JOURNAL OF APPLIED MEASUREMENT, *15*(3),

Power Analysis on the Time Effect for the Longitudinal Rasch Model

M. L. Feddag

M. Blanchin

J. B. Hardouin

V. Sebille
Université de Nantes

Statistics literature in the social, behavioral, and biomedical sciences typically stress the importance of power analysis. Patient Reported Outcomes (PRO) such as quality of life and other perceived health measures (pain, fatigue, stress,...) are increasingly used as important health outcomes in clinical trials or in epidemiological studies. They cannot be directly observed nor measured as other clinical or biological data and they are often collected through questionnaires with binary or polytomous items. The Rasch model is the well known model in the item response theory (IRT) for binary data. The article proposes an approach to evaluate the statistical power of the time effect for the longitudinal Rasch model with two time points. The performance of this method is compared to the one obtained by simulation study. Finally, the proposed approach is illustrated on one subscale of the SF-36 questionnaire.

Requests for reprints should be sent to V. Sebille, EA 4275 ``Biostatistique, Pharmacoépidemiologie et Mesures Subjectives en Santé'', Faculté de Pharmacie, Université de Nantes, 1 rue Gaston Veil, 44035 Nantes, France.

Introduction

Patient Reported Outcomes (PRO) such as quality of life and other perceived health measures (pain, fatigue, stress,...) are increasingly used as important health outcomes in clinical trials or in epidemiological studies. They cannot be directly observed nor measured as other clinical or biological data and they are often collected through questionnaires with binary or polytomous items. Item response theory (IRT) models enable to model relationship between observed and latent traits (latent variables) where the probability of answering to each item is modelled as a function of the latent trait and item parameters. Two main statistical approaches can be used to analyze PRO data: classical test theory (CTT) based on observed scores and IRT models. The Rasch model (Rasch, 1960) is the most well known used for binary responses.

The class of psychometric models presented by Georg Rasch (Rasch 1960, 1961) has gained considerable interest and has stimulated an impressive amount of research on statistical models in the social and behavioral sciences. The impact of Rasch's work on modern test theory is documented in several monographs which summarize current developement in the mathematical modeling of test data as well as new perspective in the application of test models to social science issues (e.g., Kook and Varni (2008), Catz, Itzkovich, Tesio, Biering-Sorensen et al. (2007), Smith, Wright, Rush, Stark, Velikova and Selby (2006), Van der Linden and Hambleton (1997), Fischer and Molenaar (1995)).

There are many situations in which one is interested in investigating, for the same individuals, their change in a construct over time. It could be a psychologist who is interested in whether or not their patients symptoms of depression are dissipating with therapy, or an educator who is interested in the extent to which material is learned and retained by their students. A sociologist may be interested in the stability of one's attitudes with age or an administrator in higher education may be curious as to whether or not college students gain knowledge as a result of participation in a general education curriculum. To investigate such issues, typically a scale or test is administered to the same individuals over multiple time points. Such data can be modelled with the longitudinal Rasch model which comprises repeated observations of the same dichotomous items and the same sample of patients at different occasions.

Statistics literature in the social, behavioral, and biomedical sciences typically stress the importance of power analysis. By definition, the power of a statistical test is the probability that its null hypothesis (H_0) will be rejected given that it is in fact false. Obviously, significance tests that lack statistical power are of limited use because they cannot reliably discriminate between the credibility of the H_0 assumption and its nonrejection due to a lack of power.

For cross-sectional studies comparing two groups, Hardouin, Amri, Feddag and Sebille (2012) proposed the Raschpower procedure for the Rasch model to evaluate the power of the test of group effect. The power for detecting a prespecified group effect is determined for a given sample size, inter individual variability (variance of the latent trait) at level α (Julious, 2009; Chow, 2011). The aim of this paper is the extension of this work to the evaluation of the power on the test for time effect for the longitudinal Rasch model.

The outline of the paper is as follows. We firsly present the longitudinal Rasch model for binary data. Then the methodology for the evaluation of the power for the test of time effect is given. This approach is compared to the simulation one by using a simulation study. An example of real data from one dimension of the SF-36 questionnaire is presented. We finally conclude in the last section.

The Model

We consider binary responses of a questionnaire which is administered to the same subjects at various occasions $t, t = 1, \ldots, T$. In this framework, the responses are longitudinally correlated.

Let's consider a sample of N independent $(TJ \times 1)$ random multivariate binary observations

$$Y_i = \left(Y_i^{1'},\ldots,Y_i^{T'}\right)', i = 1,\ldots,N,$$

where

J is the number of items,

$Y_i = \left(Y_i^{1'},\ldots,Y_i^{T'}\right)'$, is the response vector of individual i to the questionnaire at time t, and

Y_{ij}^t is the binary variable response of individual i to item j at time $t, t = 1,\ldots,T$.

Let $Y = (Y_1,\ldots,Y_N)$ be the vector of the variables, θ_i^t the latent trait associated with subject i at time t and $\theta_i = \left(\theta_i^1,\ldots\theta_i^T\right)$ be the multidimensional latent trait for subject i. We denote by y a realization of the random variable Y. The longitudinal mixed Rasch model which has been considered by Feddag and Mesbah (2005), and Feddag and Bacci (2009), satisfies the following assumptions:

For all $i, j, t; i = 1,\ldots,N; j = 1,\ldots,J; t = 1,\ldots,T$, the probability distribution of the random variable Y_{ij}^t is given by:

$$P\left(Y_{ij}^t = y_{ij}^t;\theta_i^t,\delta_j\right) = \frac{e^{\left(\theta_i^t-\delta_j\right)y_{ij}^t}}{1=e^{\left(\theta_i^t-\delta_j\right)}}, \tag{1}$$

For a positive response, say $Y_{ij}^t = 1$, the probability defined by equation (1) is equal to:

$$\frac{e^{\left(\theta_i^t-\delta_j\right)y_{ij}^t}}{1+e^{\left(\theta_i^t-\delta_j\right)}}.$$

This is the definition of the classical Rasch model.

Given the latent trait θ_i, $i = 1,\ldots,N$, we have the conditional independence defined by:

$$P\left(Y_i^1 = y_i^1,\ldots Y_i^T = y_i^T;\theta_i,\delta\right)$$
$$= \prod_{t=1}^T \prod_{j=1}^J P\left(Y_{ij}^t = y_{ij}^t;\theta_i^t,\delta_j\right), \tag{2}$$

where δ_j is the item difficulty parameter associated to item j and δ $(\delta_1,\ldots,\delta_J)$.

The latent traits $\theta_1,\ldots,\theta_N$, are independent and identically normally distributed with mean vector $\mu = (\mu_t)'$, $t = 1,\ldots,T$ and covariance matrix $(\boldsymbol{\omega}) = (\sigma_{jl})$, $j, l = 1,\ldots,T$, where $\boldsymbol{\omega}$ is the vector parameter.

The model as formulated above is not identifiable, so some suitable restrictions have to be imposed. The classical constraint we made on the parameters is $\mu_1 = 0$.

We are interested in estimating the mean $\mu = (\mu_1,\ldots,\mu_T)'$ where the item difficulty parameters $\delta = (\delta_1,\ldots,\delta_J)$ and the covariance matrix $\Sigma(\omega)$ are considered as fixed. The marginal likelihood for the parameter μ is given by:

$$L(\mu;\delta,\omega,\mathrm{y}) = \prod_{i=1}^N \int_{R^T} \prod_{t=1}^T$$
$$\prod_{j=1}^J \frac{e^{\left(\theta_i^t-\delta_j\right)y_{ij}^t}}{1+e^{\left(\theta_i^t-\delta_j\right)}}\varphi\left(\theta_i,\mu,\omega\right)d\theta_i, \tag{3}$$

where

$$\varphi\left(\theta_{i,\mu,\omega}\right) = \frac{1}{(2\pi)^{\frac{T}{2}}\det\left(\Sigma(\omega)\right)^{\frac{1}{2}}}$$
$$\exp\left(-\frac{1}{2}(\theta_i-\mu)'\Sigma(\omega)^{-1}(\theta_i-\mu)\right),$$

is the multivariate normal density with mean μ and covariance matrix $\Sigma(\omega)$.

From now on, we consider only two time points in this model ($T = 2$), with the mean of the distribution of the latent variables given by $\mu = (\mu_1, \mu_2)'$ and its covariance matrix supposed to be equal to

$$\Sigma = \Sigma(\rho) = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

where ρ is the correlation between the two latent variables.

The main objective is to determine the power for detecting a given time effect $\gamma = \mu_2 - \mu_1$, assuming an expected correlation between the two latent variables ρ, given the sample size N. To do so, we also need the expected variance of γ. This goal is reached using two different approaches: The proposed method, which is based on an expected sample constructed from the 2^{2J} possible profiles of responses, and a simulation approach. We present in details the methodolgy developed in the following section.

Methodology

Let $X = (x^{(p)})$ the matrix of dimension $2^{2J} \times 2^{J}$, where

$$x^{(p)} = \left(x_{p1}^{1},\ldots,x_{pJ}^{1},x_{p1}^{2},\ldots,x_{pJ}^{2}\right),$$
$$x_{pj}^{t} = 0,1\left(t = 1,2,\ j = 1,\ldots,J\right),$$

is the p^{th} binary response's pattern associated to the longitudinal Rasch model with J items and 2 time points. From the model defined before, the probability of responding to $x^{(p)}$ is given by:

$$\pi_p = \int_{R^2} \prod_{t=1}^{2} \prod_{j=1}^{J} \frac{e^{\left(\theta_p^t - \delta_j\right)x_{pj}^t}}{1 + e^{\left(\theta_p^t - \delta_j\right)}} \varphi\left(\theta_p,\mu,\rho\right) d\theta_p. \quad (4)$$

This integral is approximated by Gauss-Hermite quadrature detailed as follows:

Gauss-Hermite quadrature

The probability given by (4) could be expressed by:

$$\begin{aligned}\pi_p &= \int_{R^2} \prod_{t=1}^{2} \prod_{j=1}^{J} \frac{e^{\left(\theta_p^t - \delta_j\right)x_{pj}^t}}{1 + e^{\left(\theta_p^t - \delta_j\right)}} \varphi\left(\theta_p,\mu,\rho\right) d\theta_p \\ &= \int_{R^2} h\left(x^{(p)} \mid \delta,\mu,\theta_p^*\right) \varphi\left(\theta_p^*,\rho\right) d\theta_p^*, \quad (5)\end{aligned}$$

where $\theta^*_p = \theta_p - \mu$.

We have used the Gauss Hermite quadrature by the standardizing transformation in the form $\theta^*_p = CZ$, where C is the lower triangular Cholesky factor for $\mathbf{\Sigma}(\Sigma = CC)'$ and Z is the standardized multivariate normally distributed random variable. Then the transformed version of the integral involved in (5) takes the form:

$$\begin{aligned}\pi_p &= \det\left(C\right) \int_{R^2} h\left(x^{(p)} \mid \delta,\mu,CZ\right) \varphi\left(CZ,\rho\right) dZ \\ &= \frac{1}{2\pi} \int_{R^2} h\left(x^{(p)} \mid \delta,\mu,CZ\right) \exp\left(-\frac{1}{2} Z'Z\right) dZ. \quad (6)\end{aligned}$$

Hence this integral is approximated by:

$$l_p\left(m\right) = \frac{1}{2\pi} \sum_{t_1^p=1}^{m} \sum_{t2=1}^{m} w_{t_1^p} w_{t_2^p} h\left(x^{(p)};\delta,\mu,C\left(d_{t_1^p},d_{t_2^p}\right)'\right),$$

where

$$\left(d_{t_1^p},d_{t_2^p}\right)' \text{ and } w_{t_1^p},w_{t_2^p}$$

are Gauss-Hermite quadrature nodes and weights, respectively. The accuracy of the approximation depends in the first place on the number of nodes m and on the proximity of $h(.)$ to a polynomial of degree $2m - 1$.

Expected sample

The expected frequency n_p for each pattern p is determined as follows:

First we evaluate $n_p^* = floor(N \times \pi_p)$ with $floor(x) = n$ if $n \leq x < n + 1$, where n is an integer. Then we calculate the number of unaffected frequencies $N^* = N - \Sigma_p\ n_p^*$ and thereafter we compute the residual probabilities $\pi_p^* = \pi_p - n_p^*/N$. Then the unaffected frequencies are distributed among all the N^* pattern having the greatest values of the residual probabilities π_p^* where we add 1 to the frequency. Thus $n_p - n_p^* + 1$ for these unaffected frequencies and $n_p - n_p^*$ for the others. Hence we construct the expected sample with size N, where each pattern p is repeated n_p times ($p = 1,\ldots,2^{2J}$).

This procedure named GH is similar to the one proposed by Hardouin et al. (2012) for the classical Rasch model but is faced with the problem of the number of possible reponses patterns 2^{2J} with large number of items J. For example, for 10 items the number of response patterns is more than one million, so we are faced to computational difficulties in creating the expected sample. Thereafter, an alternative method to GH is proposed and is named POPULATION method.

In this POPULATION method, a large dataset with n individuals and $2J$ items is simulated according to the longitudinal Rasch model defined above, where the probabilities π_p are approximated by GH only for the most frequent P responses patterns in this dataset. In our case, we use $n = 10^6$ and $P = 2N$ where N is the fixed number of individuals.

Once this expected dataset is created, it is analyzed by the longitudinal Rasch model, where the difficulty items parameters and covariance matrix between the two latent variables are fixed to the expected values. The variance of the difference between the two means $\gamma = \mu_2 - \mu_1$ is approximated using the Cramer-Rao bound. The Wald test and the power of this estimation are described in the next section.

Evaluation of the power of the Wald test

This difference between means γ can be tested by the Wald test (see Greenland (1983), Hardouin et al. (2012)). We assume the case of typical null hypothesis that implies that there is no difference between means at the two time points. This test is performed on the two hypotheses: H_0: $\gamma = 0$ and H_0:$\gamma \neq 0$, and the statistic test defined by

$$\frac{\gamma}{\left(Var(\gamma)\right)^{1/2}},$$

where $Var(\gamma)$is the variance of γ. The null hypothesis is rejected at level α if

$$\frac{|\hat{\gamma}|}{\left(Var(\hat{\gamma})\right)^{1/2}} > z_{1-\alpha/2},$$

where $z_{1-\alpha/2}$ is the quantile of the cumulative standard normal distribution function, $\hat{\gamma}$ and Var($\hat{\gamma}$) are respectively the estimate of γ and its variance. The expected power of this test which is based on the Cramer-Rao bound is evaluated as follows:

$$1-\hat{\beta}_{CR} = 1\phi\left(z_{1-\frac{\alpha}{2}} - \frac{\hat{\gamma}}{\left(Var(\hat{\gamma})\right)\frac{1}{2}}\right) + \phi\left(z_{1-\frac{\alpha}{2}} - \frac{\hat{\gamma}}{\left(Var(\hat{\gamma})\right)\frac{1}{2}}\right), \quad (7)$$

where ϕ is the cumulative standard normal distribution.

Assuming $\gamma > 0$, the second part of the right hand side of this equation is close to 0, thus this power is approximated as follows:

$$1-\hat{\beta}_{CR} \simeq 1-\phi\left(z_{1-\frac{\alpha}{2}} - \frac{\hat{\gamma}}{\left(Var(\hat{\gamma})\right)\frac{1}{2}}\right). \quad (8)$$

Simulation study

A simulation study has been conducted and the different results including the estimate of the variance of γ and the statistical power are compared to those obtained from the proposed approach. The different parameters considered in this study which is based on 1000 datasets for each case are as follows:

- The number of individuals $N = 50, 100, 200, 300$ and 500.
- The number of items $J = 5, 10$ where the items difficulties are fixed as $\delta = (-1,-0.5,0,0.5,1)$ and $\delta = (-2,-1.5,-1,-0.5,0,0.5,1,1.5,2)$.
- Three values of the mean μ: (0,0.2), (0,0.5), (0,0.8). Hence the different values of γ are: 0.2, 0.5 and 0.8.
- Three values of ρ: 0.4, 0.7 and 0.9.

Estimation of the power

The estimation of the parameter γ using simulated data is obtained by the maximization of the marginal likelihood given by (3), where the integral of dimension 2 is approximated by Gauss-Hermite quadrature. The estimate of its variance is obtained, then we deduce the significance of the Wald test under H_1 for this dataset. Hence, the power estimate denoted by $1 - \beta_S$ is the rate of significant Wald tests under H_1 over the M simulated datasets. This approach is compared to the proposed one described above.

Results

The estimation of $Var(\hat{\gamma})$ and the statistical power of the Wald test obtained by these two approaches are given in Tables 1-3 respectively for ρ: 0.4, 0.7 and 0.9. These two quantities are denoted by $Var_{CR}(\hat{\gamma})$ and $(1 - \hat{\beta}_{CR})$for the proposed approach and by $Var_S(\hat{\gamma})$ and $(1 - \beta_S)$for the simulation method.

As shown in Tables 1-3, the values $Var_S(\hat{\gamma})$ and $Var_{CR}(\hat{\gamma})$ decreases as N, J and ρ increase and remain stable for all the values of γ. Hence the power follows the same trend. It increases with N, J and ρ as expected with γ.

With $N = 50$, $J = 5$, 10 and for the three values of ρ, $Var_{CR}(\hat{\gamma})$, are greater than $Var_S(\hat{\gamma})$, so with this approach the variances are largely estimated. This large estimation of the variance leads to a lower estimate of the power compared to the simulation approach. In this case the largest difference between the powers is equal to 9.5% and it correponds to $\gamma = 0.5$, $J = 10$ and $\rho = 0.4$.

With $N = 100$, and for all the cases considered, $Var_{CR}(\hat{\gamma})$are greater or equal than $Var_S(\hat{\gamma}$)and the largest difference between the powers is equal to 0.024, which corresponds to $\gamma = 0.2$, $J = 10$ and $\rho = 0.7$.

For the three other sample sizes, the estimated $Var_{CR}(\hat{\gamma})$ are very close to those obtained by simulation. They follow the same trend than those obtained by simulation. They decrease with N, J and ρ remain stable for all values of γ. For these cases, the difference between the two powers is negligible, so the two methods provide similar results.

Regarding these results, we can say that the proposed method provides a reliable estimation of the power for the test on the time effect even for moderate sample sizes, say N equal to 100.

Example

The study deals with the prospective evaluation of quality of life and nonspecific symptoms before and after cure of primary hyperparathyroidism (Caillard, Sebag, Mathonnet, Gibelin, Brunaud, Loudot et al., 2007).

This prospective study, which took place from November 2007 to June 2011 at the university hospital of Nantes includes only cured

Table 1

Estimates of $Var(\hat{\gamma})$ and the power 1-β obtained with the proposed method and using simulation method for ρ = 0.4

N	J	True γ	$Var_{CR}(\hat{\gamma})$	$1-\hat{\beta}_{CR}$	$Var_S(\hat{\gamma})$	$1-\beta_S$
50		0.2	0.041	0.165	0.040	0.172
	5	0.5	0.044	0.667	0.040	0.724
		0.8	0.047	0.957	0.042	0.977
		0.2	0.031	0.204	0.030	0.212
	10	0.5	0.038	0.730	0.030	0.825
		0.8	0.043	0.970	0.031	0.996
100		0.2	0.020	0.290	0.020	0.290
	5	0.5	0.021	0.932	0.020	0.937
		0.8	0.022	1.000	0.021	1.000
		0.2	0.015	0.362	0.015	0.373
	10	0.5	0.017	0.968	0.015	0.989
		0.8	0.020	1.000	0.015	1.000
200		0.2	0.010	0.516	0.010	0.507
	5	0.5	0.010	0.999	0.010	1.000
		0.8	0.011	1.000	0.010	1.000
		0.2	0.008	0.624	0.007	0.641
	10	0.5	0.008	1.000	0.008	1.000
		0.8	0.009	1.000	0.008	1.000
300		0.2	0.007	0.691	0.007	0.682
	5	0.5	0.007	1.000	0.007	1.000
		0.8	0.007	1.000	0.007	1.000
		0.2	0.005	0.798	0.005	0.802
	10	0.5	0.005	1.000	0.005	1.000
		0.8	0.006	1.000	0.005	1.000
500		0.2	0.004	0.887	0.004	0.894
	5	0.5	0.004	1.000	0.004	1.000
		0.8	0.004	1.000	0.004	1.000
		0.2	0.003	0.952	0.003	0.963
	10	0.5	0.003	1.000	0.003	1.000
		0.8	0.003	1.000	0.003	1.000

patients. All patients were given preoperative and postoperative French version of the SF-36 questionnaire (Leplège, Ecosse, Pouchot, Coste and Perneger, 2001) at 3, 6 and 12 months to evaluate quality of life and nonspecific symptoms.

The proposed approach is illustrated on the Role Physical (RP) dimension of this SF-36 questionnaire at the first two time points, where 140 patients answered to this dimension. It contains four dichotomous items given as follows:

1. Cut down the amount of time you spent on work or other activities
2. Accomplished less than you would like
3. Were limited in the kind of work or other activities
4. Had difficulty performing the work or other activities (for example, it took extra effort)

From the longitudinal Rasch model, the difficulty item parameters are estimated as $\hat{\delta}=$ (–0.715, 1.149, –0.179, 0.155), the time effect as $\hat{\mu}_2=1.589$ and the covariance matrix of the latent traits is estimated as $\Sigma = (11.167, 9.027,$

Table 2

Estimates of Var($\hat{\gamma}$) and the power 1-β obtained with the proposed method and using simulation method for ρ = 0.7

N	J	True γ	$Var_{CR}(\hat{\gamma})$	$1\text{-}\hat{\beta}_{CR}$	$Var_S(\hat{\gamma})$	$1\text{-}\beta_S$
50		0.2	0.037	0.177	0.036	0.179
	5	0.5	0.040	0.706	0.037	0.742
		0.8	0.045	0.964	0.038	0.987
		0.2	0.027	0.229	0.026	0.232
	10	0.5	0.032	0.803	0.026	0.883
		0.8	0.037	0.987	0.027	0.999
100		0.2	0.019	0.312	0.018	0.317
	5	0.5	0.019	0.949	0.019	0.960
		0.8	0.021	1.000	0.019	1.000
		0.2	0.013	0.410	0.013	0.434
	10	0.5	0.015	0.984	0.013	0.994
		0.8	0.017	1.000	0.013	1.000
200		0.2	0.009	0.551	0.009	0.546
	5	0.5	0.009	0.999	0.010	1.000
		0.8	0.010	1.000	0.010	1.000
		0.2	0.007	0.691	0.006	0.670
	10	0.5	0.007	1.000	0.007	1.000
		0.8	0.008	1.000	0.007	1.000
300		0.2	0.006	0.726	0.006	0.686
	5	0.5	0.006	1.000	0.006	1.000
		0.8	0.007	1.000	0.006	1.000
		0.2	0.004	0.854	0.004	0.858
	10	0.5	0.005	1.000	0.004	1.000
		0.8	0.005	1.000	0.004	1.000
500		0.2	0.004	0.912	0.004	0.927
	5	0.5	0.004	1.000	0.004	1.000
		0.8	0.004	1.000	0.004	1.000
		0.2	0.003	0.974	0.003	0.979
	10	0.5	0.003	1.000	0.003	1.000
		0.8	0.003	1.000	0.003	1.000

9.127, 17.896). We deduce that the two latent variables are highly correlated with correlation equal to 0.638.

The Rasch power procedure under Stata defined by Hardouin et al. (2012) is extended to the longitudinal Rasch model.

An example of syntax of this procedure and its outputs are given in Figure 1. With these estimates, the proposed approach evaluates the difference between means at 1.71 its variance at 0.0822 and the power using equation (8) is equal to 0.9998. We note that the large estimate of the covariance matrix is due to the considerable variability between patients. With this estimated power, the test reliably discriminate between H_0 and the alternative hypothesis H_1.

Discussion

In this paper we have proposed a power analysis on the time effect for the longitudinal Rasch model which is an extension of the one proposed by Hardouin et al. (2012) for the Rasch model with two groups. It estimates the power of the Wald test for the difference between the two

Table 3

Estimates of Var($\hat{\gamma}$) and the power 1-β obtained with the proposed method and using simulation method for ρ = 0.9

N	J	True γ	$Var_{CR}(\hat{\gamma})$	$1-\hat{\beta}_{CR}$	$Var_S(\hat{\gamma})$	$1-\beta_S$
50		0.2	0.034	0.188	0.033	0.191
	5	0.5	0.037	0.741	0.034	0.773
		0.8	0.040	0.979	0.035	0.994
		0.2	0.023	0.264	0.022	0.292
	10	0.5	0.027	0.863	0.022	0.918
		0.8	0.030	0.996	0.023	1.000
100		0.2	0.017	0.335	0.017	0.357
	5	0.5	0.018	0.964	0.017	0.972
		0.8	0.019	1.000	0.018	1.000
		0.2	0.011	0.473	0.011	0.481
	10	0.5	0.012	0.994	0.011	0.996
		0.8	0.014	1.000	0.011	1.000
200		0.2	0.008	0.587	0.008	0.614
	5	0.5	0.009	1.000	0.008	1.000
		0.8	0.009	1.000	0.009	1.000
		0.2	0.006	0.763	0.005	0.784
	10	0.5	0.006	1.000	0.005	1.000
		0.8	0.007	1.000	0.006	1.000
300		0.2	0.006	0.765	0.006	0.774
	5	0.5	0.006	1.000	0.006	1.000
		0.8	0.006	1.000	0.006	1.000
		0.2	0.004	0.908	0.004	0.908
	10	0.5	0.004	1.000	0.004	1.000
		0.8	0.004	1.000	0.004	1.000
500		0.2	0.003	0.934	0.003	0.927
	5	0.5	0.003	1.000	0.003	1.000
		0.8	0.004	1.000	0.004	1.000
		0.2	0.002	0.989	0.002	0.989
	10	0.5	0.002	1.000	0.002	1.000
		0.8	0.002	1.000	0.002	1.000

```
matrix matdiff=(-0.715,1.149,-0.179,0.155)
matrix matvar=(11.167,9.027 \ 9.027,17.896)
raschpower, longitudinal n0(140) gamma(1.589) diff(matdiff) var(matvar)

Method: GH
Number of individuals at each time: 140

Time effect: 1.589

Variance matrix of the latent trait:
symmetric matvar[2,2]
      c1          c2
r1  11.167
r2   9.027      17.896

Number of items: 4

Difficulties parameters of the items:
         item1   item2   item3   item4
delta_1  -.715   1.149   -.179    .155

Number of studied response's patterns: 256
10%..20%..30%..40%..50%..60%..70%..80%..90%..100%
-----------------------------------------------------------------------------
                                              Estimation with the
                                       Cramer-Rao bound    classical formula
----------------------------------------------------------------------------
Estimated value of the time effect           1.71
Estimation of the s.e. of the time effect     0.29
Estimation of the variance of the time effect 0.0822
Estimation of the power                      0.9998          0.9925
Number of patients for a power of 99.98%     140/140        52.06/ 52.06
Ratio of the number of patients              2.69
```

Figure 1

means of the latent variables for two time points.

This method involves the Gauss-Hermite quadrature calculations for the probabilities of the different patterns of responses. The expected data set is obtained using these approximations. When the number of items is small, the number of the possible patterns of responses is relatively moderate, we recommend to use the GH approach. However, when the number of items is large, the number of possible patterns of responses becomes very large and in this case, the POPULATION approach is preferred.

This proposed method is compared to the simulations and it is shown that they provide similar results even for moderate values of N, say N equal to 100. And in time running the GH approach is less fast than the POPULATION one when the number of items is relatively large. Regarding the different results, this proposed method provides a reliable estimation of the power of the Wald test for the difference between means. It is shown that the power increases as expected with the number of individuals, the number of items, the correlation between the two latent variables and the value of the difference between means.

It could be interesting to extend this work to the same model with more than two time points and to the longitudinal polytomous items, which are modelized by the partial credit model. In this last case, the number of possible responses patterns depends on the number of items and on the number of modalities per each item.

Acknowledgements

The authors gratefully acknowledge support from the ANR (Agence nationale de la recherche) grant N^0 PRSP 00801, held at the university of Nantes.

References

Caillard, C., Sebag, F., Mathonnet, M., Gibelin, H., Brunaud, L., Loudot, C. et al. (2007), Prospective evaluation of quality of life (SF-36) and nonspecific symptoms before and after cure of primary hyperparathyroidism (1 year follow up). *Surgery*, *141*, 153-159.

Catz, A., Itzkovich, M., Tesio, L., Biering-Sorensen, F., Weeks, C., Laramee, M.T. et al. (2007), A multicenter international study on the spinal cord independence measure, Version III. Rasch psychometric validation. *Spinal Cord*, *45*, 275-291.

Chow, S. C. (2011), Sample size calculations for clinical trials. Wiley interdisciplinary reviews: *Computational statistics*, *3*(5), 414-427.

Feddag, M. L., and Mesbah, M. (2005), Generalized estimating equations for longitudinal mixed Rasch model. *Journal of Statistical Planning and Inference*, *129*, 159-179.

Feddag, M. L., and Bacci, S. (2009), Pairwise likelihood for Longitudinal Mixed Rasch Model. *Computational Statistics and Data Analysis*, *53*, 1027-1037.

Fischer, G. H., and Molenaar, I. W. (1995), *Rasch models foundations, recents developments and applications*. New York, Springer-Verlag.

Greenland, S. (1983), Tests for interaction in epidemiologic studies: a review and a study of power. *Statistic in Medicine*, *2*, 243-251.

Hardouin, J. B., Amri, S., Feddag, M. L., and Sebille, V. (2012), Towards power and sample size calculations for the comparison of two groups of patients with item response theory models. *Statistics in Medicine*, *31*, 1277-1290.

Julious, S. A. (2009), *Sample sizes for clinical trials*. Boca Raton, FL: CRC Press—Taylor and Francis.

Kook, S. H., and Varni, J. W. (2008), Validation of the Korean version of the pediatric quality of life inventory{TM 4.0 (PedsQLTM) generic core scales in school children and adolscents using the Rasch model. *Health and Quality of Life Outcomes*, 6-41.

Lepl\ège, A., Ecosse, E., Pouchot, J., Coste, J. and Perneger, T. (2001), Le Questionnaire MOS SF-36. Manuel de l'utilisateur et guide d'interprétation des scores. Paris: Editions Scientifique Techniques et Médicales.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Danish Institute for Educational Research. (Expanded edition, 1980. Chicago: University of Chicago Press.)

Rasch, G. (1961), *On general laws and the meaning of measurement in psychology*. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, IV. Berkeley: University of California Press.

Smith, A. B., Wright, E. P., Rush, R., Stark, D. P., Velikova, G., and Selby, P. J. (2006), Rasch analysis of the dimensional structure of the hospital anxiety and depression scale. *Psycho-oncology*, *15*(9), 817-827.

van der Linden, W. J., and Hambleton, R.K. (1997), *Handbook of modern item response theory*. New York: Springer-Verlag.