International Journal of Applied Mathematics & Statistics,
Int. J. Appl. Math. Stat.; Vol. 24; Issue No. SI-11A; Year 2011,
ISSN 0973-1377 (Print), ISSN 0973-7545 (Online)

Analysis of longitudinal Patient-Reported Outcomes with informative and non-informative dropout: Comparison of CTT and Rasch-based methods.

Myriam Blanchin[1], **Jean-Benoit Hardouin**[1], **Tanguy Le Neel**[1], **Gildas Kubis**[1] **and Véronique Sébille**[1]

[1]EA 4275 "Biostatistics, Clinical Research
and Subjective Measures in Health Sciences"
Faculty of Pharmaceutical Sciences, University of Nantes, France
myriam.blanchin@etu.univ-nantes.fr

ABSTRACT

Patient-reported outcomes (PRO) are more and more used in health sciences to evaluate concepts such as health-related quality of life. These outcomes cannot be directly observed and are often referred to a latent variable. Two psychometric theories exist for the analysis of PRO: the classical test theory, the most common used in practice and the item response theory with its most used model, the Rasch model. In many studies, PRO are collected longitudinally in order to study the evolution of the outcome through time. Missing data are frequently encountered in longitudinal studies and can be potentially informative. This study aimed at comparing Classical Test Theory (CTT) and Rasch-based approaches to analyze longitudinal PRO collected from a scale validated with a Rasch model and studying the impact of dropout, informative or not, on both approaches. Data with informative dropout have shown estimation bias and have to be analyzed with more appropriate methods. For complete data and data with non-informative dropout, a method of analysis based on the Rasch model may be preferred for the analysis of longitudinal PRO collected from a scale validated with a Rasch model due to the generally observed slight gain of power and the psychometric properties of the model.

Keywords: Classical Test Theory, Rasch model, longitudinal data, dropout, informative missing data.

2010 Mathematics Subject Classification: 92B15, 62P10 .

1 Introduction

Patient-reported Outcomes, PRO, have gained major concern in the past years. This type of measures, reported by patients based on their perceptions, includes health-related quality of life (HRQoL), functional well-being, patient satisfaction with treatment,... PRO are broadly used in clinical trials as secondary outcome, especially in chronic diseases like cancer and are sometimes used as primary outcome in some contexts. The evaluated outcome (e.g. HRQoL) is often referred to a latent variable and is measured through the responses of patients to items. The special nature of PRO, which are not directly observable, beg the question of the

analysis of the data. Two psychometric theories exist for the analysis of PRO. The most common used approach is the Classical Test Theory (CTT). In this theory, the observed score, usually computed by summing item responses, is used to estimate the 'true' value of the evaluated outcome. In the second theory, Item Response Theory (IRT), the probability to answer to an item is a function of a latent variable, which represents the evaluated outcome, and item parameters. Among the wide family of IRT models, the Rasch model (Rasch, 1980; Fischer and Molenaar, 1995) is the most commonly used for dichotomous items due to its properties: parameters invariance, specific objectivity and exhaustivity of the score on the latent trait. The Rasch model is now widely used in development and validation of scales in health sciences (Lai et al., 2007; Cella et al., 1996).

Many studies including PRO are studies on chronic illness or follow-up studies after treatment or surgery. Thus, patients are evaluated at different time points to allow the analysis of the evolution of the evaluated PRO. In these longitudinal studies, the measures of each patient are therefore unlikely to be independent such as in cross-sectional studies and the correlation between measurements has to be taken into account in the analysis. One way to deal with correlated data is the use of linear mixed models (Verbeke and Molenberghs, 2000b; Fitzmaurice et al., 2009).

Missing data are frequently encountered in longitudinal studies. For instance, a patient can drop out from the study at a certain time point and so answers to questionnaire are missing for this patient after this time. Intermittent missing data can also occur when some items are not answered in a questionnaire. When the reason for missingness may be related to the evaluated outcome level of the patient, the missing data are said to be informative. Otherwise, they are called ignorable. Missing data, depending on their amount and informativity, may have an impact on the analysis and interpretation of the data. There may be a reduction of the statistical power of the analysis and a bias may be introduced leading to incorrect conclusions.

In practice, when longitudinal data coming from a scale validated with a Rasch model have to be analyzed, many methods can be considered. The researchers tend to use more often the CTT approach, probably more from habit than evidence of suitability. The purpose of this paper is to compare methods either based on CTT or Rasch model to analyze longitudinal latent variables through a simulation study. The impact of missing data in this context has also been studied.

2 Methods

2.1 Longitudinal data analysis

When measures are repeated on the same patients through time, linear mixed models are widely used for the analysis of the data. These models allow to deal with the correlation between measures of longitudinal data by specifying fixed effects (population characteristics), random effects (subject-specific effects) and structure of the variance-covariance matrix

(Verbeke and Molenberghs, 2000b). A general linear mixed model can be written as follows:

$$\begin{aligned}
&\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i,\\
&\boldsymbol{b}_i \sim N_q(0, \boldsymbol{D}),\\
&\boldsymbol{e}_i \sim N_{n_i}(0, \boldsymbol{\Sigma}_i),\\
&\boldsymbol{b}_1, ..., \boldsymbol{b}_q, \boldsymbol{e}_1, ..., \boldsymbol{e}_N \; independent\;,\\
&\boldsymbol{Y}_i \sim N_{n_i}(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i' + \boldsymbol{\Sigma}_i),
\end{aligned} \tag{2.1}$$

where $\boldsymbol{Y}_i$ is the response vector for patient i, $i = 1, ..., N$, N is the number of patients, n_i is the number of observations on patient i, p the number of fixed parameters, q the number of random parameters, $\boldsymbol{\beta}$ is a $(p \times 1)$ vector of fixed effects parameters, $\boldsymbol{X}_i$ is the $(n_i \times p)$ design matrix for fixed effects, $\boldsymbol{b}_i$ is a $(q \times 1)$ vector of random effects parameters, $\boldsymbol{Z}_i$ is the $(n_i \times q)$ design matrix for random effects, $\boldsymbol{e}_i$ is a $(n_i \times 1)$ vector of residual components, $\boldsymbol{D}$ is the $(q \times q)$ among-unit covariance matrix and $\boldsymbol{\Sigma}_i$ is the $(n_i \times n_i)$ within-unit covariance matrix.
Two methods of estimation based on the likelihood can be used to estimate the mean parameters β and variance components $\boldsymbol{\omega}$ (that contains all variances and covariance parameters found in $V_i = \boldsymbol{Z}_i\boldsymbol{D}\boldsymbol{Z}_i' + \boldsymbol{\Sigma}_i$) of the model: Maximum Likelihood estimation (ML) and Restricted maximum Likelihood estimation (REML). The estimations of variance components obtained with ML are known to be biased for finite samples. The REML estimation is used to correct the bias on ML estimates of variance components (Laird and Ware, 1982). The REML estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\omega}$ are found by maximizing the so-called REML likelihood function.

$$\begin{aligned}
L_{REML}(\boldsymbol{\beta}, \boldsymbol{\omega}) = &\left|\sum_{i=1}^{N} \boldsymbol{X}_i' V_i^{-1}(\boldsymbol{\omega}) \boldsymbol{X}_i\right|^{-1/2}\\
&\times \prod_{i=1}^{N} (2\pi)^{-n_i/2} |V_i(\boldsymbol{\omega})|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})' V_i^{-1}(\boldsymbol{\omega})(\boldsymbol{Y}_i - \boldsymbol{X}_i\boldsymbol{\beta})\right\}
\end{aligned} \tag{2.2}$$

2.2 Patient Reported Outcomes analysis

Assume that a questionnaire has J dichotomous items and that measures are repeated T times on the N patients of the study. The response of patient i ($i = 1, ..., N$) to an item j ($j = 1, ..., J$) at time t ($t = 1, ..., T$) is denoted by $Y_{ij}^{(t)}$.

2.2.1 Classical Test Theory approach

The Classical Test Theory (CTT) approach is based on a score usually computed by summing item responses. This approach was the first one developed in psychometrics and has widely spread in PRO analysis because of its simplicity to use and to interpret. The basic model assumes that the observed score is a linear function of the true score and an error term.
A method called **Score and Mixed Models (SM)** and based on the CTT approach was used for the analysis. In the first step, the score of each patient at each time was computed by summing the J item responses of the patient i at time t. A simple linear mixed model was then applied on the scores to investigate whether a time effect was plausible. The SM method was

carried out as follows:

$$\begin{aligned} \boldsymbol{S}_i^{(t)} &= \sum_{j=1}^{J} Y_{ij}^{(t)}, \\ \boldsymbol{S}_i &= (S_i^{(1)}, ..., S_i^{(T)})' = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{e}_i, \\ \boldsymbol{e}_i &\sim N(0, \boldsymbol{\Sigma}_{S,i}), \\ \boldsymbol{S}_i &\sim N_T(\boldsymbol{\mu}_S, \boldsymbol{\Sigma}_{S,i}), \end{aligned} \tag{2.3}$$

with $\boldsymbol{\mu}_S = (\mu_S^{(1)} \cdots \mu_S^{(T)})' = \boldsymbol{X}_i\boldsymbol{\beta}$. The mean parameters $\boldsymbol{\beta}$ and variance components $\boldsymbol{\omega}$ of the model were estimated using REML estimation in SAS Proc MIXED (Littell et al., 1996).

Three covariance structures $\boldsymbol{\Sigma}_{S,i}$ are often used with longitudinal data : unstructured, first-order autoregressive and heterogeneous compound symmetry denoted by UN, AR(1) and CSH respectively. The unstructured matrix is the most general possible structure. It is used when no hypothesis can be made on the structure of the covariance matrix but leads to estimate an important number of parameters. The AR(1) structure assumed that variances are constant over time and that correlation decreases when measures get further apart from each other in time. On the contrary, the choice of a CSH structure assumes that the variances are not equal but that the correlation is constant over time. For $T = 3$, the three structures of covariance can be written as follows with ρ denoting the correlation coefficient, σ_{ij} the covariance between latent variables at time i and time j ($i \neq j$) and σ_i^2 the variance of the latent variable at time i.

$$\boldsymbol{\Sigma}_{S,i} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{pmatrix} \text{ for UN} \qquad \boldsymbol{\Sigma}_{S,i} = \begin{pmatrix} \sigma^2 & \sigma^2\rho & \sigma^2\rho^2 \\ \sigma^2\rho & \sigma^2 & \sigma^2\rho \\ \sigma^2\rho^2 & \sigma^2\rho & \sigma^2 \end{pmatrix} \text{ for AR(1)}$$

$$\boldsymbol{\Sigma}_{S,i} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho & \sigma_1\sigma_3\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 & \sigma_2\sigma_3\rho \\ \sigma_1\sigma_3\rho & \sigma_2\sigma_3\rho & \sigma_3^2 \end{pmatrix} \text{ for CSH}$$

To compare non-nested models for the covariance, one of the information criteria that have been proposed is the Akaike Information Criteria also noted AIC (Akaike, 1974). This tool for model selection aims at comparing models based on their maximized log-likelihood value, ensuring that the retained models show a good fit of data. In order to select the most parsimonious model, the AIC penalizes models for the use of too many parameters. When the likelihood is estimated using REML, the AIC can be expressed as:

$$AIC = -2\hat{l} + 2c \tag{2.4}$$

where $\hat{l}$ is REML maximum log-likelihood and c is the number of covariance parameters. The most parsimonious correct model will be the model with the smallest AIC amongst the models with the same mean structure.

2.2.2 Item Response Theory and the Rasch model

Item Response Theory (IRT) emerged recently in instrument development and data analysis in health outcomes measurements due to its potential advantages over CTT such as parameters invariance (Hambleton, 2000). IRT is a family of models that express the probability of a patient's particular response to an item as a function of characteristics of the patient (latent variable θ) and characteristics of the item. The latent variable is the evaluated outcome and is considered as latent because it is not observable and must be inferred from item responses. Most of IRT models assume the unidimensionality of the construct, that is a unique latent variable explains the item responses. Among the unidimensional IRT model, the most commonly used model is the Rasch model (Rasch, 1980) due to its properties: the exhaustivity of the score on the latent trait and the specific objectivity. The exhaustivity refers to the property that the total score of a person is a sufficient statistic for the unknown latent trait. This means that no additional information is needed to estimate the person parameter θ. Each total score is associated with only one trait level in the Rasch model whatever the pattern of responses. The property of specific objectivity ensures that the difference between two latent traits does not depend on the set of items used to evaluate these traits. It allows to construct shorter versions of questionnaires or several versions of a same questionnaire to adapt the version to the patient's latest variable level.

The Rasch model expresses the probability of a response y ($y = 0$ for a negative response (the most pejorative response) and $y = 1$ for a positive response) of an individual i ($i = 1, ..., N$) to a dichotomous item j ($j = 1, ..., J$) as a logistic function of the individual value of the latent trait θ_i and one item parameter, its difficulty δ_j.

$$P(Y_{ij} = y|\theta_i, \delta_j) = \frac{exp(y(\theta_i - \delta_j))}{1 + exp(\theta_i - \delta_j)} \tag{2.5}$$

The Rasch model has been extended to situations where responses of individuals to items are observed at several points in time (Meiser, 2007). A longitudinal form of the Rasch model was used for the analysis in the method called **Longitudinal Rasch Model (LRM)**.

$$\begin{aligned} &P(Y_{ij}^{(t)} = y^{(t)}|\theta_i^{(t)}, \delta_j) = \frac{exp(y^{(t)}(\theta_i^{(t)} - \delta_j))}{1 + exp(\theta_i^{(t)} - \delta_j)} \\ &\boldsymbol{\theta_i} = (\theta_i^{(1)}, ..., \theta_i^{(T)})' \; iid \; N_T(\boldsymbol{\mu}, \boldsymbol{\Sigma}_i) \end{aligned} \tag{2.6}$$

This model assumes that the item parameters δ_j remain constant over time. The change in the latent ability $\boldsymbol{\theta}$ may be person-specific, that is the speed or direction of the evolution may be different from a person to another. As $\boldsymbol{\theta}$ is assumed to have a multinormal distribution, this model is of the family of the mixed-effects logistic models. The mean parameters $\boldsymbol{\mu}$ and covariance parameters $\boldsymbol{\Sigma}_i$ of the model are estimated using Marginal Maximum Likelihood (MML) estimation method. The marginal likelihood is expressed as

$$L(\delta_1, .., \delta_J, \boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{y}) = \prod_{i=1}^{N} \int_{\mathbb{R}^T} \prod_{t=1}^{T} \prod_{j=1}^{J} \frac{\exp\left(y_{ij}^{(t)}(\theta_i^{(t)} - \delta_j)\right)}{1 + \exp(\theta_i^{(t)} - \delta_j)} G(\boldsymbol{\theta}_i/\boldsymbol{\mu}, \boldsymbol{\Sigma}_i) d\boldsymbol{\theta}_i \tag{2.7}$$

with $G(./\boldsymbol{\mu}, \boldsymbol{\Sigma}_i)$ the multivariate normal distribution function with mean vector $\boldsymbol{\mu} = (\mu^{(1)} \cdots \mu^{(T)})'$ and an unstructured covariance matrix $\boldsymbol{\Sigma}_i$. $\mu^{(1)}$ is constrained to 0 in order to ensure the identifiability of the model. The parameters of the model were estimated using gllamm in Stata (Zheng and Rabe-Hesketh, 2007).

2.3 Missing data

Missing assessments of PRO are frequently encountered in longitudinal studies. Patients face a disease and/or treatment that have an impact on the evaluated outcome. The reasons for missingness can be totally unrelated to the subject's level of the evaluated outcome (a missed appointment, a move in another town) or may be intimely related to patient's level of evaluated PRO (side effects such as nausea or vomiting).

Missing data are often described as either 'dropout' or 'intermittent'. Dropout occurs when all observations on a subject are obtained until a certain point in time after which all measurements are missing (Diggle and Kenward, 1994). Intermittent missingness occurs when a subject misses an assessment but is later observed.

Little and Rubin (Little and Rubin, 2002) have defined three types of missing data (completely random, random, or not at random) depending on the mechanism that lead to missing data. An observation is said to be missing completely at random (MCAR) if the missingness probability is independent of all previous, current and future assessments. The missing process therefore does not depend on the values of the data, missing or observed. Data are missing at random (MAR) if the missingness probability does not depend on the missing values but only on the observed values. When data are missing not at random (MNAR) the missing data mechanism may depend on the unobserved values. Within the framework of maximum likelihood or Bayesian inference, this mechanism is often termed as 'non-ignorable' or 'informative'. When the data are 'ignorable' (MCAR or MAR), a valid analysis can be obtained through a likelihood-based analysis that ignores the dropout mechanism, provided the parameters describing the measurement process are functionally independent of the parameters describing the dropout process (Verbeke and Molenberghs, 2000b).

2.4 Simulation

The interest of longitudinal studies is in evaluating the evolution of a criteria through time. We define the time effect between time t and time $t+1$ as $d_{t,t+1} = \mu^{(t+1)} - \mu^{(t)}$. The data were simulated with a longitudinal Rasch mixed model assuming that patients were evaluated at three different times ($t = 1, 2, 3$).

$$P(Y_{ij}^{(t)} = y^{(t)} | \theta_i^{(t)}, \delta_j) = \frac{exp(y^{(t)}(\theta_i^{(t)} - \delta_j))}{1 + exp(\theta_i^{(t)} - \delta_j)} \tag{2.8}$$

where the latent trait vector $(\boldsymbol{\theta}_i^{(1)}, \boldsymbol{\theta}_i^{(2)}, \boldsymbol{\theta}_i^{(3)})'$, $(i = 1, ..., N)$ had a multivariate normal distribution $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\mu} = (-d_\theta, 0, d_\theta)'$ and $\boldsymbol{\Sigma} = \sigma^2 \begin{pmatrix} 1 & \rho_\theta & \rho_\theta^2 \\ \rho_\theta & 1 & \rho_\theta \\ \rho_\theta^2 & \rho_\theta & 1 \end{pmatrix}$.

d_θ is the value of the time effect between two consecutive times, $d_{1,2} = d_{2,3} = d_\theta$. For data simulated without time effect, $d_\theta = 0$. For data simulated with a time effect, $d_\theta = 0.2$. The first-order autoregressive structure adopted for the covariance matrix $\mathbf{\Sigma}$ means that variances are constant with time and that correlation between measures of a same patient decreases with time.

We can expect that some parameters have an impact on the performance of the two methods: datasets with different values for the sample size (N), number of items (J) and correlation of the latent variable (ρ_θ) were simulated. The data were assumed to come from a 4-item scale or a 7-item scale with dichotomous items. The values of difficulty parameters were $\delta_1 = -1$, $\delta_2 = -0.5$, $\delta_3 = 0.5$, $\delta_4 = 1$ for a 4-item scale and $\delta_1 = -1.5$, $\delta_2 = -1$, $\delta_3 = -0.5$, $\delta_4 = 0$, $\delta_5 = 0.5$, $\delta_6 = 1$, $\delta_7 = 1.5$ for a 7-item scale. The sample size could be of 100 or 200 individuals. Three different values for the correlation coefficient of the latent trait between two consecutive times ρ_θ were used: $\rho_\theta = 0.4$ (small correlation), $\rho_\theta = 0.7$, and $\rho_\theta = 0.9$ (high correlation).

To simulate the dropout of patients from the study, a latent variable denoted χ was defined as the dropout propensity. The probability that a patient drops out from the study at time t depends on its dropout propensity. The dropout process was simulated using the following model derived from a 4-parameter logistic model (Sijtsma and Hemker, 2000):

$$P(DO_i^{(t)} = 1|\chi_i^{(t)}, \pi_{min}^{(t)}, \pi_{max}^{(t)}) = \pi_{min}^{(t)} + (\pi_{max}^{(t)} - \pi_{min}^{(t)})\frac{exp(\chi_i^{(t)})}{1 + exp(\chi_i^{(t)})} \tag{2.9}$$

with $DO_i^{(t)} = 1$ represents the situation where a patient i drops out from the study at time t, $\pi_{min}^{(t)}$ the minimum individual probability of dropout at time t and $\pi_{max}^{(t)}$ the maximum individual probability of dropout at time t. $\pi_{min}^{(t)}$ and $\pi_{max}^{(t)}$ were defined such as the expected proportion of dropout at time t was $\pi^{(t)} = \frac{\pi_{min}^{(t)} + \pi_{max}^{(t)}}{2}$. We assume that data are complete at the first time of evaluation ($\pi^{(1)} = 0$). The dropout of the patients is then linear and π of the remaining patients drop out from the study at each time ($t = 2, 3$).

The dropout propensity χ_i has a multinormal distribution with mean vector (0 0 0)' and a variance covariance matrix equals to $\begin{pmatrix} 1 & \rho_{\theta\chi}^2\rho_\theta & \rho_{\theta\chi}^2\rho_\theta^2 \\ \rho_{\theta\chi}^2\rho_\theta & 1 & \rho_{\theta\chi}^2\rho_\theta \\ \rho_{\theta\chi}^2\rho_\theta^2 & \rho_{\theta\chi}^2\rho_\theta & 1 \end{pmatrix}$.

The correlation between the value of the latent variable at time t, $\theta^{(t)}$, and the dropout propensity at time t, $\chi^{(t)}$, denoted $corr(\theta^{(t)}, \chi^{(t)}) = \rho_{\theta\chi}$ and assumed constant with time, was used to determine the type of missingness of the dropout process. When the value of the latent variable for the outcome does not depend on the dropout propensity ($\rho_{\theta\chi} = 0$), the simulated dropout is of MCAR type following the definition of Little and Rubin. As the two methods are based on likelihood and ignores the dropout mechanism, we can expect that analyses on data with MCAR dropout will be valid. However, it is reasonable to assume that missing data mechanism may often be MNAR in studies including PRO, HRQoL for instance. The simulated dropout is MNAR when $\rho_{\theta\chi} \neq 0$. Furthermore, the patients with worse HRQoL, due to disease progression or increase of side effects, may be most likely to dropout from the study than other patients (Troxel, Fairclough, Curran and Hahn, 1998). So, we assume that $\rho_{\theta\chi} < 0$ to simulate

MNAR dropout.
To study the behaviour of SM and LRM in case of missing data, the proportion of dropout in the simulated datasets could be $\pi^{(t)} = 0\%$ (complete data), 5%, 10% or 20% ($t = 2, 3$). The correlation between the value of the latent variable θ and the dropout propensity were $\rho_{\theta\chi} = 0$ (MCAR dropout), $\rho_{\theta\chi} = -0.4; -0.7; -0.9$ (MNAR dropout with increasing informativity). The different values of the parameters led to consider 312 different cases. Five hundred simulated datasets were generated and analyzed for each case.

To compare the two methods, a test for time effect was defined using an approximate Wald test:
$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu \Leftrightarrow L\boldsymbol{\mu} = 0$
$H_1 : \exists i | \mu_i \neq \mu \Leftrightarrow L\boldsymbol{\mu} \neq 0$
$\boldsymbol{L} = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$
Under H_0, $T_L = (\boldsymbol{L}\hat{\boldsymbol{\mu}})'(\boldsymbol{L}\hat{\boldsymbol{V}}\boldsymbol{L}')^{-1}\boldsymbol{L}\hat{\boldsymbol{\mu}}$ has approximately a χ^2_r distribution (Verbeke and Molenberghs, 2000a) where $\hat{\boldsymbol{\mu}}$ is the estimate of $\boldsymbol{\mu}$, r is the rank of $\boldsymbol{L}$ and $\hat{\boldsymbol{V}}$ is the estimated covariance matrix.

In order to compare the methods to analyze longitudinal PRO data, three criteria were studied: the type I error, the power and the bias of the time effect estimation. The type I error of the tests were classically computed as the proportion of rejection of H_0 under the null hypothesis. Rejection of H_0 was based on a test of simultaneous equality of mean estimations, i.e. the absence of time effect. This criteria allowed the study of the aptitude of the method to avoid falsely detecting a time effect. The power calculation used the same tests but calculated the proportion of rejection of H_0 under the alternative hypothesis. On the opposite, the power allowed to study the aptitude of the method to correctly detect the presence of time effect. A one-sided McNemar's test for paired data (McNemar, 1947) was used to compare the power observed for each method.
$H_0 : power_{LRM} = power_{SM}$
$H_1 : power_{LRM} > power_{SM}$

The comparison of the estimated time effect to the simulated 'true' time effect gave the bias of time effect and informed about the quality of the parameters estimation. The comparison held for the LRM method as the known true value of time effect had been fixed for the latent variable. For SM method, based on score, the 'true' time effect was not known and was estimated by d_S, the difference of the computed expected score at each time. For $t = 2, 3$,

$$d_S = E\left(S_i^{(t)}\right) - E\left(S_i^{(t-1)}\right) \tag{2.10}$$

with

$$
\begin{aligned}
E\left(S_i^{(t)}\right) &= E\left(\sum_j Y_{ij}^{(t)}\right) \\
&= \sum_j E\left(Y_{ij}^{(t)}\right) \\
&= \sum_j P\left(Y_{ij}^{(t)}=1\right) \\
&= \sum_j \int_{\mathbb{R}} P\left(Y_{ij}^{(t)}=1\right) G(\theta_i^{(t)}/\mu^{(t)},\sigma^2)d\theta_i \\
&= \sum_j \int_{\mathbb{R}} \frac{\exp\left(\theta_i^{(t)}-\delta_j\right)}{1+\exp(\theta_i^{(t)}-\delta_j)} G(\theta_i^{(t)}/\mu^{(t)},\sigma^2)d\theta_i
\end{aligned}
\tag{2.11}
$$

where $G(\theta_i^{(t)}/\mu^{(t)},\sigma^2)$ the normal distribution with mean $\mu^{(t)}$ and variance σ^2. These integrals can be estimated using Gauss-Hermite quadratures. Considering the simulated item parameters, we obtained the following estimations. For a simulated time effect between two consecutive times on the latent variable $d_\theta = 0$, the time effect on the score d_S is also 0. For a simulated time effect on the latent variable $d_\theta = 0.2$, the time effect on the score d_S is 0.15 when $J = 4$ and 0.25 when $J = 7$.

3 Results

To determine which structure of covariance is the most adequate to use for the analysis of data with SM method, we compared the Akaike Information Criteria (AIC) of the three structures of covariance matrix, UN, AR(1) and CSH for each of the 312 cases. When $\rho_\theta = 0.4$ or $\rho_\theta = 0.7$, the AIC was more often minimized by the choice of an AR(1) structure for the covariance matrix. When $\rho_\theta = 0.9$, the CSH structure more often minimized the AIC of the models than the other structures. Results presented further for SM method come from analyses with an AR(1) structure. LRM method used an unstructured covariance matrix.

3.1 Type I error rate and power

Table 1 shows the type I error of the test for time effect for each of the methods depending on the value of all simulation parameters: sample size, number of items, latent variable correlation, proportion of dropout and type of dropout. Both methods, LRM and SM, give comparable values of type I error whatever the value of the simulation parameters. The value of the sample size, the number of items and the latent variable correlation don't seem to have an impact on the values of the type I error since the type I errors only show small variations as the values of these three parameters change. When the data are complete ($\pi = 0$), all type I error are close to the expected value of 5%. In this case, the values range from 4.4% and 3.6% for LRM and SM respectively when N=100, J=7 and $\rho_\theta = 0.4$ to 6.6% and 6.0% for LRM and SM respectively when N=200, J=4 and $\rho_\theta = 0.9$. Whatever the proportion of dropout, data subject to MCAR dropout ($\rho_{\theta\chi} = 0$) show type I errors close to complete data ones. When dropout of

Table 1: Type I error of the tests of time effect for Score Mixed model (SM) and Longitudinal Rasch Mixed model (LRM) methods for different values of sample size (N), number of items (J), latent variable correlation (ρ_θ), proportion of dropout (π) and type of dropout ($\rho_{\theta\chi}$). Results from analyses with an AR(1) structure for the covariance matrix of SM method and an unstructured covariance matrix for LRM method.

				no dropout		MCAR		MNAR					
								$\rho_{\theta\chi}=-0.4$		$\rho_{\theta\chi}=-0.7$		$\rho_{\theta\chi}=-0.9$	
N	J	ρ_θ	π	LRM	SM	LRM	SM	LRM	SM	LRM	SM	LRM	SM
100	4	0.4	0	0.054	0.054								
			0.05			0.054	0.050	0.066	0.066	0.060	0.054	0.038	0.038
			0.1			0.046	0.042	0.050	0.046	0.048	0.052	0.054	0.054
			0.2			0.050	0.048	0.072*	0.064	0.080*	0.078*	0.070	0.066
		0.7	0	0.044	0.040								
			0.05			0.048	0.042	0.040	0.034	0.042	0.044	0.036	0.034
			0.1			0.048	0.038	0.044	0.044	0.034	0.032	0.058	0.052
			0.2			0.068	0.060	0.052	0.060	0.052	0.046	0.076*	0.080*
		0.9	0	0.048	0.038								
			0.05			0.048	0.048	0.032	0.034	0.048	0.050	0.050	0.054
			0.1			0.040	0.046	0.024*	0.030*	0.032	0.038	0.034	0.036
			0.2			0.048	0.046	0.034	0.048	0.050	0.046	0.044	0.038
	7	0.4	0	0.044	0.036								
			0.05			0.054	0.048	0.066	0.058	0.048	0.050	0.040	0.038
			0.1			0.060	0.058	0.054	0.058	0.068	0.064	0.072*	0.066
			0.2			0.046	0.046	0.054	0.046	0.062	0.062	0.088*	0.080*
		0.7	0	0.058	0.046								
			0.05			0.056	0.046	0.038	0.038	0.054	0.052	0.056	0.042
			0.1			0.054	0.040	0.052	0.046	0.050	0.042	0.056	0.036
			0.2			0.052	0.054	0.064	0.044	0.090*	0.084*	0.090*	0.072*
		0.9	0	0.046	0.044								
			0.05			0.049	0.050	0.052	0.038	0.056	0.046	0.036	0.036
			0.1			0.038	0.034	0.046	0.046	0.048	0.040	0.064	0.060
			0.2			0.034	0.034	0.064	0.050	0.054	0.054	0.054	0.068
200	4	0.4	0	0.052	0.060								
			0.05			0.054	0.052	0.040	0.046	0.062	0.060	0.048	0.052
			0.1			0.052	0.044	0.066	0.058	0.080*	0.072*	0.046	0.048
			0.2			0.046	0.050	0.052	0.052	0.074*	0.076*	0.106*	0.104*
		0.7	0	0.056	0.048								
			0.05			0.042	0.038	0.040	0.042	0.080*	0.072*	0.050	0.044
			0.1			0.054	0.050	0.044	0.048	0.062	0.058	0.086*	0.078*
			0.2			0.038	0.040	0.044	0.044	0.050	0.056	0.088*	0.086*
		0.9	0	0.066	0.060								
			0.05			0.044	0.046	0.052	0.044	0.050	0.048	0.032	0.030*
			0.1			0.057	0.050	0.044	0.042	0.057	0.058	0.060	0.052
			0.2			0.046	0.046	0.063	0.052	0.069	0.070	0.074*	0.082*
	7	0.4	0	0.050	0.050								
			0.05			0.056	0.050	0.052	0.052	0.064	0.058	0.064	0.062
			0.1			0.042	0.038	0.064	0.054	0.066	0.052	0.058	0.052
			0.2			0.066	0.070	0.080*	0.072*	0.058	0.054	0.106*	0.100*
		0.7	0	0.048	0.036								
			0.05			0.054	0.054	0.066	0.052	0.046	0.042	0.068	0.066
			0.1			0.056	0.038	0.060	0.046	0.058	0.054	0.064	0.056
			0.2			0.052	0.032	0.084*	0.074*	0.086*	0.084*	0.080*	0.086*
		0.9	0	0.058	0.048								
			0.05			0.056	0.050	0.054	0.046	0.059	0.058	0.060	0.060
			0.1			0.044	0.028*	0.050	0.054	0.038	0.036	0.048	0.044
			0.2			0.053	0.046	0.059	0.040	0.088*	0.080*	0.103*	0.104*

* indicates that the 95% confidence interval of the type I error does not contain the expected value of 5%.

MNAR type occurs, type I errors increase with the proportion of dropout and the value of the correlation between the latent variable and the dropout propensity. As the values of these two parameters rise in absolute value, the number of 95% confidence intervals of the type I error that do not contain the expected 5% also rises. The type I error can reach 10% in the worst cases when $\pi = 20\%$ and $\rho_{\theta\chi} = -0.9$.

Table 2 shows the results of power of the test for time effect for both methods depending on the value of all simulation parameters: sample size, number of items, latent variable correlation, proportion of dropout and type of dropout. As observed for the type I errors, values of power are close to each other for LRM and SM method for fixed values of simulation parameters but the powers for LRM seem to be generally slightly higher than the power for SM when the correlation $\rho_\theta = 0.4$. LRM powers seem to be systematically slightly higher than the SM powers when the correlation $\rho_\theta = 0.7$ or 0.9. In these cases, the McNemar's tests are always significant at 5% concluding that LRM power is higher than SM power. Powers increases with the sample size, number of items and latent variable correlation. For example, powers range from 38.8% and 38.4% for LRM and SM respectively when $N = 100$, $J = 4$, $\rho_\theta = 0.4$ and $\pi = 0$ to 95.5% and 91.8% for LRM and SM respectively when $N = 200$, $J = 7$, $\rho_\theta = 0.9$ and $\pi = 0$. Powers of data presenting MCAR dropout are lower than the corresponding powers of complete data. This loss of power is highest when the proportion of dropout and the latent variable correlation are high, $\pi = 20\%$ and $\rho_\theta = 0.9$. The fall is up to -15.7% and -20.8% for LRM and SM respectively when $N = 100$, $J = 7$, $\rho_\theta = 0.9$ and $\pi = 20\%$. Powers of data with MNAR dropout are higher than powers of complete data. As the proportion of dropout and the informativity of the missing data (as ρ_{θ_χ} decreases to -0.9) increase, the powers for MNAR case get higher than for complete case.

3.2 Time effect estimation

The observed effect of the proportion and the informativity of dropout on type I error and power suggest a bias of the time effect estimation. Figure 1 shows the estimation of the time effect between the two first times of evaluation for different values of sample size, number of items and latent variable correlation when the simulated time effect on the latent trait was null. Figure 1 shows that time effect seems to be well estimated for complete data and data with MCAR dropout since all bias estimations for these cases are close to 0. For data with MNAR dropout, LRM and SM seem to overestimate the time effect. The overestimation is more marked when the latent variable correlation is low. The bias produced by the SM method depends on the number of items J. When $J = 4$, the overestimation for SM is lower than when $J = 7$. As a consequence, the overestimation for SM is lower than for LRM when $J = 4$ and higher than for LRM when $J = 7$.

Figure 2 shows the estimation of the time effect between the two first times of evaluation for different values of sample size, number of items and latent variable correlation when the simulated time effect on the latent trait was equal to 0.2. As described before, the time effect on score was estimated to 0.15 for $J = 4$ and to 0.25 for $J = 7$. On figure 2, the estimations for

Table 2: Power of the tests for Score Mixed model (SM) and Longitudinal Rasch Mixed model (LRM) methods for different values of sample size (N), number of items (J), latent variable correlation (ρ_θ), proportion of dropout (π) and type of dropout ($\rho_{\theta\chi}$). Results from analyses with an AR(1) structure for the covariance matrix of SM method and an unstructured covariance matrix for LRM method.

				no dropout		MCAR		MNAR					
								$\rho_{\theta\chi}=-0.4$		$\rho_{\theta\chi}=-0.7$		$\rho_{\theta\chi}=-0.9$	
N	J	ρ_θ	π	LRM	SM	LRM	SM	LRM	SM	LRM	SM	LRM	SM
100	4	0.4	0	0.388	0.384								
			0.05			0.399	0.400	0.356	0.354	0.383$^\sharp$	0.360	0.404	0.404
			0.1			0.335	0.320	0.411	0.408	0.439	0.426	0.440	0.426
			0.2			0.325	0.316	0.384	0.376	0.474	0.464	0.513	0.508
		0.7	0	0.423$^\sharp$	0.372								
			0.05			0.439$^\sharp$	0.382	0.426$^\sharp$	0.390	0.457$^\sharp$	0.400	0.498$^\sharp$	0.448
			0.1			0.441$^\sharp$	0.406	0.453$^\sharp$	0.400	0.480	0.464	0.499$^\sharp$	0.466
			0.2			0.399$^\sharp$	0.368	0.399$^\sharp$	0.378	0.486$^\sharp$	0.466	0.543$^\sharp$	0.524
		0.9	0	0.508$^\sharp$	0.424								
			0.05			0.471$^\sharp$	0.390	0.505$^\sharp$	0.432	0.511$^\sharp$	0.432	0.560$^\sharp$	0.474
			0.1			0.462$^\sharp$	0.398	0.427$^\sharp$	0.366	0.511$^\sharp$	0.436	0.508$^\sharp$	0.462
			0.2			0.379$^\sharp$	0.342	0.477$^\sharp$	0.434	0.509	0.496	0.598$^\sharp$	0.578
	7	0.4	0	0.470$^\sharp$	0.428								
			0.05			0.454$^\sharp$	0.424	0.488$^\sharp$	0.468	0.504	0.492	0.518$^\sharp$	0.496
			0.1			0.482$^\sharp$	0.454	0.522$^\sharp$	0.494	0.510$^\sharp$	0.496	0.556	0.546
			0.2			0.398	0.396	0.500	0.482	0.534$^\sharp$	0.514	0.600	0.588
		0.7	0	0.568$^\sharp$	0.522								
			0.05			0.558$^\sharp$	0.502	0.586$^\sharp$	0.548	0.598$^\sharp$	0.534	0.626$^\sharp$	0.576
			0.1			0.541$^\sharp$	0.470	0.549$^\sharp$	0.510	0.623$^\sharp$	0.574	0.607$^\sharp$	0.562
			0.2			0.446$^\sharp$	0.414	0.568$^\sharp$	0.518	0.655$^\sharp$	0.632	0.685$^\sharp$	0.658
		0.9	0	0.688$^\sharp$	0.564								
			0.05			0.670$^\sharp$	0.548	0.679$^\sharp$	0.604	0.723$^\sharp$	0.622	0.725$^\sharp$	0.620
			0.1			0.635$^\sharp$	0.526	0.686$^\sharp$	0.584	0.743$^\sharp$	0.654	0.719$^\sharp$	0.632
			0.2			0.531$^\sharp$	0.452	0.687$^\sharp$	0.590	0.745$^\sharp$	0.674	0.737$^\sharp$	0.684
200	4	0.4	0	0.654$^\sharp$	0.634								
			0.05			0.654	0.644	0.682	0.668	0.718$^\sharp$	0.690	0.678$^\sharp$	0.658
			0.1			0.631	0.628	0.658	0.646	0.738	0.726	0.718	0.714
			0.2			0.552	0.540	0.694	0.688	0.745	0.736	0.820	0.818
		0.7	0	0.721$^\sharp$	0.678								
			0.05			0.729$^\sharp$	0.694	0.747$^\sharp$	0.700	0.787$^\sharp$	0.742	0.753$^\sharp$	0.702
			0.1			0.701$^\sharp$	0.670	0.752$^\sharp$	0.724	0.806$^\sharp$	0.772	0.825$^\sharp$	0.786
			0.2			0.669$^\sharp$	0.626	0.745$^\sharp$	0.710	0.843$^\sharp$	0.830	0.820$^\sharp$	0.808
		0.9	0	0.826$^\sharp$	0.756								
			0.05			0.765$^\sharp$	0.686	0.807$^\sharp$	0.736	0.809$^\sharp$	0.764	0.823$^\sharp$	0.744
			0.1			0.768$^\sharp$	0.704	0.797$^\sharp$	0.740	0.857$^\sharp$	0.784	0.847$^\sharp$	0.806
			0.2			0.698$^\sharp$	0.616	0.781$^\sharp$	0.744	0.861$^\sharp$	0.846	0.875	0.872
	7	0.4	0	0.822$^\sharp$	0.810								
			0.05			0.758	0.750	0.778	0.778	0.826$^\sharp$	0.812	0.798$^\sharp$	0.782
			0.1			0.778$^\sharp$	0.760	0.794	0.788	0.826	0.826	0.846	0.832
			0.2			0.696	0.698	0.812$^\sharp$	0.786	0.874	0.870	0.896	0.894
		0.7	0	0.916$^\sharp$	0.880								
			0.05			0.858$^\sharp$	0.826	0.882$^\sharp$	0.846	0.892$^\sharp$	0.860	0.880$^\sharp$	0.850
			0.1			0.814$^\sharp$	0.786	0.894$^\sharp$	0.878	0.904$^\sharp$	0.870	0.936$^\sharp$	0.910
			0.2			0.776$^\sharp$	0.732	0.876$^\sharp$	0.856	0.900$^\sharp$	0.890	0.958$^\sharp$	0.944
		0.9	0	0.955$^\sharp$	0.918								
			0.05			0.935$^\sharp$	0.882	0.966$^\sharp$	0.932	0.946$^\sharp$	0.910	0.938$^\sharp$	0.912
			0.1			0.931$^\sharp$	0.876	0.956$^\sharp$	0.932	0.946$^\sharp$	0.932	0.962$^\sharp$	0.942
			0.2			0.893$^\sharp$	0.810	0.922$^\sharp$	0.900	0.956$^\sharp$	0.942	0.960$^\sharp$	0.950

$^\sharp$ indicates that LRM power is significantly higher than SM power at 5% with a McNemar's test.

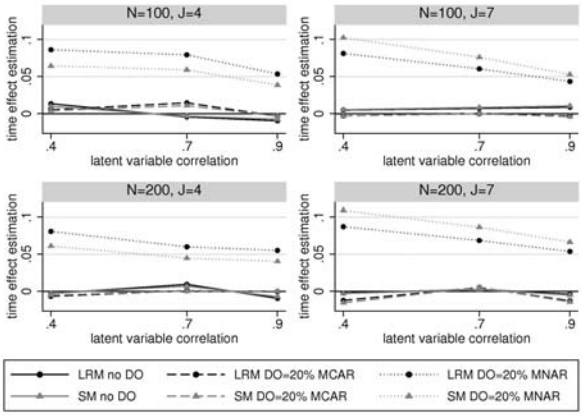

Figure 1: Time effect estimations between time 1 and time 2 for Score and Mixed models (SM) and Longitudinal Rasch Mixed model (LRM) methods for different values of sample size (N), number of items (J) and latent variable correlation (ρ_θ). No time effect simulated ($d_\theta = 0$). Data without dropout (DO), with 20% of MCAR dropout or 20% of MNAR dropout. Analyses performed with an AR(1) structure for the covariance matrix in SM method and an unstructured covariance matrix for LRM method.

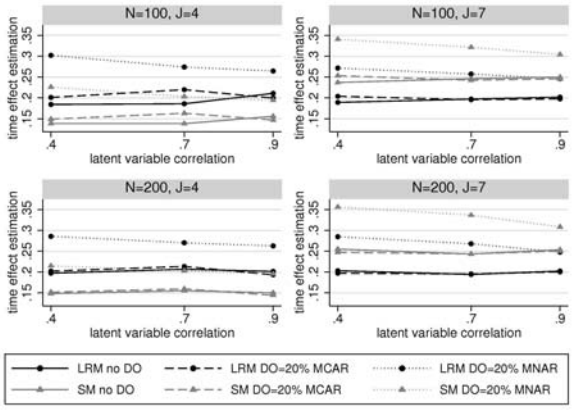

Figure 2: Time effect estimations between time 1 and time 2 for Score and Mixed models (SM) and Longitudinal Rasch Mixed model (LRM) methods for different values of sample size (N), number of items (J) and latent variable correlation (ρ_θ). Time effect simulated ($d_\theta = 0.2$). Data without dropout (DO), with 20% of MCAR dropout or 20% of MNAR dropout. Analyses performed with an AR(1) structure for the covariance matrix in SM method and an unstructured covariance matrix for LRM method.

complete data and data with MCAR dropout are also unbiased whatever the method. LRM and SM methods overestimate the time effect when MNAR dropout occurs compared to estimations on complete data.
The overestimation of the time effect grows with the proportion of dropout (results not shown). When $\pi = 5\%$, the time effect is overestimated for half of the cases with $\rho_\theta = -0.7$ or $\rho_\theta = -0.9$. At 10% of dropout, all the cases show an overestimation of time effect when $\rho_\theta = -0.7$ or $\rho_\theta = -0.9$. At 20% of dropout, all the cases show an overestimation of time effect when $\rho_\theta = -0.4$, $\rho_\theta = -0.7$ or $\rho_\theta = -0.9$.
Results on time effect estimation between time 2 and time 3 are comparable to results on time effect estimation between time 1 and time 2 (results not shown).

4 Discussion

Patient-Reported Outcomes are increasingly used in health sciences. Two methods to analyze longitudinal latent variables following a Rasch model were compared: Score and Mixed models (SM) and Longitudinal Rasch Model (LRM) methods. They have shown comparable results in term of type I error. LRM seemed to generally have a slightly but significantly higher power than SM when the latent variable correlation was medium or high. As expected, powers increased with the sample size, number of items and latent variable correlation.
Before choosing what method to use, statisticians have to keep in mind the underlying hypotheses and properties of each method. Although CTT has the advantage to be simple to use and to interpret, this approach produces ordinal measures (Embretson, 1996). In fact, only ordinal level measurement can be achieved with CTT, that is we can only observe that a person has a higher score than another. But the distance between two scores has no meaning because it depends on the population and the set of items. The Rasch model rests on strong assumptions but presents interesting psychometric properties: the exhaustivity of the score on the latent trait, the specific objectivity and the interval-level measurement. The interval-level measurement property means that the relative distance between two latent traits is maintained across questionnaires of different difficulties. In the Rasch model, the comparisons of the distance between pairs of person that have answered different set of items are possible.

Many assumptions had to be made to perform this simulation study. First, the data were assumed to follow a Rasch model because the purpose of this study was to compare CTT-based and Rasch-based models to determine which approach is the most adequate to analyze longitudinal latent variables when the data actually follow a Rasch model. The results of this study are only valid in this context. The choice of the Rasch model among all possible IRT models comes from its psychometric properties and its wide use in development and validation.
Second, the longitudinal Rasch mixed model used to simulate the data assumes a linear time effect for the evolution of the outcome and item parameters constant with time. It is reasonable to assume a linear time effect for long-term studies of quality of life occurring after treatment completion and where the quality of life is expected to improve. However, many studies include time points before and after treatment. With measurements before the beginning of the treat-

ment, at the end of the treatment and during follow-up period, the quality of life will probably decrease during the treatment period and increase after the completion of treatment leading to a non-linear time effect. Both methods estimate time effect for each period between two consecutive times without assuming that time effect is linear. LRM and SM can be used in the case where time effect decreases first and increases afterwards and we could expect similar results than for linear time effect.

This study aimed at evaluating the impact of type and proportion of dropout on CTT-based and Rasch-based approaches. For complete data and data with MCAR dropout, the type I error rates were well maintained to the expected 5%. As expected, MCAR case have shown close results to complete case due to the use of methods based on likelihood. However, a loss of power as well as an underestimation of the time effect has been observed for MCAR data in certain cases.

In the case where the missing data are non-ignorable, both methods have shown poor results. The type I error rates were not maintained to the expected 5% and could reach 10% in the worst cases, that is when the proportion and the informativity of dropout were high. Powers were also higher than complete case and increased with the proportion and informativity of the dropout as a direct result of the overestimation of the time effect. When MNAR data are encountered, both methods can't be used without taking into account the dropout process in the analysis. Several methods have been proposed in the litterature to handle missing data such as imputation, selection models and pattern-mixture models. Imputation consists in filling in the missing values using observed values. Several simple imputation methods exist based on information on the same subjects or from other subjects. Most of them lead to biased estimates and underestimation of the variability. In multiple imputation (Rubin, 2004), each missing value is replaced by several imputed values. Each imputed dataset is then analysed and the results are combined. This method overcomes the problems encountered with single imputation. Selection and pattern-mixture models are likelihood-based methos for handling missing data. These models combine linear model for the response with a suitable dropout model. Selection models and pattern-mixture models (Little, 1995) were proposed to model nonignorable nonresponse. They are an interesting way of dealing with MNAR missing data process by modeling explicitly the missing data mechanism. These models have to be used carefully because untestable assumptions have to be made on the missing data process for selection models and untestable identifying restrictions are used in pattern-mixture models to ensure their identifiability. For the use of such models, it is recommended to consider estimation coming from many selection/pattern-mixture models with different assumptions rather than only one model.

Missing at random (MAR) mechanism was not examined in this study. The dropout is said to be MAR if the dropout process does not depend on the missing values but only on the observed values. To simulate MAR dropout, the dropout propensity at time t ($\chi^{(t)}$) should be independent of the value of the latent variable at time t ($\theta^{(t)}$) and so $\rho_{\theta\chi}$ should be equal to 0. Furthermore, the dropout propensity at time t should be dependent on the previous values of latent variable already observed. The model could be extended by including it as a covariate. The MAR miss-

ing mechanism is ignorable, such as MCAR missing mechanism, if the separability condition is met (Verbeke and Molenberghs, 2000b). Both methods are likelihood-based analysis and ignores the dropout mechanism, we can expect that LRM and SM will lead to a valid analysis such as for MCAR case, provided that the parameters describing the measurement process are functionally independent of the parameters describing the dropout process.

This article focused on one type of missing data: the dropout. An important further development to this study concerns intermittent missing data. The missing data are intermittent if the patient has not answered all the items of the questionnaire. As dropout, intermittent missing data can lead to a loss of power and possible bias. The causes for intermittent missing data are multiple. For example, the patient may have not seen the question and the value is missing completely at random. The item can bother the patient because its content concerns religion, politics, sexual life. Thus, the patient choose not to answer this particular item and the missing data is informative. The informative dropout seems to be linked with the quality of life level of the patient whereas informative intermittent missing data might be related to the characteristics of the item. The ways to deal with intermittent missing data are complete case analysis, available case analysis, imputation. The complete case analysis only includes measurements that are complete in the analysis. The available case analysis uses as much data as possible to take advantage of all available information. Imputation methods can be simple or multiple. The problem that arises with intermittent missing data in CTT is the computation of the score. The most commonly used questionnaires (SF-36, EORTC QLQ-C30) have general guidelines regarding treatment of missing data. Generally, the score can still be computed if the patient has filled in half or more of the items of the scale. The patients with more missing items can't be used in the analysis. The Rasch model uses all available items in the analysis. In studies with a high proportion of intermittent missing data, we can expect that Rasch-based approach perform better than CTT-based approach because Rasch model may use more information than CTT. Furthermore, the property of specific objectivity of the Rasch model may ensure that the latent variable may be estimated consistently even for patients with missing items. We can expect that the occurrence of ignorable intermittent missing data (MCAR and MAR) will lead to valid analysis because both approaches are based on likelihood. Estimation problems will probably be observed for non-ignorable dropout but maybe in different extent for each approach.

In health sciences, longitudinal studies evaluating PRO often include two or more groups of patients in order to compare the evolution of the outcome between the groups. For example, study of a new strategy of treatment for a cancer may compare the long-term quality of life of patients receiving the new strategy and the long-term quality of life of patients who have received the usual strategy. This study has to be extended to compare both approaches in the context of longitudinal PRO collected in different groups of patients.

The purpose of this study was to compare CTT-based and Rasch-based models to determine which approach is the most adequate to analyze longitudinal latent variables when the

data were actually collected from a scale validated with a Rasch model. Data with informative dropout have shown parameter estimation problems. They have to be analyzed with appropriate methods taking into account of the dropout process such as selection models or pattern-mixture models. For complete data and data with non-informative dropout, the method of analysis based on the Rasch model may be preferred than the method based on CTT due to the generally observed slight gain of power and the psychometric properties of the Rasch model.

Acknowledgment

This work was supported by the Ligue Nationale Contre le Cancer.

References

Akaike, H. 1974. A new look at the statistical model identification, *IEEE Transactions on Automatic Control* **19**: 716–723.

Cella, D. F., Dineen, K., Arnason, B., Reder, A., Webster, K. A., Karabatsos, G., Chang, C., Lloyd, S., Steward, J. and Stefoski, D. 1996. Validation of the functional assessment of multiple sclerosis quality of life instrument, *Neurology* **47**(1): 129–139.

Diggle, P. and Kenward, M. G. 1994. Informative Drop-Out in longitudinal data analysis, *Applied Statistics* **43**(1): 49–93.

Embretson, S. E. 1996. The new rules of measurement, *Psychological Assessment* **8**(4): 341–349.

Fischer, G. H. and Molenaar, I. W. 1995. *Rasch models*, Springer.

Fitzmaurice, G. M., Davidian, M., Verbeke, G. and Molenberghs, G. 2009. *Longitudinal data analysis*, Chapman and Hall/CRC.

Hambleton, R. K. 2000. Emergence of item response modeling in instrument development and data analysis, *Medical Care* **38**(9 Suppl II): 60–65.

Lai, J., Cella, D., Kupst, M. J., Holm, S., Kelly, M. E., Bode, R. K. and Goldman, S. 2007. Measuring fatigue for children with cancer: Development and validation of the pediatric functional assessment of chronic illness Therapy-Fatigue (pedsFACIT-F), *Journal of Pediatric Hematology/Oncology* **29**(7): 471–479.

Laird, N. M. and Ware, J. H. 1982. Random-Effects models for longitudinal data, *Biometrics* **38**(4): 963–974.

Littell, R. C., Milliken, G. A., Stroup, W. W. and Wolfinger, R. D. 1996. *SAS system for mixed models*, SAS Institute, Inc., Cary, NC.

Little, R. J. A. 1995. Modeling the Drop-Out mechanism in Repeated-Measures studies, *Journal of the American Statistical Association* **90**(431): 1112–1121.

Little, R. J. A. and Rubin, D. B. 2002. *Statistical analysis with missing data*, Wiley.

McNemar, Q. 1947. Note on the sampling error of the difference between correlated proportions or percentages, *Psychometrika* **12**(2): 153–157.

Meiser, T. 2007. Rasch models for longitudinal data, *in* M. von Davier and C. H. Carstensen (eds), *Multivariate and Mixture Distribution Rasch Models*, Springer, New York, pp. 191–199.

Rasch, G. 1980. *Probabilistic models for some intelligence and attainment tests*, University of Chicago Press.

Rubin, D. B. 2004. *Multiple imputation for nonresponse in surveys*, Wiley-IEEE.

Sijtsma, K. and Hemker, B. T. 2000. A taxonomy of IRT models for ordering persons and items using simple sum scores, *Journal of Educational and Behavioral Statistics* **25**(4): 391–415.

Troxel, A. B., Fairclough, D. L., Curran, D. and Hahn, E. A. 1998. Statistical analysis of quality of life with missing data in cancer clinical trials, *Statistics in Medicine* **17**(5-7): 653–66.

Verbeke, G. and Molenberghs, G. 2000a. Inference for the marginal model, *Linear Mixed Models for Longitudinal Data*, Springer, New York, pp. 55–76.

Verbeke, G. and Molenberghs, G. 2000b. *Linear Mixed Models for Longitudinal Data*, corrected edn, Springer.

Zheng, X. and Rabe-Hesketh, S. 2007. Estimating parameters of dichotomous and ordinal item response models with gllamm, *Stata Journal* **7**(3): 313–333.