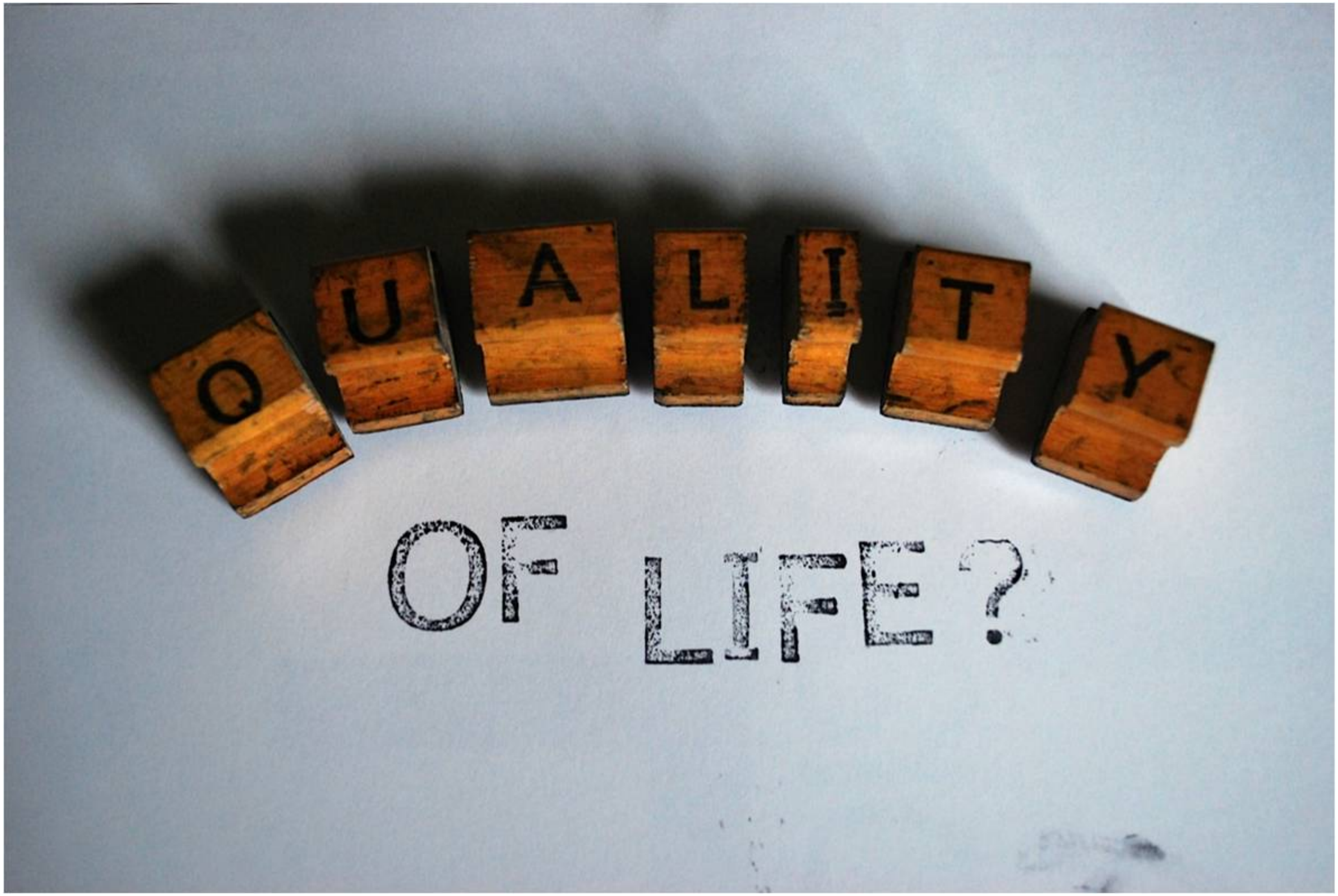
The impact of measurement bias on the assessment of change

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

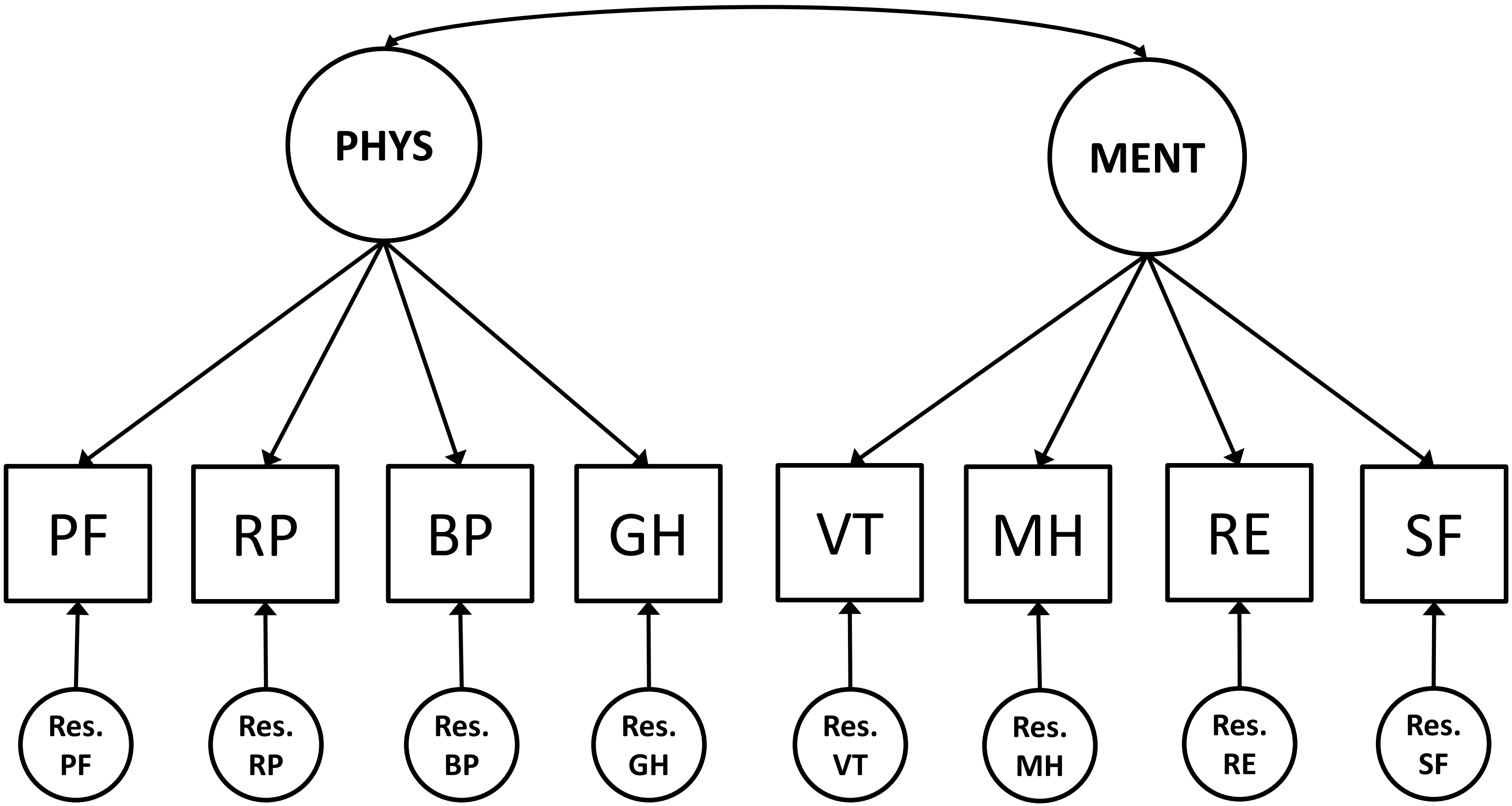
Health-related quality of life

"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



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

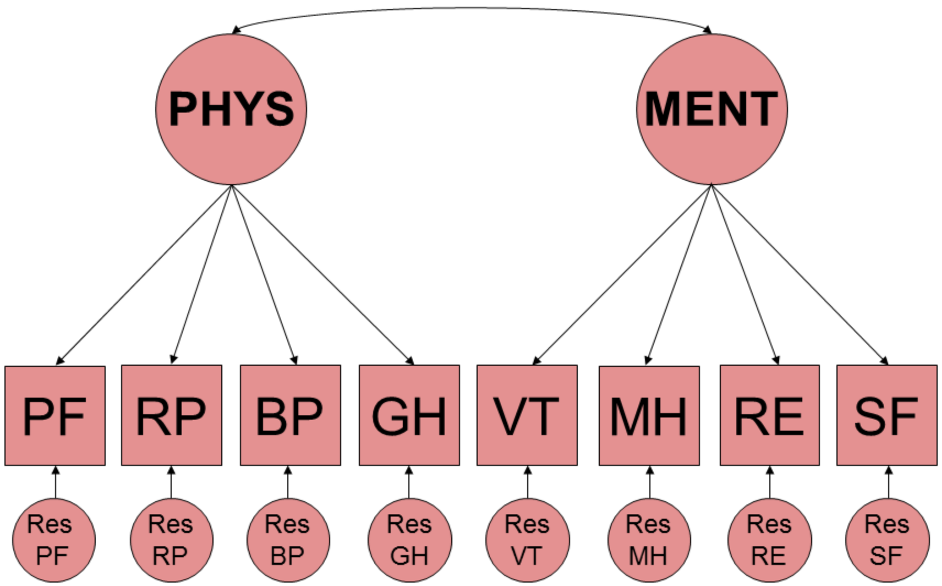
- Measurement bias / Response shift

  “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)

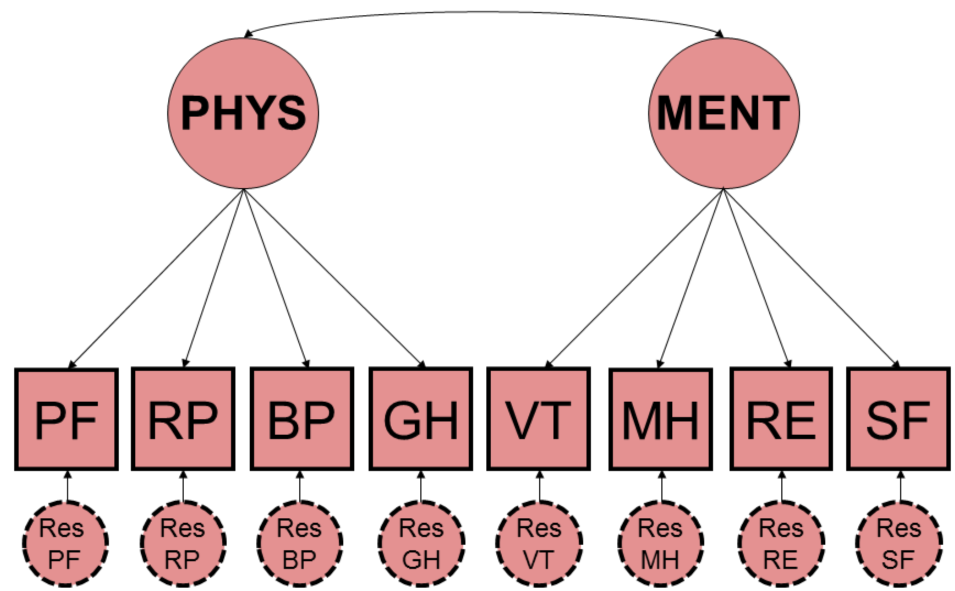
Measurement bias detection

- Recalibration

  A change in respondents' internal standard of measurement

➢ Intercepts (uniform)

➢ *Residual variances*



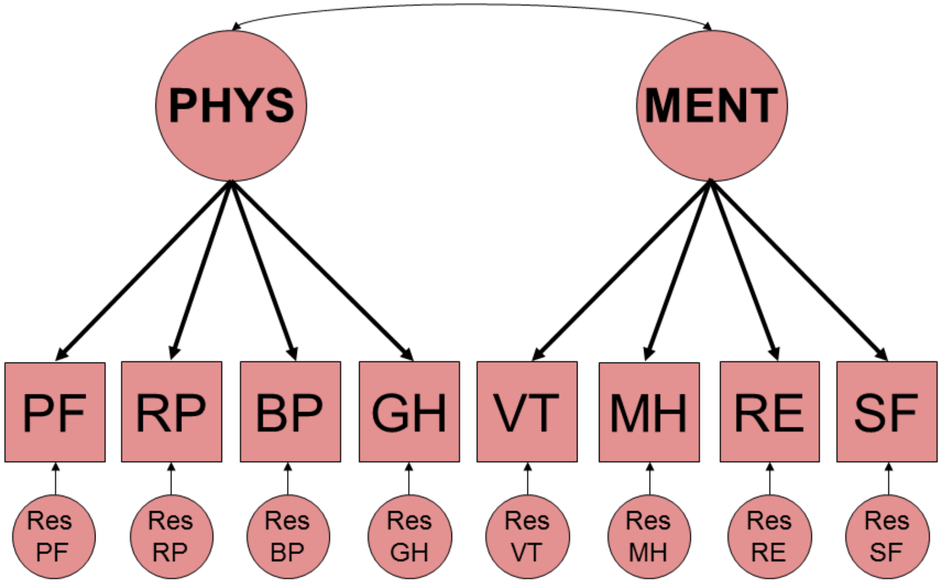
Oort (2005)

Measurement bias detection

- Reprioritization

  A change in respondents' values regarding the relative importance of subdomains

➢ Factor loadings (size)



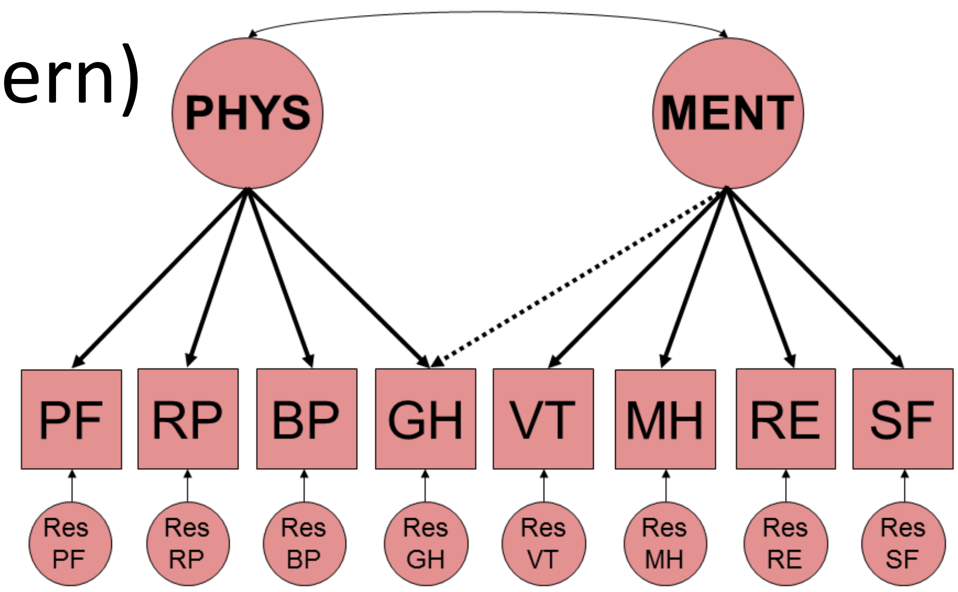
Oort (2005)

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?

→ Comparing common factor means before/after bias detection

→ Effect-size indices using a decomposition of change

Decomposition of change

Change due to changes in common factor means

Change due to changes in factor loadings

$$\mu_{post} - \mu_{pre} = \Lambda_{pre}\,\alpha_{post} + (\tau_{post} - \tau_{pre}) + (\Lambda_{post} - \Lambda_{pre})\,\alpha_{post}$$

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* = $\dfrac{\bar{x}_2 - \bar{x}_2}{s_{x_2 - x_1}}$

Using SEM estimates : $\dfrac{\hat{\mu}_{post} - \hat{\mu}_{pre}}{\hat{\sigma}_{post-pre}} = \dfrac{\hat{\mu}_{post} - \hat{\mu}_{pre}}{\sqrt{\hat{\sigma}^2_{post} + \hat{\sigma}^2_{pre} - 2\hat{\sigma}_{post,pre}}}$

Decomposition of change

$$\frac{\mu_{\text{post}} - \mu_{\text{pre}}}{\sqrt{\hat{\sigma}^2_{post} + \hat{\sigma}^2_{pre} - 2\hat{\sigma}_{post,pre}}} = \frac{\Lambda_{\text{pre}}\,\alpha_{\text{post}}}{\sqrt{\hat{\sigma}^2_{post} + \hat{\sigma}^2_{pre} - 2\hat{\sigma}_{post,pre}}} + \frac{(\tau_{\text{post}} - \tau_{\text{pre}})}{\sqrt{\hat{\sigma}^2_{post} + \hat{\sigma}^2_{pre} - 2\hat{\sigma}_{post,pre}}} + \frac{(\Lambda_{\text{post}} - \Lambda_{\text{pre}})\,\alpha_{\text{post}}}{\sqrt{\hat{\sigma}^2_{post} + \hat{\sigma}^2_{pre} - 2\hat{\sigma}_{post,pre}}}$$

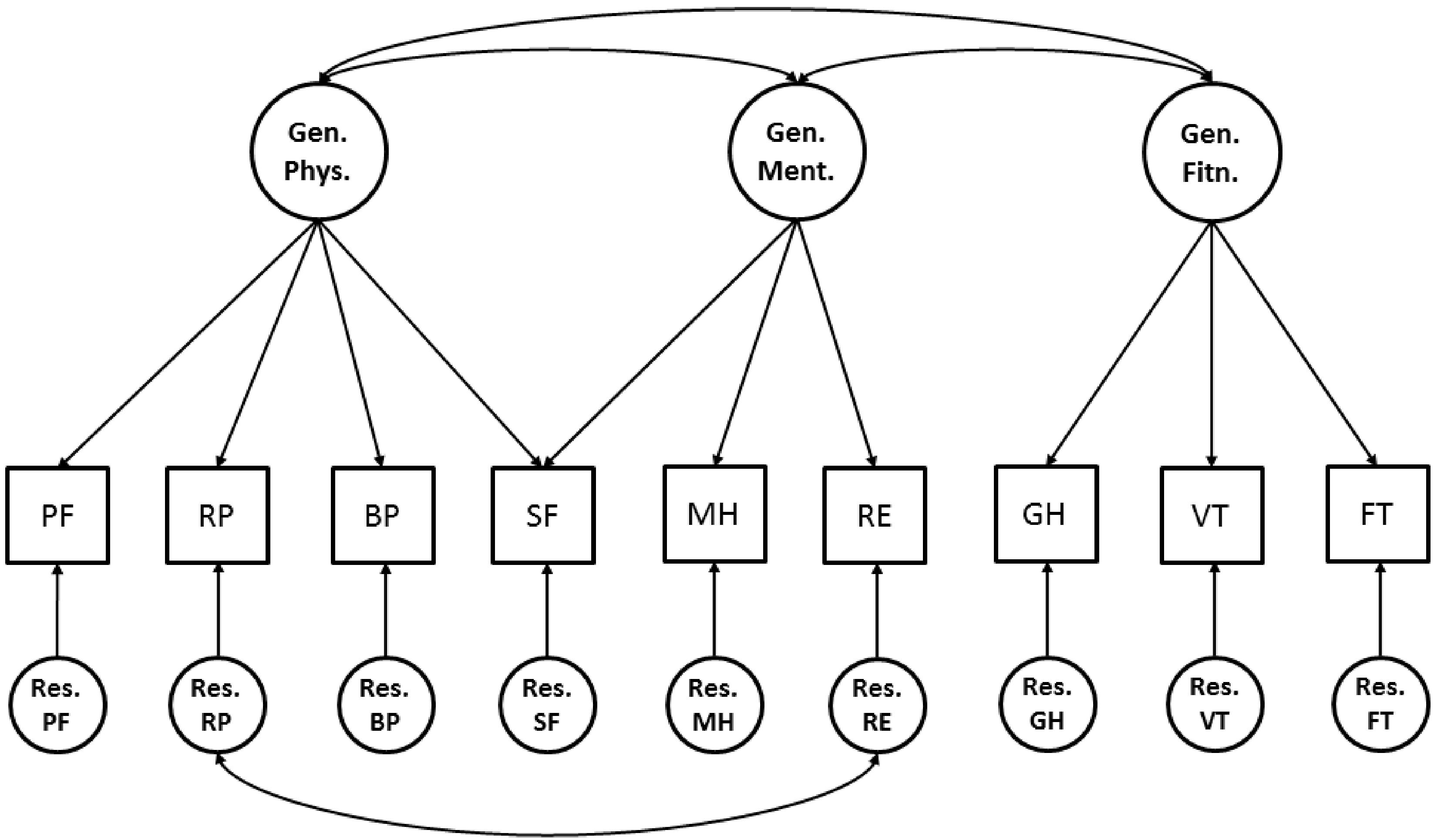
→ <u>Contribution to change in terms of effect-size indices</u>

Application in HRQL

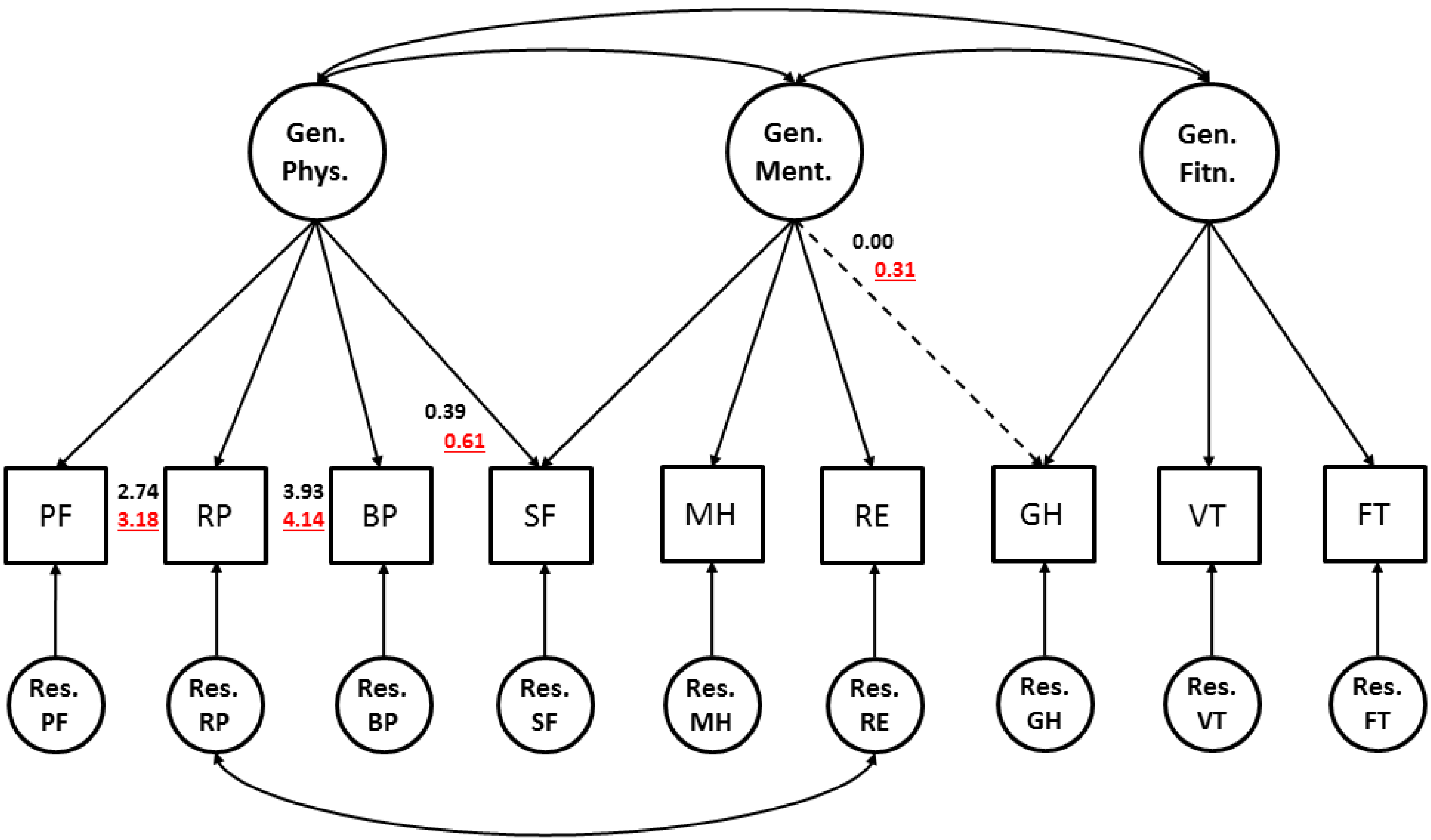
<u>Sample</u>: 170 newly diagnosed cancer patients undergoing invasive surgery. 87 men and 83 women. Ages ranging from 27 to 83 (M = 57.5, SD=14.1).

<u>Procedure</u>: Questionnaires were administered prior to surgery (pre-test), and three months following surgery (post-test)

Application in HRQL



Application in HRQL



Decomposition of change

Scale	Observed change	True change	Recal RS	Repri RS	Recon RS
PF	-0.51**	-0.51**	-	-	-
RP	-0.28**	-0.47**	0.19**	-	-
BP	-0.25**	-0.42**	0.17**	-	-
SF	-0.09	0.01	-	-0.10*	-
MH	0.37**	0.37**	-	-	-
RE	0.26**	0.26**	-	-	-
GH	-0.01	-0.15**	-	-	0.14**
VT	-0.31**	-0.31**	-	-	-
FT	-0.32**	-0.32**	-	-	-

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

Scale	Observed change	True change	Recal RS	Repri RS	Recon RS
PF	-0.51**	-0.51**	-	-	-
RP	-0.28**	-0.47**	0.19**	-	-
BP	-0.25**	-0.42**	0.17**	-	-
SF	-0.09	0.01	-	-0.10*	-
MH	0.37**	0.37**	-	-	-
RE	0.26**	0.26**	-	-	-
GH	-0.01	-0.15**	-	-	0.14**
VT	-0.31**	-0.31**	-	-	-
FT	-0.32**	-0.32**	-	-	-

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

Scale	Observed change	True change	Recal RS	Repri RS	Recon RS
PF	-0.51**	-0.51**	-	-	-
RP	-0.28**	-0.47**	0.19**	-	-
BP	-0.25**	-0.42**	0.17**	-	-
SF	-0.09	0.01	-	-0.10*	-
MH	0.37**	0.37**	-	-	-
RE	0.26**	0.26**	-	-	-
GH	-0.01	-0.15**	-	-	0.14**
VT	-0.31**	-0.31**	-	-	-
FT	-0.32**	-0.32**	-	-	-

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

Decomposition of change

Scale	Observed change	True change	Recal RS	Repri RS	Recon RS
PF	-0.51**	-0.51**	-	-	-
RP	-0.28**	-0.47**	0.19**	-	-
BP	-0.25**	-0.42**	0.17**	-	-
SF	-0.09	0.01	-	-0.10*	-
MH	0.37**	0.37**	-	-	-
RE	0.26**	0.26**	-	-	-
GH	-0.01	-0.15**	-	-	0.14**
VT	-0.31**	-0.31**	-	-	-
FT	-0.32**	-0.32**	-	-	-

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

Decomposition of change

Scale	Observed change	True change	Recal RS	Repri RS	Recon RS
PF	-0.51**	-0.51**	-	-	-
RP	-0.28**	-0.47**	0.19**	-	-
BP	-0.25**	-0.42**	0.17**	-	-
SF	-0.09	0.01	-	-0.10*	-
MH	0.37**	0.37**	-	-	-
RE	0.26**	0.26**	-	-	-
GH	-0.01	-0.15**	-	-	0.14**
VT	-0.31**	-0.31**	-	-	-
FT	-0.32**	-0.32**	-	-	-

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

$$\mu_{post} - \mu_{pre} = \Lambda_{pre}\, \alpha_{post} + (\tau_{post} - \tau_{pre}) + (\Lambda_{post} - \Lambda_{pre})\, \alpha_{post}$$

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?

$$d = \frac{\hat{\mu}_{post} - \hat{\mu}_{pre}}{\hat{\sigma}_{post-pre}} = \frac{\hat{\mu}_{post} - \hat{\mu}_{pre}}{\sqrt{\hat{\sigma}^2_{post} + \hat{\sigma}^2_{pre} - 2\hat{\sigma}_{post,pre}}}$$

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

Cohen's *d*	CLES	SRD	NNT
0.0	0.50	0.00	∞
0.1	0.54	0.08	12.6
0.2	0.58	0.16	6.31
0.3	0.62	0.24	4.24
0.4	0.66	0.31	3.22
0.5	0.69	0.38	2.61
0.6	0.73	0.45	2.21
0.7	0.76	0.52	1.94
0.8	0.79	0.58	1.74
0.9	0.82	0.63	1.58
1.0	0.84	0.68	1.46
∞	1.00	1.00	1.00

Converting Cohen's d to z:

$z = d / \sqrt{2} / \sqrt{1-r}$

(if $sd = sd_{pooled}$)

Rules of thumb apply to correlations between measurements of 0.5

Discussion

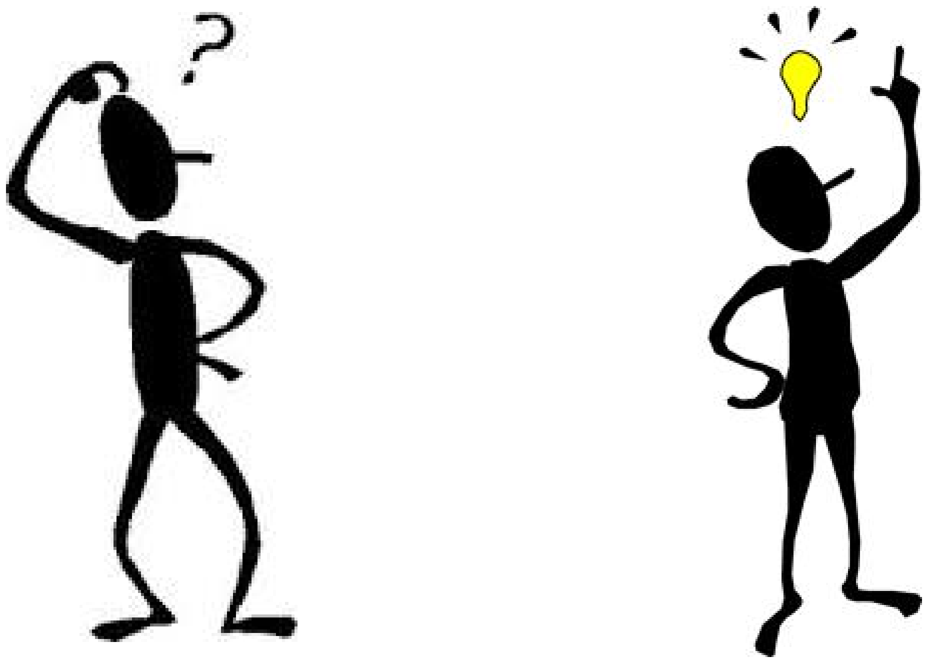
Clinically meaningful?

- “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 interpretabilty and comparability
- CLES preferred to develop insights, whereas NNT most intuitive to interpet clinical significance

<u>Effect-size indices are not a panacea</u>

Norman, Sloan, & Wyrwich (2003)
Norman, Wyrwich, & Patrick (2007)
Kraemer & Kupfer (2006)

Questions / Suggestions?



References

De Haes, J. C. J. M., Olschewski, M., Fayers P., Visser, M. R. M., Cull, A., Hopwood, P., & Sanderman, R. (2012). *Measuring the quality of life of cancer patients with the Rotterdam Symptom Checklist (RSCL): A manual.* 2nd Ed. Northern Centre for Healthcare Research.

Kraemer, H. C., & Kupfer, D. J. (2006). Size of treatment effects and their importance to clinical research and practice. *Biological psychiatry, 59,* 990-996.

Norman, G. R., Wyrwich, K. W., & Patrick, D. L. (2007). The mathematical relationship among different forms of responsiveness coefficients. *Quality of Life Research, 16*, 815-822.

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.

Oort, F. J. (2005). Using structural equation modeling to detect response shift and true change. *Quality of Life Research, 14, 587-598*.

Sprangers, M. A. G., & Schwartz, C. E. (1999). Integrating response shift into health-related quality of life research: a theoretical model. *Social Science & Medicine*, 48, 1507-1515.

Ware, J. E., Snow, K. K., Kosinski M., & Gandek, B. (1993). *SF-36 health survey: Manual and interpretation guide*. Boston, MA: The Health Institute, New England Medical Center.