Identification of response shift at the item level on Patient-reported outcomes changes using Rasch Measurement Theory models

The study has received ethics approval and is supported by a grant from the Institut National du Cancer, under reference ‘INCA_6931’

Karima Hammas, Myriam Blanchin, Jean-Benoit Hardouin, Véronique Sébille

veronique.sebille@univ-nantes.fr

INSERM U1246 SPHERE “methodS in Patient-centered outcomes & HEalth ResEarch” Nantes & Tours Universities, France

http://www.sphere-nantes.fr/
Background

- A better understanding of the experience of patients using **Patient reported outcomes (PRO)** is essential to assess the effectiveness of health care.

- **PRO** is directly reported by the patients without interpretation of their responses by a clinician or anyone else.
  - Measures of perceived health, QoL, fatigue, well-being...

- **Measurement and interpretation of PRO** involves conceptual, methodological, interpretational, and practical issues.
Background

- **Measurement and interpretation of PRO – Some issues**
  - Cognitive processes involved in completing PRO are **complex**
  - PRO are often **multidimensional** with multiple items and dimensions
  - **Missing data** are often **non-ignorable** (e.g. patients might be too tired to fill in the fatigue questionnaire)
  - **Measurement non-invariance**
    - Patients might not respond to PRO consistently & might not be comparable between groups (Differential Item Functioning, **DIF**) and over time (Response Shift, **RS**)
Measurement non-invariance

- **Measurement non-invariance** between groups & over time
  - PRO data include *patients' perceptions of the items* which cannot be directly measured but can influence their responses

- **DIF**: perception varies between groups (*e.g. gender, age*) ⇒ can alter the properties of questionnaires such as reliability, validity or ability to detect “true” differences

- **RS**: perception varies over time (*e.g. change in meaning, in life priorities*) ⇒ erroneous conclusions for the detection of “true” change
Response Shift & clinical interpretation

Should Response Shift be only reduced to measurement bias?

Example: in the context of cancer

- Likely that patients might regularly adapt to their illness ⇒ might give different answers to the questionnaires over time...

- Not only because their health has changed, but also because their perception of what QoL means to them has changed

Methods for Response Shift

- Mostly "sample-level" methods
  - *Dimension-level* ⇒ *e.g.* Structural equation Modeling (SEM), Classification and Regression Trees (CART)
  - *Item-level* ⇒ *e.g.* Rasch Measurement Theory / Item Response Theory (RMT/IRT)

- Assume that the **whole sample** has been affected by the **same changes** in the perception of QoL over time BUT among a sample
  - Only **some individuals** might be affected by RS, ≠ types of RS might affect ≠ individuals to ≠ extent, might depend on known or unknown covariates

Methods for Response Shift

Alternative approaches

- **Dimension or item-level subgroup RS analyses**
  - Multigroup SEM, Growth Mixture Modelling (GMM) ± SEM, RMT and Guttman errors

- **Pros and cons**
  - Investigating differential RS in subgroups ± DIF simultaneously
  - Known and unknown covariates (latent classes) interpretation sometimes tricky
  - Multiple testing issues + MNAR data

Objectives & Motivations

Objective

- Assessing the impact of known covariates on DIF and RS at item-level on PRO changes using longitudinal Rasch models

Statistical motivations

- Rasch models ⇒ specific objectivity property *(robust to MNAR data, simulation studies)*, interval measurements, item-level
- DIF and RS ⇒ jointly in the same modelling process *(latent regression)* ⇒ DIF stability over time + differential RS in subgroups
- Multiple testing ⇒ accounted for

Objectives & Motivations

Clinical motivation

- The term “cancer” includes various diseases and may generate ≠ social beliefs about prognosis and perceived dangerousness
- Breast cancer & melanoma patients might experience QoL changes and adaptation to their illness in a ≠ way during and after treatment ⇒ ≠ healthcare needs

Clinical objective

- Assess the impact of breast cancer and melanoma on DIF, RS and QoL changes (Emotional Functioning, EF)

Extension of the "RespOnse Shift ALg orithm for Item-level" (ROSALI) to incorporate covariates for assessing:

- **DIF** between groups (breast cancer and melanoma)
- Covariate effect on **DIF**, **RS** and **QoL changes** between 2 measurement occasions (*RS differentially estimated between groups*)

Cross-sectional and longitudinal Partial Credit Models (**PCM**) to detect non-uniform and uniform recalibration **RS**

- Iterative steps: DIF detection → RS detection ± DIF including *Bonferroni correction*

Guilleux et al. *Qual Life Res* 2015;24:553-64.
DIF detection – PCM at T1

Step A
DIF measurement model

\[
P(X_{ij} = h | c_i, \theta_i, \delta_{j1}, ..., \delta_{jm}, \gamma_{j1}, ..., \gamma_{jm}) = \frac{\exp(h\theta_i - [\sum_{p=1}^{h} (\delta_{jp} + \gamma_{jp}c_i)])}{\sum_{l=0}^{m_j} \exp(l\theta_i - [\sum_{p=1}^{l} (\delta_{jp} + \gamma_{jp}c_i)])}
\]

- \(X_{ij}\): response of patient i to item j
- \(\Theta \sim N(0, \sigma^2); \theta_i\): latent trait level of patient i
- \((\delta_{jp})\): item difficulties of item j
- \(m_j\): number of positive response categories for item j
- \(c_i\): group covariate
- \(\gamma_{jp}\): DIF parameter (*uniform if \(\gamma_{jp} = \gamma_j \, \forall p; \, non-uniform\, otherwise*)
DIF detection – PCM at T1

Step B
No DIF model

\[ P(X_{ij} = h | \beta, c_i, \theta_i, \delta_{j_1}, \ldots, \delta_{jm_j}) = \frac{\exp(h(\beta c_i + \theta_i) - \sum_{p=1}^{h} \delta_{jp})}{\sum_{l=0}^{m_j} \exp(l(\beta c_i + \theta_i) - \sum_{p=1}^{l} \delta_{jp})} \]

- \( X_{ij} \): response of patient i to item j
- \( \Theta \sim N(0, \sigma^2); \theta_i \): latent trait level of patient i
- \( (\delta_{jp}) \): item difficulties of item j
- \( m_j \): number of positive response categories for item j
- \( c_i \): group covariate
- \( \beta \): regression parameter
DIF detection → RS detection

**Step A**
DIF measurement model

**Likelihood ratio test (LRT)**

- **Significant LRT**
  - Steps C & D
    - DIF Items
    - DIF stability over time
    - *Longitudinal PCM*

- **LRT not significant**
  - RS detection
    - *Longitudinal PCM*

**Step B**
No DIF model
RS detection – Longitudinal PCM T1/T2

Step 1
RS measurement model ± DIF

\[ P \left( X_{ij}^{(t)} \right) = h | \beta, c_i, \theta_i^{(t)}, \delta_{j1}, ..., \delta_{jm}, \eta_1^{(t)}, ..., \eta_{jm}, \gamma_1^{(t)}, ..., \gamma_{jm} \right) \]

\[ = \frac{\exp \left( h(\beta c_i + \theta_i^{(t)}) - \left[ \sum_{p=1}^{h} (\delta_{jp} + \eta_{jp}^{(t)} + \gamma_{jp}^{(t)} c_i) \right] \right)}{\sum_{l=0}^{m_j} \exp \left( l(\beta c_i + \theta_i^{(t)}) - \left[ \sum_{p=1}^{l} (\delta_{jp} + \eta_{jp}^{(t)} + \gamma_{jp}^{(t)} c_i) \right] \right)} \]

Constraints
\[ \eta_{jp}^{(1)} = 0 \quad \forall j, p \]

Based on steps C & D
• No DIF: \( \gamma_{jp}^{(1)} = 0 \quad \forall j, p \)
• Uniform DIF: \( \gamma_{jp}^{(t)} = \gamma_j^{(t)} \); Non-uniform DIF: \( \gamma_{jp}^{(1)} = \gamma_{jp}^{(2)} \)

Recalibration RS parameters

\[ \begin{bmatrix} \Theta^{(1)} \\ \Theta^{(2)} \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix} ; \Sigma \right) \]
RS detection – Longitudinal PCM T1/T2

Step 2
No RS model ± DIF

\[
\begin{bmatrix}
\Theta^{(1)} \\
\Theta^{(2)}
\end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ \mu^{(2)} \end{bmatrix}; \Sigma \right)
\]

\[
P \left( X_{ij}^{(t)} = h | \beta, \beta_{\text{time*group}}^{(t)}, c_i, \theta_i^{(t)}, \delta_{j1}, ..., \delta_{jm_j}, \eta_{j1}^{(t)}, ..., \eta_{jm_j}^{(t)}, \gamma_{j1}^{(t)}, ..., \gamma_{jm_j}^{(t)} \right)
\]

\[
= \frac{\exp \left( h (\beta c_i + \beta_{\text{time*group}}^{(t)} c_i + \theta_i^{(t)}) - \left[ \sum_{p=1}^{h} (\delta_{jp} + \gamma_{jp}^{(t)} c_i) \right] \right)}{\sum_{l=0}^{m_j} \exp \left( l (\beta c_i + \beta_{\text{time*group}}^{(t)} c_i + \theta_i^{(t)}) - \left[ \sum_{p=1}^{l} (\delta_{jp} + \gamma_{jp}^{(t)} c_i) \right] \right)}
\]

Constraints
\[
\eta_{jp}^{(t)} = 0 \ \forall j, p, t
\]

Based on steps C & D
- No DIF: \( \gamma_{jp}^{(1)} = 0 \ \forall j, p \)
- Uniform DIF: \( \gamma_{jp}^{(t)} = \gamma_j^{(t)} \); Non-uniform DIF: \( \gamma_{jp}^{(1)} = \gamma_{jp}^{(2)} \)
DIF detection → RS detection

**Step 1**
RS measurement model ± DIF

**Step 2**
No RS model ± DIF

**Likelihood ratio test (LRT)**

- **Significant LRT**
  - **Step 3**
    - Items with (differential) RS

- **LRT not significant**
  - **Step 4**
    - Latent trait change + impact of covariates
Material & methods – ELCCA study

Study design
- Prospective longitudinal study, 2 year-follow-up: *within 1 month after diagnosis (T1), 12 (T2) and 24 (T3) months after → focus on first year*

Inclusion criteria
- Adults, early stage non-metastatic (stages I and II) melanoma and breast cancer (BC), informed consent

PRO Measures
- Cancer-specific Quality of Life Questionnaire (QLQ-C30 version 3.0), emotional functioning (EF) scale with 4 items
  - *During the past week:* Did you feel tense? Did you worry? Did you feel irritable? Did you feel depressed?
  - *4-point Likert scale:* Not at all, A little, Quite a bit, and Very much
Results – One (T1) and 12 (T2) months after diagnosis

- Breast cancer & melanoma ⇒ Uniform recalibration - Item “Did you worry?”

η² = -0.95

⇒ For a same level of emotional functioning, patients reported lower worry levels at T2 as compared to T1
Results – Breast cancer only ⇒ Non-uniform recalibration - Item “Did you feel irritable?”

For a same level of emotional functioning, breast cancer patients reported higher irritability levels at T2 as compared to T1.
Results – Latent trait change

EF levels at T1
• Breast cancer < Melanoma

Change in EF levels
• Breast cancer → Increasing
• Melanoma → Stable

Significant group + interaction group*time effects
Discussion

- **DIF & Recalibration RS**
  - DIF at T1: for a same level of EF, higher irritability levels for melanoma as compared to BC patients → not stable over time (no DIF at T2) → RS
  - Similar RS *(worry)* and differential RS *(irritable)* evidenced for breast cancer (BC) and melanoma patients

- **QoL change**
  - Increasing for BC → reaching the QoL level of melanoma patients which remained stable during the first year

- **Interpretation of RS “effect sizes”**
  - What is a meaningful RS effect? What is the MCID of RS?
Recalibration RS → Reprioritization & Reconceptualization?

Statistical perspective
- Reprioritization → Generalized PCM with discrimination parameters → IRT not RMT anymore, *no specific objectivity*
- Reconceptualization → Multidimensional RMT or IRT models → convergence problems, large sample sizes

Conceptual perspective
- Reprioritization → Interpretation at dimension-level (e.g. social >> physical); does it makes sense at item-level?
Conclusion

**Perspectives**

- Enhance the development of methods for identification of RS at more **individual levels**
- **ROSALI2**: differential impact of covariates on RS and latent trait change at item-level
- Combine methods for the analysis of RS at **dimension** and **item-level**

- **Simulation studies** are needed to assess the performance of the methods