The impact of measurement bias on the assessment of change

Calculation of effect-size indices

Mathilde Verdam
Frans Oort & Mirjam Sprangers
University of Amsterdam & Academic Medical Center

SEM meeting Zürich
7-8 april 2016
QUALITY OF LIFE?
Health-related quality of life

- WHO definition of Health (1948):
  “A state of complete physical, mental and social well-being and not merely the absence of disease or infirmity.”

- Not merely ‘objective’ medical outcomes
Health-related quality of life

“Quality of life is regarded as a subjective report of the patients’ experience of disease and treatment.”

• SF-36
  – Physical health
    • Physical functioning, bodily pain, general health, role limitations due to physical health
  – Mental health
    • Mental health, social functioning, vitality, role limitations due to emotional health

De Haes et al. (2012)
Ware et al. (1996)
Structural Equation Modeling

Diagram:

- PHYS
  - PF
  - RP
  - BP
  - GH
- MENT
  - VT
  - MH
  - RE
  - SF

Variables:
- Res. PF
- Res. RP
- Res. BP
- Res. GH
- Res. VT
- Res. MH
- Res. RE
- Res. SF
The impact of measurement bias on the assessment of change
Outline

• Assessment of change in health-related quality of life (HRQL)
• Investigation of measurement bias (or response shift)
• Calculation of effect-size indices using a decomposition of change
• Relation to other effect-size indices
Measurement bias

• Measurement bias / Response shift
  “A change in the frame of reference by which individuals assess their HRQL”

Sprangers & Schwartz (1999)
Structural Equation Modeling

• Measurement bias detection
  – Intercepts
  – Factor loadings
  – Residual variances

Oort (2005)
Measurement bias detection

• Recalibration
  A change in respondents’ internal standard of measurement

➢ Intercepts (uniform)
➢ Residual variances

Oort (2005)
Measurement bias detection

• Reprioritization
  A change in respondents’ values regarding the relative importance of subdomains

➢ Factor loadings (size)

Oort (2005)
Measurement bias detection

• Reconceptualization
  A change in definition of the target construct

➢ Factor loadings (pattern)

Oort (2005)
Measurement bias detection

• Detect response shift / measurement bias
  – Reconceptualization
  – Reprioritization
  – Recalibration

• Take into account measurement bias

• A more valid assessment of change

Oort (2005)
Measurement bias detection

A more valid assessment of change

• But what is the impact of potential response shifts on the assessment of change??

→ Is ‘more valid’ also ‘more informative’?
Measurement bias detection

Assessment of significance
- Chi-square difference test
- Significance of model parameters

Assessment of relevance
- Impact on the assessment of change?
  \[\rightarrow\] Comparing common factor means before/after bias detection
  \[\rightarrow\] Effect-size indices using a decomposition of change
Decomposition of change

\[ \mu_{\text{post}} - \mu_{\text{pre}} = \Lambda_{\text{pre}} \alpha_{\text{post}} + (\tau_{\text{post}} - \tau_{\text{pre}}) + (\Lambda_{\text{post}} - \Lambda_{\text{pre}}) \alpha_{\text{post}} \]

- Change due to changes in common factor means
- Change due to changes in factor loadings
- Change in means of the indicators
- Change due to changes in intercepts
- Observed change = True change + Recalibration + (Reprioritization & Reconceptualization)

Residual variances do not feature in the mean structure
Decomposition of change

\[ \mu_{\text{post}} - \mu_{\text{pre}} = \Lambda_{\text{pre}} \alpha_{\text{post}} + (\tau_{\text{post}} - \tau_{\text{pre}}) + (\Lambda_{\text{post}} - \Lambda_{\text{pre}}) \alpha_{\text{post}} \]

Observed change = True change + Recalibration + (Reprioritization & Reconceptualization)

Calculation of effect-size indices

Cohen’s \( d = \frac{\bar{x}_2 - \bar{x}_1}{S_{x_2-x_1}} \)

Using SEM estimates:

\[ \frac{\hat{\mu}_{\text{post}} - \hat{\mu}_{\text{pre}}}{\hat{\sigma}_{\text{post-pre}}} = \frac{\hat{\mu}_{\text{post}} - \hat{\mu}_{\text{pre}}}{\sqrt{\hat{\sigma}_{\text{post}}^2 + \hat{\sigma}_{\text{pre}}^2 - 2\hat{\sigma}_{\text{post,pre}}}} \]
Decomposition of change

\[ \mu_{\text{post}} - \mu_{\text{pre}} = \Lambda_{\text{pre}} \alpha_{\text{post}} + (\tau_{\text{post}} - \tau_{\text{pre}}) + (\Lambda_{\text{post}} - \Lambda_{\text{pre}}) \alpha_{\text{post}} \]

→ Contribution to change in terms of effect-size indices
Application in HRQL

**Sample:** 170 newly diagnosed cancer patients undergoing invasive surgery. 87 men and 83 women. Ages ranging from 27 to 83 (M = 57.5, SD=14.1).

**Procedure:** Questionnaires were administered prior to surgery (pre-test), and three months following surgery (post-test).
Application in HRQL
Application in HRQL
### Decomposition of change

<table>
<thead>
<tr>
<th>Scale</th>
<th>Observed change</th>
<th>True change</th>
<th>Recal RS</th>
<th>Repri RS</th>
<th>Recon RS</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF</td>
<td>-0.51**</td>
<td>-0.51**</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RP</td>
<td>-0.28**</td>
<td>-0.47**</td>
<td>0.19**</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BP</td>
<td>-0.25**</td>
<td>-0.42**</td>
<td>0.17**</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SF</td>
<td>-0.09</td>
<td>0.01</td>
<td>-</td>
<td>-0.10*</td>
<td>-</td>
</tr>
<tr>
<td>MH</td>
<td>0.37**</td>
<td>0.37**</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RE</td>
<td>0.26**</td>
<td>0.26**</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>GH</td>
<td>-0.01</td>
<td>-0.15**</td>
<td>-</td>
<td>-</td>
<td>0.14**</td>
</tr>
<tr>
<td>VT</td>
<td>-0.31**</td>
<td>-0.31**</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>FT</td>
<td>-0.32**</td>
<td>-0.32**</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**General Physical Health:**
\[ d = -0.51 \] (\( d = -0.46 \))

**General Mental Health:**
\[ d = 0.39 \] (\( d = 0.33 \))

**General Fitness:**
\[ d = -0.34 \] (\( d = -0.33 \))
### Decomposition of change

<table>
<thead>
<tr>
<th>Scale</th>
<th>Observed change</th>
<th>True change</th>
<th>Recal RS</th>
<th>Repri RS</th>
<th>Recon RS</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF</td>
<td>-0.51**</td>
<td>-0.51**</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RP</td>
<td>-0.28**</td>
<td>-0.47**</td>
<td>0.19**</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BP</td>
<td>-0.25**</td>
<td>-0.42**</td>
<td>0.17**</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SF</td>
<td>-0.09</td>
<td>0.01</td>
<td>-</td>
<td>-0.10*</td>
<td>-</td>
</tr>
<tr>
<td>MH</td>
<td>0.37**</td>
<td>0.37**</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RE</td>
<td>0.26**</td>
<td>0.26**</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>GH</td>
<td>-0.01</td>
<td>-0.15**</td>
<td>-</td>
<td>-</td>
<td>0.14**</td>
</tr>
<tr>
<td>VT</td>
<td>-0.31**</td>
<td>-0.31**</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>FT</td>
<td>-0.32**</td>
<td>-0.32**</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**General Physical Health:**
- $d = -0.51$ ($d = -0.46$)

**General Mental Health:**
- $d = 0.39$ ($d = 0.33$)

**General Fitness:**
- $d = -0.34$ ($d = -0.33$)
## Decomposition of change

Patients score higher on RP and BP after treatment, as compared to the other indicators of general physical health ($d = .19$, $d = .17$).

<table>
<thead>
<tr>
<th>Scale</th>
<th>Observed change</th>
<th>True change</th>
<th>Recal RS</th>
<th>Repri RS</th>
<th>Recon RS</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF</td>
<td>-0.51**</td>
<td>-0.51**</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RP</td>
<td>-0.28**</td>
<td>-0.47**</td>
<td>0.19**</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BP</td>
<td>-0.25**</td>
<td>-0.42**</td>
<td>0.17**</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SF</td>
<td>-0.09</td>
<td>0.01</td>
<td>-</td>
<td>-0.10*</td>
<td>-</td>
</tr>
<tr>
<td>MH</td>
<td>0.37**</td>
<td>0.37**</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RE</td>
<td>0.26**</td>
<td>0.26**</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>GH</td>
<td>-0.01</td>
<td>-0.15**</td>
<td>-</td>
<td>-</td>
<td>0.14**</td>
</tr>
<tr>
<td>VT</td>
<td>-0.31**</td>
<td>-0.31**</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>FT</td>
<td>-0.32**</td>
<td>-0.32**</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

- **General Physical Health:** $d = -0.51$ ($d = -0.46$)
- **General Mental Health:** $d = 0.39$ ($d = 0.33$)
- **General Fitness:** $d = -0.34$ ($d = -0.33$)
## Decomposition of change

<table>
<thead>
<tr>
<th>Scale</th>
<th>Observed change</th>
<th>True change</th>
<th>Recal RS</th>
<th>Repri RS</th>
<th>Recon RS</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF</td>
<td>-0.51**</td>
<td>-0.51**</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RP</td>
<td>-0.28**</td>
<td>-0.47**</td>
<td>0.19**</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BP</td>
<td>-0.25**</td>
<td>-0.42**</td>
<td>0.17**</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SF</td>
<td>-0.09</td>
<td>0.01</td>
<td>-</td>
<td><strong>-0.10</strong></td>
<td>-</td>
</tr>
<tr>
<td>MH</td>
<td>0.37**</td>
<td>0.37**</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RE</td>
<td>0.26**</td>
<td>0.26**</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>GH</td>
<td>-0.01</td>
<td>-0.15**</td>
<td>-</td>
<td>-</td>
<td>0.14**</td>
</tr>
<tr>
<td>VT</td>
<td>-0.31**</td>
<td>-0.31**</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>FT</td>
<td>-0.32**</td>
<td>-0.32**</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

→ Patients SF becomes more important to the measurement of general physical health after treatment ($d = -.10$)

**General Physical Health:**

$d = -0.51$ ($d = -0.46$)

**General Mental Health:**

$d = 0.39$ ($d = 0.33$)

**General Fitness:**

$d = -0.34$ ($d = -0.33$)
## Decomposition of change

<table>
<thead>
<tr>
<th>Scale</th>
<th>Observed change</th>
<th>True change</th>
<th>Recal RS</th>
<th>Repri RS</th>
<th>Recon RS</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF</td>
<td>-0.51**</td>
<td>-0.51**</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RP</td>
<td>-0.28**</td>
<td>-0.47**</td>
<td>0.19**</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BP</td>
<td>-0.25**</td>
<td>-0.42**</td>
<td>0.17**</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SF</td>
<td>-0.09</td>
<td>0.01</td>
<td>-</td>
<td>-0.10*</td>
<td>-</td>
</tr>
<tr>
<td>MH</td>
<td>0.37**</td>
<td>0.37**</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RE</td>
<td>0.26**</td>
<td>0.26**</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>GH</td>
<td>-0.01</td>
<td>-0.15**</td>
<td>-</td>
<td>-</td>
<td>0.14**</td>
</tr>
<tr>
<td>VT</td>
<td>-0.31**</td>
<td>-0.31**</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>FT</td>
<td>-0.32**</td>
<td>-0.32**</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

- **General Physical Health:**
  \[ d = -0.51 \ (d = -0.46) \]

- **General Mental Health:**
  \[ d = 0.39 \ (d = 0.33) \]

- **General Fitness:**
  \[ d = -0.34 \ (d = -0.33) \]

→ Patients GH becomes indicative of the measurement of general mental health after treatment \( d = .14 \)
Decomposition of change

\[ \mu_{post} - \mu_{pre} = \Lambda_{pre} \alpha_{post} + (\tau_{post} - \tau_{pre}) + (\Lambda_{post} - \Lambda_{pre}) \alpha_{post} \]

Dependent on change in common factor

→ Impact may differ over samples with different amount of change in the underlying common factors
Decomposition of change

\[ \mu_{\text{post}} - \mu_{\text{pre}} = \Lambda_{\text{pre}} \alpha_{\text{post}} + (\tau_{\text{post}} - \tau_{\text{pre}}) + (\Lambda_{\text{post}} - \Lambda_{\text{pre}}) \alpha_{\text{post}} \]

Significance (CI’s) of decomposition difficult to calculate

→ Using estimated SE’s from SEM program (Sobel’s test)?
→ Regard chi-square test / significance parameter as sufficient?
Relation to other effect-sizes

Cohen’s $d$
Intuitive / Interpretable?

Other effect-size indices
- Common Language Effect Size (CLES)
- Success Rate Difference (SRD)
- Number Needed to Treat (NNT)

Other suggestions?
Relation to other effect-sizes

**Common Language Effect Size (CLES)** = P(post > pre)
→ The probability that a random sampled person scores better at post-assessment than at pre-assessment

**Success Rate Difference (SRD)** = P(post > pre) – P(post < pre)
→ Net probability that someone scores better at post-assessment as compared to pre-assessment

**Number Needed to Treat (NNT)** = 1 / SRD
→ Number of patients that need to be treated to have one person score better at post-assessment as compared to pre-assessment
Relation to other effect-sizes

<table>
<thead>
<tr>
<th>Cohen’s $d$</th>
<th>CLES</th>
<th>SRD</th>
<th>NNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.50</td>
<td>0.00</td>
<td>$\infty$</td>
</tr>
<tr>
<td>0.1</td>
<td>0.54</td>
<td>0.08</td>
<td>12.6</td>
</tr>
<tr>
<td>0.2</td>
<td>0.58</td>
<td>0.16</td>
<td>6.31</td>
</tr>
<tr>
<td>0.3</td>
<td>0.62</td>
<td>0.24</td>
<td>4.24</td>
</tr>
<tr>
<td>0.4</td>
<td>0.66</td>
<td>0.31</td>
<td>3.22</td>
</tr>
<tr>
<td>0.5</td>
<td>0.69</td>
<td>0.38</td>
<td>2.61</td>
</tr>
<tr>
<td>0.6</td>
<td>0.73</td>
<td>0.45</td>
<td>2.21</td>
</tr>
<tr>
<td>0.7</td>
<td>0.76</td>
<td>0.52</td>
<td>1.94</td>
</tr>
<tr>
<td>0.8</td>
<td>0.79</td>
<td>0.58</td>
<td>1.74</td>
</tr>
<tr>
<td>0.9</td>
<td>0.82</td>
<td>0.63</td>
<td>1.58</td>
</tr>
<tr>
<td>1.0</td>
<td>0.84</td>
<td>0.68</td>
<td>1.46</td>
</tr>
<tr>
<td>$\infty$</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Converting Cohen’s $d$ to $z$:  
$$z = d / \sqrt{2} / \sqrt{1-r} \quad (\text{if } sd = sd_{pooled})$$

Rules of thumb apply to correlations between measurements of 0.5
Discussion

Clinically meaningful?

- “Remarkably universality” among estimates of clinical significance that centre around +/- Cohen’s $d$ of 0.5
- Recommendation to use Cohen’s $d$ as a measure of responsiveness to ensure interpretability and comparability
- CLES preferred to develop insights, whereas NNT most intuitive to interpret clinical significance

Effect-size indices are not a panacea

Norman, Sloan, & Wyrwich (2003)
Norman, Wyrwich, & Patrick (2007)
Kraemer & Kupfer (2006)
Questions / Suggestions?


Norman, G. et al. (2003). Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Medical Care, 41, 582-292.

