2 ( \frac{1}{2} ) hours</td>
<td>&gt; 4 ( \frac{1}{2} ) hours</td>
</tr>
</tbody>
</table>

“more than 2\( \frac{1}{2} \)” → set 1 = 16%  
set 2 = 38%
Item Terminology/Structure: 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)

“ACA 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?”

“With their precarious health, it really isn’t fair to expect transplant recipients to return to work.”
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 Survey 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)
  – http://www.measurementexperts.org/