Validity and reliability of measurements
The Yin and the Yang of Research
Components in a dataset

Data in the Research Data set

Correct Scores

Errors
Random error

- The difference, due to chance alone, of an observation on a sample from the true population value

- Leading to lack of precision in the measurement of association

- Three major sources
  - Individual biological variation
  - Sampling variation
  - Measurement variation
Random error ...........

- Random error can never be completely eliminated since we can study only a sample of the population.

- **Random error can be reduced by the careful measurement**

- Sampling error can be reduced by optimal sampling procedures and by increasing sample size
Do you already have challenges / questions related to measurement?
2. Why bother
Examples

- **Unreliable** histological score – the score is not measuring what we were planning to measure.

- **Imprecise or Inaccurate** protein measurement –
  - too big variation to be able to draw conclusion
  - too low/ high detection in our samples to be able to compare differences.
Examples

• **Unreliable** weighing scale
  Is my intervention that aimed at weight loss really unsuccessful, or was the weighing scale unreliable?

• **Imprecise** blood heart rate monitor
  Did the medication really change blood pressure?

• **Inaccurate** observation
  Did I really observe an improvement in scarring?

• **Instable** depression questionnaire
  Did these patients really get worse after counselling?
2. What is reliability?
Components in a dataset

- Data in the Research Data set
  - Correct Scores
  - Errors
• **Reliable measures (instruments)** maximize the true score component and minimize error

• **Terms:**
  - Stability
  - Consistency
  - Accuracy
  - Dependability
Reliability

"Reliability refers to the degree of consistency or accuracy with which it measures the target attribute."
• The idea behind **reliability** is that any significant results must be more than a one-off finding and be repeatable.

• Other people must be able to perform exactly the same test, under the same conditions and generate the same results.
Reliability

- Stability
- Internal consistency
- Equivalence
Test-retest (stability)

Can be used in:

- Self-report
- Observational
- Physiologic
Development and psychometric evaluation of the Thirst Distress Scale for patients with heart failure

Nana Waldréus¹, Tiny Jaarsma²,³, Martje HL van der Wal²,⁴ and Naoko P Kato²,⁵

Abstract
Background: Patients with this in patients with heart fail
Heart Failure (TDS-HF) and

Data collection
The study was performed between 2012–2014 in Sweden, 2014–2015 in the Netherlands, and between 2015–2016 in Japan. Patients were asked to fill in the TDS-HF after the visit at the outpatient HF clinic or during hospital admission. To evaluate the test-retest reliability of the TDS-HF, patients from Sweden and Japan filled in the questionnaire again after two weeks.
Table 5. Stability of the eight-item Thirst Distress Scale for patients with Heart Failure (TDS-HF) over time (n=77).

<table>
<thead>
<tr>
<th>Item</th>
<th>Weighted kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. My thirst bothers me a lot</td>
<td>0.45</td>
</tr>
<tr>
<td>2. I am very uncomfortable when I am thirsty</td>
<td>0.33</td>
</tr>
<tr>
<td>3. My mouth feels like sandpaper when I am thirsty</td>
<td>0.41</td>
</tr>
<tr>
<td>4. My mouth feels dry when I am thirsty</td>
<td>0.29</td>
</tr>
<tr>
<td>5. My saliva is very thick when I am thirsty</td>
<td>0.39</td>
</tr>
<tr>
<td>6. When I drink less water, my thirst gets worse</td>
<td>0.32</td>
</tr>
<tr>
<td>7. I am so thirsty I could drink water uncontrollably</td>
<td>0.45</td>
</tr>
<tr>
<td>8. My thirst feels difficult to overcome</td>
<td>0.60</td>
</tr>
</tbody>
</table>

Intra-class correlation coefficient (95% CI) 0.88 (0.81–0.93)<sup>a</sup>

Cl: confidence interval.

<sup>a</sup>p-value<0.001.
Questions with stability

- What is the best time frame between test-restest
- How many people should do the test-retest
- How to ask /prepare people
- EPN?
- Same sample /new sample?
Internal consistency

- Ensure that each part of the test generates similar results, and that each part of a test measures the correct construct.
- For example, a test of anxiety should measure anxiety only, and every single question must also contribute.
- If people’s responses to Item 1 are completely unrelated to their responses to Item 2, then it does not seem reasonable to think that these two items are both measuring anxiety … so they probably should not be on the same test.
Cronbach’s alpha

The coefficient alpha is defined as follows

\[ \alpha = \left( \frac{K}{K-1} \right) \left( 1 - \frac{\sum V_i}{V_T} \right) \]

where \( K \) represents the number of the item; \( V_1, V_2, \ldots, V_k \) represent the variation of 1, 2, \ldots, \( K \) item points, respectively; and \( V_T \) represents the variation of scale (total) score. When the scale items are scored as 0 or 1, Equation (4) turns into the formula Kuder-Richardson KR-20.
Or in plain english…

• The most common way to assessing and reporting internal consistency for multi item scales

• A function of two things:
  • The average inter-item correlation
  • The number of items

• A common interpretation of the $\alpha$ coefficient
  • $<$0.7 Unsatisfactory
  • $\geq$0.7 Acceptable
  • $\geq$0.8 Good
  • $\geq$0.9 Excellent
A common misconception

- Cronbach’s alpha is a measure of uni-dimensionality
- No! This is a misconception as multidimensional scales can yield a high Cronbach’s alpha coefficient

Why?

- High $\alpha$-values will be produced if there are sub-dimensions with high correlations between items in each subset but low inter-item correlations across sub-dimensions
Internal consistency

• Item-item correlation
• Item-total correlation (corrected)
• Split-half
Internal consistency
Internal consistency
<table>
<thead>
<tr>
<th>Number</th>
<th>Item stems</th>
<th>Median (q&lt;sub&gt;1&lt;/sub&gt;–q&lt;sub&gt;3&lt;/sub&gt;)</th>
<th>12-item version</th>
<th>Item-total correlation&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Cronbach’s α if item deleted</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I weigh myself every day</td>
<td>3 (2–5)</td>
<td>0.37</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>If I get SOB I take it easy</td>
<td>1 (1–2)</td>
<td>0.25</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>If SOB increases I contact my doctor or nurse</td>
<td>2 (1–3)</td>
<td>0.60</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>If leg/feet are more swollen, I contact doctor or nurse</td>
<td>2 (1–3)</td>
<td>0.64</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>If I gain weight I contact doctor or nurse</td>
<td>2 (1–4)</td>
<td>0.65</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>I limit the amount of fluids</td>
<td>2 (1–3)</td>
<td>0.49</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>I take a rest during the day</td>
<td>1 (1–2)</td>
<td>0.24</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>If I experience fatigue I contact doctor or nurse</td>
<td>2 (1–3)</td>
<td>0.60</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>I eat a low salt diet</td>
<td>2 (1–3)</td>
<td>0.40</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>I take my medication as prescribed</td>
<td>1 (1–2)</td>
<td>0.33</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>I get a flu shot every year</td>
<td>1 (1–3)</td>
<td>0.14</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>I exercise regularly</td>
<td>3 (2–4)</td>
<td>0.27</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Total scale</strong></td>
<td>26 (20–31)</td>
<td><strong>0.77&lt;sup&gt;b&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Item-total correlation computed using item scores from the 12-item version.

<sup>b</sup> Cronbach’s α computed using item scores from the 12-item version.
Reliability

- Stability
- Internal consistency
- Equivalence: the degree to which two or more independent observers or coders agree about scoring
Inter-Rater or Inter-Observer Reliability

**SIMPLE**
Number of agreement
----------------------------------
Number of agreement + disagreement

**Cohens Kappa**
Adjusts for chance agreement
Relationship of Agreement to Disagreement in Scores based on Squared Kappa or Percent Agreement Statistics

<table>
<thead>
<tr>
<th>Value</th>
<th>$\nu^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.20</td>
<td>4%</td>
</tr>
<tr>
<td>.40</td>
<td>16%</td>
</tr>
<tr>
<td>.60</td>
<td>36%</td>
</tr>
<tr>
<td>.80</td>
<td>64%</td>
</tr>
<tr>
<td>.90</td>
<td>81%</td>
</tr>
<tr>
<td>1.00</td>
<td>100%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Value of Kappa</th>
<th>Level of Agreement</th>
<th>% of Data that are Reliable</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–.20</td>
<td>None</td>
<td>0–4%</td>
</tr>
<tr>
<td>.21–.39</td>
<td>Minimal</td>
<td>4–15%</td>
</tr>
<tr>
<td>.40–.59</td>
<td>Weak</td>
<td>15–35%</td>
</tr>
<tr>
<td>.60–.79</td>
<td>Moderate</td>
<td>35–63%</td>
</tr>
<tr>
<td>.80–.90</td>
<td>Strong</td>
<td>64–81%</td>
</tr>
<tr>
<td>Above.90</td>
<td>Almost Perfect</td>
<td>82–100%</td>
</tr>
</tbody>
</table>
Interrater reliability between scorers from eight European sleep laboratories in subjects with different sleep disorders

HEIDI DANKER-HOPFE¹, D. KUNZ², G. GRUBER³, G. KLÖSCH⁴, J. L. LORENZO⁵, S. L. HIMANEN⁶, B. KEMP⁷, T. PENZEL⁸, J. RÖSCHKE⁹, H. DORN¹, A. SCHLÖGL¹⁰, E. TRENKER¹¹ and G. DORFFNER¹²
Consistency: Intra Class Correlation (ICC)

- the assessment of consistency or reproducibility of quantitative measurements made by different observers measuring the same quantity

- Since the *intraclass correlation coefficient* gives a composite of intra-observer and inter-observer variability

- Example: How correct can ER doctors predict the severity of disease based on echo results: how consistent the scores are to each other?
What would possibly effect reliability in YOUR research?

(What would possibly effect reliability in research you heard about?)
3. What is validity?
Validity

Test validity is the degree to which an instrument measures what it is supposed to measure.
Does this instrument measure what I want to measure?

- How is the self care?

<table>
<thead>
<tr>
<th>Hvor stor en del af tiden</th>
<th>Hele tiden</th>
<th>Det meste af tiden</th>
<th>Lidt over halvdelen af tiden</th>
<th>Lidt under halvdelen af tiden</th>
<th>Lidt af tiden</th>
<th>På intet tidspunkt</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Har du følt dig trist til moøre, ked af det?</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2 Har du manglet interesse for dine daglige gøremål?</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3 Har du følt at du mangrede energi og kræfter?</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4 Har du haft mindre selvstilling?</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5 Har du haft dårlig samvittighed eller skyldfølelse?</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>6 Har du følt, at livet ikke var værd at leve?</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7 Har du haft besvær med at koncentrere dig, fx at læse avis eller følge med i fjernsyn?</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8a Har du følt dig rastløs?</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8b Har du følt dig mere stille?</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>9 Har du haft besvær med at sove om natten?</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>10a Har du haft nedsat appetit?</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>10b Har du haft øget appetit?</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Not valid

.... To measure self care
Does this instrument measure what I want to measure?

- What is her temperature?

Not valid

.... To measure temperature
Does this instrument measure what I want to measure?

- What is the digoxin level in the blood?

Not valid
.... To measure digoxin level
Experimental examples of Validity

• Adapting a new methodology for a new set of data or samples
  • Golden standard?
  • Species dependent?
Validity

Content validity

Criterion-Related Validity

Construct Validity
Content validity

How much a measure represents every single element of a construct

e.g. A heart failure knowledge test with strong content validity will represent the subjects that are actually related to heart failure rather than asking unrelated questions
Content validity

- Based on judgement
- Expert panels of users?
- Patient interviews
  - What is self care for you?
Criterion-related validity

The degree to which scores on an instrument are correlated with some external criterion

(Polite & Hungler, 1999)
Compare to other criteria

For example

- Measures on a newly developed personality test

Compare to

- Behavioral criterion on which psychologists agree
Construct validity

The degree to which an instrument measures the construct under investigation

(Polite & Hungler, 1999)
Construct validity

- Known groups technique
- Factor analysis
Factor analyses

- Based on classical test theory (CTT) / traditional test theory (TTT)
- Are used to explore or confirm dimensionality in data
- Different forms of factor analyses
  - Principal component analysis
  - Exploratory factor analysis
  - Confirmatory factor analysis
Figure 1.
Exploratory Factor Analysis (EFA)
- A method of data reduction which infers presence of latent factors which are responsible for the shared variance in a set of observed variables / items.
- EFA is by definition ‘exploratory’ – the user does not specify a structure, and assumes each item/ variable could be related to each latent factor.

Confirmatory Factor Analysis (CFA)
- User defines which observed variables /items are related to the specified constructs or latent factors – based on a priori theory or the results of EFA
example of EFA (SCHFI)

Note. Chi-Square = 36.096; DF = 27; p = 0.11; CFI = .971; TLI .952; RMSEA = .035; SRMR = .042. All relationships are significant at p level < 0.001.
What would possibly effect VALIDITY in YOUR research?

(What would possibly effect validity in research you heard about?)
1. Why bother (examples from research)
2. What is reliability?
3. What is validity?
4. How should I treat missing values?
5. I cannot find an instrument: should I develop my own, translate, adapt?
4. How should I treat missing values?

Reasons for missing data

• Participants can fail to respond to questions
• Equipment malfunction
• Power failure
• Sample loss
• Data collecting or recording mechanisms can malfunction
• Subjects can withdraw from studies before they are completed
• Incomplete registry data
• Data entry errors
• Attrition due to social/natural processes
  • Example:
  • Accident of the patients, snowfall during data collection,
  • Patients I not able to drive to the clinic, death
  • Power failure in the lab, sick lab technician
• Skip pattern in survey
  • Example: Certain questions only asked to respondents who indicate they have symptoms (see SCFHI)
• Attrition due to social/natural processes
  • Example: accident, not able to drive to the clinic, death
• Skip pattern in survey
  • Example: Certain questions only asked to respondents who indicate they have symptoms (see SCFHI)
• Random data collection issues
• Respondent refusal/Non-response
  • Is it the same patients who misses all
  • Are there certain item? (e.g. MLwHFQ)
Missing Data Mechanisms

- Missing Completely at Random (MCAR) \( (uniform \ non-response) \)
- Missing at Random (MAR)
- Missing not at Random (NMAR) \( non\-ignorable \)
Missing Data Mechanisms

- Missing Completely at Random (MCAR) (*uniform non-response*)
  - Data missing are **independent** both of observable variables and of unobservable parameters of interest
  - Example: postal questionnaire is missing, sample is dropped
Missing Data Mechanisms

- Missing Completely at Random (MCAR) (*uniform non-response*)
- Missing at Random (MAR)
  - Missingness is related to a particular variable, but it is not related to the value of the variable that has missing data
  - Example: male participants are more likely to refuse to fill out the depression survey, but it does not depend on the level of their depression.
  - Example: technical error of a machine or handling of sample during measurements
Missing Data Mechanisms

- Missing Completely at Random (MCAR) *(uniform non-response)*
- Missing at Random (MAR)
- Missing not at Random (NMAR) *non-ignorable*
  - The probability of a missing value depends on the variable that is missing

Example:
- Respondents with high income less likely to report income
- A particular treatment causes discomfort, a patient is more likely to drop out of the study. This missingness is not at random
4. How to treat missing values/data?

- Deletion Methods
  - Listwise deletion, pairwise deletion
- Imputation?

Methods
- Mean/mode substitution, dummy variable method, single regression
- Model-Based Methods
- Maximum Likelihood, Multiple imputation
Pairwise deletion

- Analysis of all cases in which the variables of interest are present
- Keeps as many cases as possible
- Keeps many information

But

- Different ‘n’ per analysis

<table>
<thead>
<tr>
<th>Gender</th>
<th>8th grade math test score</th>
<th>12th grade math score</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>45</td>
<td>.</td>
</tr>
<tr>
<td>M</td>
<td>99</td>
<td>86</td>
</tr>
<tr>
<td>F</td>
<td>55</td>
<td>88</td>
</tr>
<tr>
<td>F</td>
<td>85</td>
<td>75</td>
</tr>
<tr>
<td>F</td>
<td>80</td>
<td>82</td>
</tr>
<tr>
<td>.</td>
<td>81</td>
<td>82</td>
</tr>
<tr>
<td>F</td>
<td>75</td>
<td>80</td>
</tr>
<tr>
<td>M</td>
<td>95</td>
<td>.</td>
</tr>
<tr>
<td>M</td>
<td>86</td>
<td>90</td>
</tr>
<tr>
<td>F</td>
<td>70</td>
<td>75</td>
</tr>
<tr>
<td>F</td>
<td>85</td>
<td>.</td>
</tr>
</tbody>
</table>
Listwise deletion

- Only analyzes cases with data on each variable
- Simple + comparability in all analysis

But
- Losing power
- Losing information
- Bias if data are not MCAR
IMPUTATION

Rather than removing variables or observations with missing data, another approach is to fill in or “impute” missing values.

Examples

• Last value carried forward
• Mean imputation
• Imputation based on logical rules
• Using regression predictions to perform deterministic imputation
<table>
<thead>
<tr>
<th>Approach</th>
<th>Description</th>
<th>Missingness Pattern</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete cases</td>
<td>Only cases without missing observations in analysis</td>
<td>MCAR&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Basic, single</td>
</tr>
<tr>
<td>Mean imputation</td>
<td>Imputes missing observations with listwise mean for each variable</td>
<td>MCAR&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Basic, single</td>
</tr>
<tr>
<td>LOCF</td>
<td>Imputes the last available observation in the current data collection wave</td>
<td>-</td>
<td>Basic, single</td>
</tr>
<tr>
<td>Regression imputation</td>
<td>Imputes missing observations by prediction based on other variables in a regression model</td>
<td>MAR, MCAR</td>
<td>Advanced, single</td>
</tr>
<tr>
<td>EM imputation</td>
<td>Imputes missing observations using expectation maximization algorithm</td>
<td>MAR, MCAR</td>
<td>Advanced, single</td>
</tr>
<tr>
<td>NORM</td>
<td>Multiple imputes missing observations under a normal model</td>
<td>MAR, MCAR</td>
<td>Advanced, multiple</td>
</tr>
<tr>
<td>MICE</td>
<td>Multiple imputes missing observations using chained equations</td>
<td>MAR, MCAR</td>
<td>Advanced, multiple</td>
</tr>
<tr>
<td>SPSS MI</td>
<td>Multiple imputes missing observations under a normal model in SPSS</td>
<td>MAR, MCAR</td>
<td>Advanced, multiple</td>
</tr>
<tr>
<td>Amelia II</td>
<td>Multiple imputes missing observations using a bootstrapping-based algorithm</td>
<td>MAR, MCAR</td>
<td>Advanced, multiple</td>
</tr>
</tbody>
</table>

**Journal of Medical Internet Research**

Original Paper

**Missing Data Approaches in eHealth Research: Simulation Study and a Tutorial for Nonmathematically InclinedResearchers**

Matthijs Blankers<sup>1,2</sup>, MSc; Maarten W J Koeter<sup>2</sup>, PhD; Gerard M Schippers<sup>1,2</sup>, PhD

<sup>1</sup>Arkin Academy, Amsterdam, The Netherlands  
<sup>2</sup>Amsterdam Institute for Addiction Research (AIAR), Academic Medical Center, University of Amsterdam, Department of Psychiatry, Amsterdam, The Netherlands
5. I cannot find an instrument: should I develop my own, translate, adapt?

Discuss

Advantages

Disadvantages

Consequences
Developing a new questionnaire

*Phase 1*
• Decision on theoretical framework

*Phase 2*
• Item development
  • Systematic review or qualitative analysis of the literature
  • Qualitative interviews of patients, physician, nurses

*Phase 3*
• Pilot testing
• Initial validity and reliability assessment
• Reduction or revision of items or scale structure
• Scoring

*Phase 4*
• Further field testing, more advanced validity and reliability estimation
Translating
Culture and language

- Translation = translation?
- Translation = cultural?
- Translation = validation?
Concerns in translation

• Use of expressions
  • E.g. Thrist scale:
    • ’feels like cotton in the mouth’
Problems in Cross-Cultural Use of the Hospital Anxiety and Depression Scale: “No Butterflies in the Desert”

Gemma A. Maters¹*, Robbert Sanderman¹, Aimee Y. Kim², James C. Coyne¹,³

¹Health Psychology Section, Department of Health Sciences, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.
³Institute for Health, Health Care Policy and Aging Research, Rutgers, the State University of New Jersey, New Brunswick, New Jersey, United States of America.

Abstract

Objective: The Hospital Anxiety and Depression Scale (HADS) is widely used to screen for anxiety and depression. A large literature is citable in support of its validity, but difficulties are increasingly being identified, such as inexplicably discrepant optimal cutpoints and inconsistent factor-structures. This article examines whether these problems could be due to the construction of the HADS that poses difficulties for translation and cross-cultural use.
Can I add /delete, change procedures?
<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>4. If my feet/legs become more swollen than usual I contact my doctor or nurse.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>5. If I gain 2 kg in one week I contact my doctor or nurse.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>6. I limit the amount of fluids I drink (not more than 1½-2 l/day)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>7. I take a rest during the day</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>8. If I experience increased fatigue I contact my doctor or nurse</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>9. I eat a low salt diet</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>10. I take my medication as prescribed</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>11. I get a flu shot every year</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>12. I exercise regularly</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

*The European Heart Failure Self-care Behavior Scale (Jaarsma, Stromberg, Martensson, Dracup, 1999)*
European Heart Failure Self-Care Behavior Scale_9. Russian version.

Эта шкала содержит утверждения по поводу самоконтроля людей с сердечной недостаточностью. Дайте ответ на каждое утверждение, обведите число, которое, на Ваш взгляд, наиболее соответствует Вам. Обратите внимание, что ответы составляют шкалу с диапазоном от "Полностью согласен" (1) до "Совершенно не согласен" (5). Даже если Вы сомневаетесь с ответом на какое-либо утверждение, обведите то число, которое Вы считаете наиболее подходящим для Вас.

<table>
<thead>
<tr>
<th>Полностью согласен/согласна</th>
<th>Совершенно не согласен /не согласна</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Я взвешиваюсь каждый день</td>
<td>2 3 4 5</td>
</tr>
</tbody>
</table>

2. Если моя одышка усиливается,
ОБЩЕСТВО СПЕЦИАЛИСТОВ ПО СЕРЕДНЕЙ НЕДОСТАТОЧНОСТИ

ЕВРОПЕЙСКАЯ ШКАЛА ОЦЕНКИ СПОСОБНОСТИ К САМОПОМОЩИ ПАЦИЕНТОВ С СЕРЕДНЕЙ НЕДОСТАТОЧНОСТЬЮ, ШОСЧН_9
(ЕВРОПЕЙСКАЯ ИНСЕКТНОЕ ФАКУЛЬТЕТ СЕЛЕКЦИИ ИНСЕКТОБЕЕВ СЕЛЕКЦИИ_9)

Эта шкала содержит утверждения по поводу самоконтроля людей с сердечной недостаточностью. Дайте ответ на каждое утверждение, обведя число, которое, на Ваш взгляд, наиболее соответствует Вам. Обратите внимание, что ответы составляют шкалу с диапазоном от «Полностью согласен» (☹☹) до «Совершенно не согласен» (☺☺). Даже если Вы сомневаетесь с ответом на какое-либо утверждение, обведите то число, которое Вы считаете наиболее подходящим для Вас.

<table>
<thead>
<tr>
<th>Утверждение</th>
<th>Полностью согласен/согласна</th>
<th>Совершенно не согласен/не согласен</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Я взвешиваюсь каждый день.</td>
<td>☹☹</td>
<td>☹</td>
</tr>
<tr>
<td>2. Если моя одежда усииливается, я обращаюсь к доктору или медицинской сестре.</td>
<td>☹☹</td>
<td>☹</td>
</tr>
<tr>
<td>3. Если у меня усииливается отеки ступней/ног, я связываюсь с врачом или медсестрой.</td>
<td>☹☹</td>
<td>☹</td>
</tr>
<tr>
<td>4. Если мой вес увеличивается более чем на 2 кг в неделю, я обращаюсь к врачу или медсестре.</td>
<td>☹☹</td>
<td>☹</td>
</tr>
<tr>
<td>5. Я ограничивываю количество выпиваемой жидкости (не более 1,5–2 л/день).</td>
<td>☹☹</td>
<td>☹</td>
</tr>
<tr>
<td>6. Если я испытываю повышенную утомляемость, я обращаюсь к своему врачу или медсестре.</td>
<td>☹☹</td>
<td>☹</td>
</tr>
<tr>
<td>7. Я соблюдаю диету с низким содержанием соли.</td>
<td>☹☹</td>
<td>☹</td>
</tr>
<tr>
<td>8. Я принимаю лекарства так, как мне назначено.</td>
<td>☹☹</td>
<td>☹</td>
</tr>
<tr>
<td>9. Я поддерживаю рекомендованный уровень физической активности.</td>
<td>☹☹</td>
<td>☹</td>
</tr>
</tbody>
</table>
Adding items / changing an instrument

Yes you can do that
• Consider ‘revised version’
• Validity and reliability
• Legal rights / copy rights

Consequences:
• Comparability?
• Reliability, validity
The COSMIN checklist

The COSMIN checklist contains standards for design requirements and preferred statistical methods of studies on the measurement properties of health measurement instruments. The checklist can be used to determine if a study on measurement properties meets the standards for good methodological quality.

We recommend users of the COSMIN checklist to read the COSMIN manual before using the checklist. The manual contains user-friendly data extraction forms and detailed instructions for how to complete the COSMIN checklist. In addition, background information is provided on the development and validation of the checklist and the rationale behind all items. Also possible applications of the COSMIN checklist are described.

In 2011 a COSMIN checklist with 4-point rating scale was developed which can be used to calculate total quality scores per measurement property, e.g. when using the checklist in a systematic review of measurement properties. Read more...

- Download the COSMIN manual
- Download the original COSMIN checklist
- Download the COSMIN checklist with 4-point rating scale

Focus of the checklist

The focus of the checklist is on standards for studies on the measurement properties of Health-Related Patient-Reported Outcomes (HR-PROS). However, the same measurement properties...
Step 3. Determining if a study meets the standards for good methodological quality

<table>
<thead>
<tr>
<th>Box A. Internal consistency</th>
<th>excellent</th>
<th>good</th>
<th>fair</th>
<th>poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  Does the scale consist of effect indicators, i.e. is it based on a reflective model?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Design requirements</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2  Was the percentage of missing items given?</td>
<td>Percentage of missing items described</td>
<td>Percentage of missing items NOT described</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3  Was there a description of how missing items were handled?</td>
<td>Not described but it can be deduced how missing items were handled</td>
<td>Not clear how missing items were handled</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4  Was the sample size included in the internal consistency analysis adequate?</td>
<td>Adequate sample size (≥100)</td>
<td>Good sample size (50-99)</td>
<td>Moderate sample size (30-49)</td>
<td>Small sample size (&lt;30)</td>
</tr>
<tr>
<td>5  Was the unidimensionality of the scale checked? i.e. was factor analysis or IRT model applied?</td>
<td>Factor analysis performed in the study population</td>
<td>Authors refer to another study in which factor analysis was performed in a similar study population</td>
<td>Authors refer to another study in which factor analysis was performed, but not in a similar study population</td>
<td>Factor analysis NOT performed and no reference to another study</td>
</tr>
<tr>
<td>6  Was the sample size included in the unidimensionality analysis adequate?</td>
<td></td>
<td>7* items and</td>
<td>5* items and</td>
<td>5* items but</td>
</tr>
</tbody>
</table>
“Every line is the perfect length if you don't measure it.”

Marty Rubin