Test-Retest Correlations and Construct Validity

Patricia Markham Risica, DrPH, RD
October 5, 2015
Learning Objectives:

- At the completion of the webinar participants will be able to:
  - describe the steps for demonstrating reliability of an instrument and one example based in nutrition assessment
  - describe the steps for demonstrating validity of an instrument and one example based in nutrition assessment
  - identify characteristics of construct validity and one example based in nutrition assessment
Reliability = Repeatability

- **Definition of RELIABILITY.** 1. : the quality or state of being **reliable**. : the extent to which an experiment, test, or measuring procedure yields the same results on repeated trials. (Merriam-Webster)

- Also called **test-retest reliability**
Other kinds of Reliability

- Interrater Reliability
- Parallel Forms Reliability
- Internal Consistency Reliability
Inter-rater Reliability

- Would two observers or measurers find the same results?
Parallel Forms Reliability

Form A

=*

Form B
## Internal Consistency Reliability

### Measure

<table>
<thead>
<tr>
<th>Item 1</th>
<th>Item 2</th>
<th>Item 3</th>
<th>Item 4</th>
<th>Item 5</th>
<th>Item 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Average Inter-Item Correlation

<table>
<thead>
<tr>
<th></th>
<th>$I_1$</th>
<th>$I_2$</th>
<th>$I_3$</th>
<th>$I_4$</th>
<th>$I_5$</th>
<th>$I_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_1$</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$I_2$</td>
<td>0.89</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$I_3$</td>
<td>0.94</td>
<td>0.92</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$I_4$</td>
<td>0.88</td>
<td>0.93</td>
<td>0.94</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$I_5$</td>
<td>0.94</td>
<td>0.88</td>
<td>0.92</td>
<td>0.84</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>$I_6$</td>
<td>0.85</td>
<td>0.96</td>
<td>0.89</td>
<td>0.91</td>
<td>0.94</td>
<td>1.0</td>
</tr>
</tbody>
</table>
Test- Retest Reliability

Measure A

Time 1

Measure A

Time 2
Reliability and Validity of Nutrition Knowledge Questionnaire for Adults
Anna Marie Jones, PhD1; Cathi Lamp, MS, MPH, RD2; Marisa Neelon, MS3; Yvonne Nicholson, MS4; Connie Schneider, PhD, RD5; Patti Wooten Swanson, PhD6; Sheri Zidenberg-Cherr, PhD (J Nutr Educ Behav. 2015;47:69-74)

- Initial Questions Came From:
  - General Nutrition Knowledge Questionnaire
  - New Questions on DGA and My Pyramid

- Final Question Sections:
  - Familiarity with My Pyramid
  - Nutrition Content of Foods
  - Diet-Disease Relationships
  - Overall
Reliability and Validity of Nutrition Knowledge Questionnaire for Adults

Anna Marie Jones, PhD1; Cathi Lamp, MS, MPH, RD2; Marisa Neelon, MS3; Yvonne Nicholson, MS4; Connie Schneider, PhD, RD5; Patti Wooten Swanson, PhD6; Sheri Zidenberg-Cherr, PhD (J Nutr Educ Behav. 2015;47:69-74)

- **Reliability**

  “Test–retest reliability was moderate to high (r = 0.93, 0.92, and 0.84, respectively for Familiarity with MyPyramid, Nutrition Content of Foods, and Diet–Disease Relationships) with total reliability equal to 0.95 (P < .001 for all correlations)”

- **Pearson’s Correlations (r):**
  - Familiarity with My Pyramid 0.93
  - Nutrition Content of Foods 0.92
  - Diet–Disease Relationships 0.84
  - Overall 0.95
Pearson’s Correlation?

- **Pearson’s correlation**
  - Any parametric or non-parametric correlation coefficients measure LINEAR TREND, but do NOT measure AGREEMENT.

- Instead the Intraclass Correlation Coefficient (ICC or $r_i$) should be used to correct for the overall number of chance expected agreements.
Validity and test-retest reliability of a short dietary questionnaire to assess intake of saturated fat and free sugars: a preliminary study.

Francis H1, Stevenson R. (J Hum Nutr Diet. 2013 Jun;26(3):234-42.)

Measures

- **Dietary Fat and free Sugar – Short Questionnaire (DFS)**
  - ~10 days after the first administration, 29 ppt repeated the DFS

- Commonwealth Scientific and Industrial Research Organisation Food Frequency Questionnaire (C-FFQ)

- 4-Day Diet Diary (DD)
Validity and test-retest reliability of a short dietary questionnaire to assess intake of saturated fat and free sugars: a preliminary study.

Francis H¹, Stevenson R. (J Hum Nutr Diet. 2013 Jun;26(3):234-42.)

Analysis

- The intraclass correlation coefficient (ICC) was calculated as an indication of test–retest reliability

- ICC assesses within subject variability
Validity

- Many types of validity
  - Psychology research
  - Epidemiology research

- Construct Validity
  - The degree to which a method measures what it says that it measures (Klaver, et al. 1988)
  - How well a test measures up to its claims (https://explorable.com/types-of-validity)
Validation

Truth

Measure
Validation

Truth → Measure → Measure
Validation

Truth

Measure

Gold Standard

Validation

Calibration

Measure

Validation in Nutrition Research

- What is a Gold Standard?
  - Biological Samples – Nutritional Status
  - Doubly Labeled Water - Metabolism
  - Under Water Weighing – Body Composition
  - CT Scan - Body Fat Distribution
  - Dietary Intake – Weighed and Measured Food Intake
Why Not Use the Gold Standard??

- Expensive
- Hard to do
- Intrusive
- High Participant Burden
- Does not represent a free living population
Validity and test-retest reliability of a short dietary questionnaire to assess intake of saturated fat and free sugars: a preliminary study.

Francis H¹, Stevenson R. (J Hum Nutr Diet. 2013 Jun;26(3):234-42.)

- **Measures**
  - Dietary Fat and free Sugar – Short Questionnaire (DFS)
    - ~10 days after the first administration, 29 ppt repeated the DFS
  - Commonwealth Scientific and Industrial Research Organisation Food Frequency Questionnaire (C-FFQ)
  - 4-Day Diet Diary (DD)
Validity and test-retest reliability of a short dietary questionnaire to assess intake of saturated fat and free sugars: a preliminary study.

Francis H¹, Stevenson R. (J Hum Nutr Diet. 2013 Jun;26(3):234-42.)

Measures

- Commonwealth Scientific and Industrial Research Organisation Food Frequency Questionnaire (C-FFQ)
- 4-Day Diet Diary (DD)
  - Both were converted to percentage Kj of mean daily Kj
  - Reflect nutrient composition, rather than higher or lower intake
  - Recoded to reflect “Free Sugar” based on author definition
  - Because not normally distributed, used
    - Spearman’s Rank Order Correlation
Validation Results

Table 2  Spearman rank-order correlations between Dietary Fat and free Sugar – Short Questionnaire score and nutrient estimates from the full-length food frequency questionnaire and diet diary (n = 40)

<table>
<thead>
<tr>
<th></th>
<th>DD</th>
<th>C-FFQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy (kJ)</td>
<td>0.35*</td>
<td>0.40†</td>
</tr>
<tr>
<td>Protein</td>
<td>−0.21</td>
<td>−0.18</td>
</tr>
<tr>
<td>Fat</td>
<td>0.35*</td>
<td>0.63†</td>
</tr>
<tr>
<td>Saturated fat</td>
<td>0.46†</td>
<td>0.71†</td>
</tr>
<tr>
<td>Polyunsaturated fat</td>
<td>−0.03</td>
<td>−0.30</td>
</tr>
<tr>
<td>Monounsaturated fat</td>
<td>0.13</td>
<td>−0.25</td>
</tr>
<tr>
<td>Carbohydrate</td>
<td>−0.05</td>
<td>−0.24</td>
</tr>
</tbody>
</table>
Conclusions

- Reliability and Validity are very different constructs
- **Reliability** = Repeatability
  - Assesses within person differences
- **Validity** = Is it measuring what you think it is?
  - Comparison with Gold Standard or an instrument more respected than the one being testing
  - Validity is an imperfect and imprecise science