Using SEM to detect measurement bias in longitudinal data: An application to the measurement of change in the quality-of-life of cancer patients

B.L. King-Kallimanis, F.J. Oort, M.R.M. Visser, M.A.G. Sprangers
Introduction

- Self-report questionnaires are used to assess change in respondents over time in quality of life research

- Key assumption:
  - respondents interpretation of the questionnaire is consistent over time, or that the questionnaire is measurement invariant.
Response Shift

- Response shift is a special case of measurement bias and occurs when investigating the change in an attribute longitudinally.

- Can be found in either the measurement or explanation of health related quality of life.
General Health Item for the SF-36

My health is excellent

Is this true change in Quality of Life or is it Response Shift?
Objective

• To demonstrate how SEM can be used to detect measurement bias in longitudinal data with the SF-36.

• To investigate whether “violator” variables also have the potential to cause measurement bias on the SF-36.
Methods

- N = 202 cancer patients
- Two time points -> pre and post surgery

- SF-36 measure of HRQL
  - Eight domains (item composites from 36 items)
  - Two factors -> Physical and Mental Health

- Violator variables:
  - cancer site, health status, sex, age, optimism, and social comparison.
Statistical Analysis

A three step procedure was used to investigate bias/response shift in the structural model

- **Step 1:** Establish a measurement model

- **Step 2:** Investigate response shift by examining measurement (in)variance in HRQL scales across time

- **Step 3:** Investigate measurement bias/response shift by examining violator variables
## Results

<table>
<thead>
<tr>
<th>Factor loadings</th>
<th>Before surgery</th>
<th>After surgery</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PHYS HRQL</td>
<td>MENT HRQL</td>
</tr>
<tr>
<td>PF</td>
<td>0.47</td>
<td></td>
</tr>
<tr>
<td>RP</td>
<td>0.61</td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>SF</td>
<td>0.27</td>
<td>0.48</td>
</tr>
<tr>
<td>MH</td>
<td></td>
<td>0.72</td>
</tr>
<tr>
<td>RE</td>
<td></td>
<td>1.10</td>
</tr>
<tr>
<td>VT</td>
<td>0.36</td>
<td>0.30</td>
</tr>
<tr>
<td>GH</td>
<td>0.17</td>
<td></td>
</tr>
</tbody>
</table>
Measurement Model

\[ \chi^2(84) = 113.67 \]
\[ RMSEA = 0.042, 90\% CI (0.019 ; 0.060) \]

\[ \chi^2(202) = 295.51 \]
\[ RMSEA = 0.048, 90\% CI (0.036 ; 0.059) \]

All observed variables covary across measurement occasions.
# Results

<table>
<thead>
<tr>
<th>Measurement Bias</th>
<th>Before surgery</th>
<th>After surgery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age → PF</td>
<td>-0.22</td>
<td>-0.22</td>
</tr>
<tr>
<td><strong>Response Shift</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimism → GH</td>
<td><strong>0.21</strong></td>
<td>0.00</td>
</tr>
<tr>
<td>Sex (female) → GH</td>
<td>0.00</td>
<td><strong>0.36</strong></td>
</tr>
<tr>
<td>Upward Comp. → GH</td>
<td>0.00</td>
<td><strong>-0.19</strong></td>
</tr>
</tbody>
</table>

\[ \chi^2 (218) = 326.03 \]

RMSEA = 0.050, 90%CI (0.038 ; 0.061)
Lung cancer  Panc. cancer  Esop. cancer  Health state 1  Health state 2

Age  Optim.  Social comp.  Sex

Observed measurement variables (X)

PHYS HRQL 1
M ENT HRQL 1

PHYS HRQL 2
M ENT HRQL 2

Attributes (A)

Sex

Lung cancer
Panc. cancer
Esop. cancer
Health state 1
Health state 2

Age
Optim.
Social comp.

PHYS HRQL 1
M ENT HRQL 1

PHYS HRQL 2
M ENT HRQL 2

PF1  RP1  BP1  SF1  VT1  GH1  M H1  RE1

PF2  RP2  BP2  SF2  VT2  GH2  M H2  RE2

Observed measurement variables (X)
Discussion

- Detection and interpretation of bias and response shift is not always as straightforward as it may first appear.
Physical Functioning items include:
- Ability to walk a block
- Climb a flight of stairs
- Carry groceries
Discussion

• Detection and interpretation of bias and response shift is not always as straightforward as it may first appear

• Order of modifications to the model also introduces a source of subjective decision making

• General items appear more prone to bias
  ▫ For example, the General Health domain has very broad items, different groups of respondents appear to attach different meanings to these items