



**HAL**  
open science

## Continuous-time modeling of self-reported outcome data: a dynamic Item Response Theory model

Cécile Proust-Lima, Viviane Philipps, Bastien Perrot, Myriam Blanchin,  
Véronique Sébille

► **To cite this version:**

Cécile Proust-Lima, Viviane Philipps, Bastien Perrot, Myriam Blanchin, Véronique Sébille. Continuous-time modeling of self-reported outcome data: a dynamic Item Response Theory model. *Methods*, inPress, 10.1016/j.ymeth.2022.01.005 . hal-03355105

**HAL Id: hal-03355105**

**<https://hal.science/hal-03355105>**

Submitted on 27 Sep 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Continuous-time modeling of self-reported outcome data: a dynamic Item Response Theory model

Cécile Proust-Lima\*

*Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, UMR1219, F-33000  
Bordeaux, France*

Viviane Philipps

*Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, UMR1219, F-33000  
Bordeaux, France*

Bastien Perrot

*Université de Nantes, Université de Tours, INSERM, SPHERE U1246, Nantes, France  
Methodology and Biostatistics Unit, CHU Nantes, Nantes, France*

Myriam Blanchin

*Université de Nantes, Université de Tours, INSERM, SPHERE U1246, Nantes, France*

Véronique Sébille

*Université de Nantes, Université de Tours, INSERM, SPHERE U1246, Nantes, France  
Methodology and Biostatistics Unit, CHU Nantes, Nantes, France*

---

## Abstract

Item Response Theory (IRT) models have received growing interest in health science for analyzing latent constructs such as depression, anxiety, quality of life or cognitive functioning from the information provided by each individual's items responses. However, in the presence of repeated item measures, IRT methods usually assume that the measurement occasions are made at the exact same time for all patients. In this paper, we show how the IRT methodology can be combined with the mixed model theory to provide a dynamic IRT model

---

\*Corresponding author

*Email address:* [cecile.proust-lima@inserm.fr](mailto:cecile.proust-lima@inserm.fr) (Cécile Proust-Lima)

which exploits the information provided at item-level for a measurement scale while simultaneously handling observation times that may vary across individuals. The latent construct is a latent process defined in continuous time that is linked to the observed item responses through a measurement model at each individual- and occasion-specific observation time; we focus here on a Graded Response Model for binary and ordinal items. The Maximum Likelihood Estimation procedure of the dynamic IRT model is available in the R package `lcmm`. The proposed approach is contextualized in a clinical example in end-stage renal disease, the PREDIALA study. The objective is to study the trajectories of depressive symptomatology (as measured by 7 items of the Hospital Anxiety and Depression scale) according to the time on renal transplant waiting list and the renal replacement therapy. We also illustrate how the method can be used to assess Differential Item Functioning and lack of measurement invariance over time.

*Keywords:* Item Response Theory, Mixed Model, Longitudinal data, Measurement Invariance, Latent Process Model

---

### **Highlights**

- The dynamic IRT model provides a flexible solution to analyze repeated item responses measuring a latent construct over time
- The dynamic IRT model relies on a mixed model to capture the continuous-time nature of the underlying construct
- The dynamic IRT model can investigate the item psychometric properties and measurement invariance
- Estimation of the dynamic IRT model is made available in the R package `lcmm` with a companion vignette

- The case study describes the depressive symptomatology of patients with end-stage renal disease on the transplant waiting list

## 1. Introduction

Item Response Theory (IRT) models, which exploit the information provided by each individual's items responses, have received growing interest in health science for capturing latent constructs of interest such as depression, anxiety, fatigue, quality of life, or cognitive functioning [1, 2, 3, 4]. IRT models have interesting properties compared to models coming from classical measurement theory such as Classical Test Theory (CTT) models which aggregate the items into a global score or score per domain. In particular, CTT produces ordinal measurements while IRT generates interval measurements. Hence, with IRT, a unit difference characterizes the same amount when measured from different initial levels on the latent construct scale. IRT also allows for a finer granularity of the level of analysis, done at the item level, which enables a better understanding of the item psychometric properties, and an in depth description of patients' experience.

In health research, the interest often lies in the longitudinal changes of latent constructs. Examples include the trajectory of anxiety or fatigue in clinical research [5, 6] or the trajectory of functional dependency in epidemiological research on aging [7] based on the observed patients' repeated responses to questionnaires either self-reported (and named Patient-Reported Outcomes (PRO)) or reported by the clinician (and named Clinician-Reported Outcomes (CRO)). IRT models have been extended to account for repeated item measurements but most of the time, the measurement occasions are necessarily considered as occurring at the exact same time for all patients [8]. This is, however, rarely the case in practice. For instance, in cohort studies, even though visits are planned,

25 the exact timing may differ substantially across individuals. The timescale may  
also be the time from a specific health event (e.g. diagnosis, registration on a  
waiting list), independent from the time of study inclusion, so that the timescale  
becomes per se continuous. This is the case in aging studies where trajectories  
are assessed according to age, or more generally in all the situations where en-  
30 try in the study does not correspond to a clearly defined time zero. Models  
from linear mixed model theory [9] are particularly suited for the analysis of  
outcomes repeatedly measured over time. They can model the outcome trajec-  
tory in continuous time while accounting for the within-subject correlation. We  
describe in this paper how IRT modeling can be combined with linear mixed  
35 model theory for the analysis of item responses measured repeatedly over time  
when observation times vary across individuals. We show how to operational-  
ize the latent construct as a latent process defined in continuous time and how  
to link it with the observed items responses through a measurement model for  
graded response, the Graded Response Model (GRM) [10], at each individual-  
40 and occasion-specific observation time. Beyond the description of a latent con-  
struct trajectory over time from repeated item measurements, this dynamic IRT  
model may also help to assess the item and scale properties as done with cross-  
sectional IRT methods, and investigate lack of measurement invariance, that  
is Differential Item Functioning (DIF) between groups, or Response Shift (RS)  
45 over more than two measurement times.

The proposed approach is contextualized in a clinical example in end-stage  
renal disease, the PREDIALA study [11, 12]. The PREDIALA study aims  
at studying the experience of patients with end-stage renal disease (e.g. qual-  
ity of life, anxious and depressive symptoms) on the renal transplant waiting  
50 list. From a psychological perspective, the waiting list period can be long and  
anxiety-provoking for patients because of the uncertainty of waiting, the hope

of being called for a transplant and the disappointment, and sometimes distress, of not being called. In addition, the onset or worsening of depressive/anxious symptoms can occur as the time on the waiting list increases [13]. Moreover, patients' experience may also differ according to their renal replacement therapy, that is, whether they are dialyzed or not (i.e. preemptive) [12]. As patients with chronic diseases have to live and adapt to their illness, its stability or progression over time, they may also understand or interpret the items of the questionnaires differently according to socio-demographic or clinical characteristics (e.g. type of renal replacement therapy) or over time, despite having similar health outcomes. The former situation may induce DIF [14], while the latter may produce RS [15]. From a methodological perspective, the relevant timescale was the time elapsed since registration on the waiting list. Since entry in the study occurred after different periods of time spent on the waiting list across individuals, this timescale substantially differed from the classical follow-up time. It induces large inter-individual variations of measurement times between patients and calls for a dynamic IRT model that can handle individual-specific measurement times. Such models can also enable the investigation of measurement invariance between groups (DIF) or over time (RS).

## 2. Methods

### 2.1. Dynamic IRT model

In a sample of  $N$  subjects, let consider a set of  $K$  items belonging to the same scale that are measured repeatedly over time, with  $Y_{ikj}$  designating the response to item  $k$  ( $k = 1, \dots, K$ ) for subject  $i$  ( $i = 1, \dots, N$ ) at repeated occasion  $j$  ( $j = 1, \dots, n_{ki}$ ). Note that here, we consider the general framework where the number of measurements may differ from one individual to the other (and possibly from one item to the other) so that item measurement  $Y_{ikj}$  is to be

associated with its actual time of measurement  $t_{ikj} \in \mathbb{R}$ .

### 2.1.1. Structural model

80 As in classical IRT methodology, we assume that the  $K$  items measure the same underlying construct called  $\Lambda$ . The major difference is that in a longitudinal IRT setting, the construct is now a latent process  $(\Lambda_i(t))_{t \in \mathbb{R}}$  defined in continuous time.

Its trajectory over time can be described by a linear mixed model (LMM) to 85 account for the within individual correlation (and between-individual variability):

$$\Lambda_i(t) = \mathbf{X}_i(\mathbf{t})^\top \boldsymbol{\beta} + \mathbf{Z}_i(\mathbf{t})^\top \mathbf{b}_i, \quad \forall t \in \mathbb{R} \quad (1)$$

where  $\mathbf{X}_i(\mathbf{t})$  is a vector of variables including functions of time  $t$ , associated with parameters  $\boldsymbol{\beta}$ , which describes the shape of trajectory over time of the construct of interest at the population level and its association with covariates; 90  $\mathbf{Z}_i(\mathbf{t})$  is a vector of variables that almost always includes exclusively functions of time and is associated with the vector of individual random-effects  $\mathbf{b}_i$  with  $\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{B})$  ( $\mathbf{B}$  is usually left unstructured). This second part  $\mathbf{Z}_i(\mathbf{t})^\top \mathbf{b}_i$  models the individual departure from the mean trajectory  $\mathbf{X}_i(\mathbf{t})^\top \boldsymbol{\beta}$ . In some contexts, a Gaussian stochastic process may be added to better reflect the local 95 variations in the individual trajectories. For sake of readability, we do not cover this aspect in the present paper.

Identifiability constraint is added to this model to determine the dimension of the latent process. Usually, the intercept is removed from the model (i.e.,  $\mathbf{X}_i(\mathbf{t})$  does not include any intercept) so that the mean of the latent process 100 is 0 in the category of reference, and the variance of the first random-effect is constrained to 1. This first random effect being usually an intercept, this

corresponds to assuming that the conditional variance given the covariates is 1 at time  $t = 0$ .

### 2.1.2. Measurement model

105 The latent process is linked to the observations of the  $K$  items using an item-specific measurement model. In this work, we focus mainly on binary and ordinal items even though the methodology could also apply to continuous items (see Proust-Lima et al. [16] for more details). We assume that item  $k$  is defined with  $L_k$  ordinal levels from 0 to  $L_k - 1$ .

110 The probability to observe the level  $l$  for item  $Y_{ikj}$  is defined by a cumulative probit model:

$$\mathbb{P}(Y_{ikj} \leq l \mid \Lambda_i(t_{ikj})) = \Phi(\sigma_k^- (\eta_{kl+1} - \Lambda_i(t_{ikj}))) \quad (2)$$

where  $\Phi$  is the Gaussian cumulative distribution function,  $\sigma_k^-$  is a parameter defining the discrimination of item  $k$  and  $(\eta_{kl})_{l=1, \dots, L_k-1}$  are the  $L_k - 1$  location parameters which correspond to the thresholds defining the change in the successive levels of item  $k$ . For an ordinal item, we assume  $\eta_{k1} \leq \eta_{k2} \leq \dots \leq \eta_{kL_k-1} \leq \eta_{kL_k} = +\infty$ .

Equation (2) comes from the idea that item  $k$  takes the level  $l$  if the underlying construct plus a measurement error  $\epsilon_{ikj}$  of variance  $\sigma_k^2$  lies in the interval  $[\eta_{kl}, \eta_{kl+1})$  as shown below:

$$\begin{aligned} \mathbb{P}(Y_{ikj} = l \mid \Lambda_i(t_{ikj})) &= \mathbb{P}(\eta_{kl} \leq \Lambda_i(t_{ikj}) + \epsilon_{ikj} < \eta_{kl+1}) \\ &= \mathbb{P}\left(\frac{\eta_{kl} - \Lambda_i(t_{ikj})}{\sigma_k} \leq \frac{\epsilon_{ikj}}{\sigma_k} < \frac{\eta_{kl+1} - \Lambda_i(t_{ikj})}{\sigma_k}\right) \\ &= \mathbb{P}\left(\frac{\epsilon_{ikj}}{\sigma_k} < \frac{\eta_{kl+1} - \Lambda_i(t_{ikj})}{\sigma_k}\right) - \mathbb{P}\left(\frac{\epsilon_{ikj}}{\sigma_k} < \frac{\eta_{kl} - \Lambda_i(t_{ikj})}{\sigma_k}\right) \end{aligned} \quad (3)$$



120 By considering that the measurement error  $\epsilon_{ikj}$  is Gaussian, it induces that

$$\mathbb{P}(Y_{ikj} = l \mid \Lambda_i(t_{ikj})) = \Phi\left(\frac{\eta_{kl+1} - \Lambda_i(t_{ikj})}{\sigma_k}\right) - \Phi\left(\frac{\eta_{kl} - \Lambda_i(t_{ikj})}{\sigma_k}\right) \quad (4)$$

and Equation (2) comes by denoting  $\sigma_k^- = \frac{1}{|\sigma_k|}$ .

Note that, by considering logistic errors instead of Gaussian errors, one would obtain the logistic ogive model, also very popular in IRT methodology.

Equation (2) defines a model for graded response, also known as Graded Re-  
125 sponse Model (GRM) [10, 17]. We focus on this type of model in the remaining  
of the work but acknowledge that any alternative measurement model could be  
considered instead depending on the distributional assumption and the items  
type (e.g., binary, continuous). See for instance Saulnier et al. in the same  
special [REF] issue and Barbieri et al. [18].

### 130 2.1.3. Measurement invariance

The dynamic IRT model defined with equations (1) and (2) assumes that  
all the items have the same functioning, meaning that the common underlying  
construct captures all the information of the items and what remains specific  
to the item is only its location and discrimination/error; this is made clear  
135 with equation (3). However, sometimes a different functioning of the items may  
be suspected according to a covariate at a given time or over time. In IRT  
methodology, this is called differential item functioning (DIF) or Response Shift  
(RS) when lack of measurement invariance occurs over time. DIF and RS can  
also be investigated in the dynamic IRT by completing the measurement model

140 as follows:

$$\mathbb{P}(Y_{ikj} \leq l \mid \Lambda_i(t_{ikj})) = \Phi(\sigma_k^- (\eta_{kl} - (\Lambda_i(t_{ikj}) + (\mathbf{X}_i^{\text{DIF}}(\mathbf{t}_{ikj}))^\top \boldsymbol{\gamma}_k))) \quad (5)$$

where  $\mathbf{X}_i^{\text{DIF}}(\mathbf{t})$  is the vector of covariates for which a DIF is suspected and  $\boldsymbol{\gamma}_k$  the associated parameters. When  $\mathbf{X}_i^{\text{DIF}}(\mathbf{t})$  is also part of the structural model in Equation (1) (i.e.,  $\mathbf{X}_i^{\text{DIF}}(\mathbf{t}) \subset \mathbf{X}_i(\mathbf{t})$ ), an identifiability constraint is added to the  $\boldsymbol{\gamma}_k$  parameters making them correspond to contrasts (i.e., deviations to the mean effect) with  $\sum_{k=1}^K \boldsymbol{\gamma}_k = 0$ . The total effect of  $\mathbf{X}_i^{\text{DIF}}(\mathbf{t})$  becomes the sum of its common effect on the latent process (part of  $\boldsymbol{\beta}$ ) and its item-specific contrast  $\boldsymbol{\gamma}_k$ .

With longitudinal data, two sorts of item-specific functionings may be investigated:

- 150 • classical DIF with  $\mathbf{X}_i^{\text{DIF}}(\mathbf{t})$  including time-independent covariates only; this explores how differently level parameters of a specific item differ according to individual characteristics;
- item response shift with  $\mathbf{X}_i^{\text{DIF}}(\mathbf{t})$  including functions of time  $t$ ; this explores how level parameters of a specific item change over time.

## 155 2.2. Maximum Likelihood Estimation

We consider here a maximum likelihood framework for the estimation of the dynamic IRT model.

### 2.2.1. Parameterization of the vector of parameters

Let denote  $\boldsymbol{\theta}$  the total vector of parameters defined in the structural part of the model described in (1) and in the  $K$  measurement equations described in 160 (2). This vector includes:

- the fixed effects  $\beta$  (except the intercept for identifiability).
- the parameters specifying the Variance-Covariance matrix  $\mathbf{B}$  of the random effects. To ensure that  $\mathbf{B}$  is positive definite, we consider the parameters of the Cholesky upper triangular transformation  $\mathbf{C}$  (i.e.,  $\mathbf{B} = \mathbf{C}^\top \mathbf{C}$ ) with first element fixed at 1 for identifiability.
- the discrimination parameters. We consider parameters  $(\sigma_k)_{k=1, \dots, K}$  so that the item discriminations  $\sigma_k^- = \frac{1}{|\sigma_k|}$ .
- the vector of item  $L_k - 1$  locations of each item  $k$ . To account for the constraint that  $\eta_{k1} \leq \eta_{k2} \leq \dots \leq \eta_{kL_k-1}$ , we consider the vector  $(\eta_{kl}^*)_{l=1, \dots, L_k-1}$  so that  $\eta_{k1} = \eta_{k1}^*$  and  $\eta_{kl} = \eta_{k1}^* + \sum_{m=2}^l (\eta_{km}^*)^2$  for  $l \geq 2$ .
- In case of differential effect of covariates on items (Equation (5)), the vector of parameters  $(\gamma_{\mathbf{k}})_{\mathbf{k}=1, \dots, K-1}$  with  $\gamma_{\mathbf{K}} = -\sum_{\mathbf{k}=1}^{K-1} \gamma_{\mathbf{k}}$ .

### 2.2.2. Contribution to the likelihood

Let denote  $\mathcal{Y}_i = \{Y_{ik}(t_{ikj}), \forall k \in \{1, \dots, K\}, \forall j \in \{1, \dots, n_{ik}\}\}$  all the repeated item information of subject  $i$ . The contribution of subject  $i$  to the likelihood is

$$\begin{aligned}
l_i(\boldsymbol{\theta}) &= \mathbb{P}(\mathcal{Y}_i | \boldsymbol{\theta}) \\
&= \int_{\mathbb{R}^p} \mathbb{P}(\mathcal{Y}_i | \boldsymbol{\theta}, \mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i \\
&= \int_{\mathbb{R}^p} \prod_{k=1}^K \prod_{j=1}^{n_{ik}} \prod_{l=0}^{L_k-1} \mathbb{P}(Y_{ik}(t_{ikj}) = l | \boldsymbol{\theta}, \mathbf{b}_i)^{\mathbb{1}_{Y_{ik}(t_{ikj})=l}} f(\mathbf{b}_i) d\mathbf{b}_i \\
&= \int_{\mathbb{R}^p} \prod_{k=1}^K \prod_{j=1}^{n_{ik}} \prod_{l=1}^{L_k} \left\{ \Phi(\sigma_k^-(\eta_{kl+1} - (\mathbf{X}_i(t_{ikj})^\top \boldsymbol{\beta} + \mathbf{Z}_i(t_{ikj})^\top \mathbf{b}_i))) \right. \\
&\quad \left. - \Phi(\sigma_k^-(\eta_{kl} - (\mathbf{X}_i(t_{ikj})^\top \boldsymbol{\beta} + \mathbf{Z}_i(t_{ikj})^\top \mathbf{b}_i))) \right\} f(\mathbf{b}_i) d\mathbf{b}_i
\end{aligned} \tag{6}$$

where the integration over the distribution of the p-vector of random effects is obtained by Quasi Monte-Carlo approximation following proposals of Philipson et al. [19]. We systematically considered 1000 points in this work.

180 The maximum likelihood estimators of  $\boldsymbol{\theta}$  are obtained by maximizing the log-likelihood  $\mathcal{L} = \sum_{i=1}^N \log(l_i(\boldsymbol{\theta}))$ . This is achieved with the Marquardt-Levenberg algorithm, a robust Newton-like algorithm, with stringent convergence criteria on the parameters, the log-likelihood and the first and second derivatives of the log-likelihood (see Philipps et al. [20] for details).

185 The maximum likelihood estimates are denoted  $\hat{\boldsymbol{\theta}}$  and their variance, obtained by the inverse of the Hessian, is denoted  $\widehat{\mathbf{V}}(\hat{\boldsymbol{\theta}})$ .

### 2.2.3. Software

The dynamic IRT model can be estimated with the `multlcm` function of R package `lcmm` [21]. A package vignette provides a tutorial that fully describes 190 the present dynamic IRT model estimation and posterior computations on a simulated dataset that mimics the PREDIALA data.

### 2.3. Posterior computations

We can compute several posterior quantities from the estimates  $\hat{\boldsymbol{\theta}}$ , and confidence intervals around these quantities can be obtained by approximating the 195 posterior distribution by Monte-Carlo simulations using the asymptotic distribution of the parameters  $\boldsymbol{\theta} \sim \mathcal{N}(\hat{\boldsymbol{\theta}}, \widehat{\mathbf{V}}(\hat{\boldsymbol{\theta}}))$ . We consider for this 2000 random draws. In the following, we omit  $\hat{\boldsymbol{\theta}}$  or more generally  $\boldsymbol{\theta}$  in the equations for better readability.

#### 2.3.1. Predicted trajectories of the construct

200 Predicted trajectories of the construct can be computed either at the population level (i.e., marginally to the random-effects) or at the individual level (i.e., conditionally to the individual random-effects). The predicted trajectory at the population level is computed for a profile of covariates  $\mathbf{x}(\mathbf{t})$ :

$$\mathbb{E}(\Lambda_i(t) | \mathbf{X}_i(\mathbf{t}) = \mathbf{x}(\mathbf{t})) = \mathbf{x}(\mathbf{t})^\top \boldsymbol{\beta}$$

The predicted trajectory at the individual level is computed given the individual covariates  $\mathbf{X}_i(t)$ ,  $\mathbf{Z}_i(t)$  and all the information on the items  $\mathcal{Y}_i = \{Y_{ik}(t_{ikj}), \forall k \in \{1, \dots, K\}, \forall j \in \{1, \dots, n_{ik}\}\}$ :

$$\mathbb{E}(\Lambda_i(t) | \mathbf{X}_i(t), \mathbf{Z}_i(t), \mathcal{Y}_i) = \mathbf{X}_i(t)^\top \boldsymbol{\beta} + \mathbf{Z}_i(t)^\top \mathbb{E}(\mathbf{b}_i | \mathcal{Y}_i)$$

where the expected random-effect  $\mathbb{E}(\mathbf{b}_i | \mathcal{Y}_i)$  is approximated by the mode of the posterior distribution  $f(\mathbf{b}_i | \mathcal{Y}_i)$ .

### 2.3.2. Item Characteristic Curves

With binary items, the Item Characteristic Curve (ICC) describes the probability of the highest item level according to the underlying construct level. With ordinal items, ICC translates into two curves:

- the item category probability curve also known as category characteristic curve which describes the probability of a response in a given item category according to the underlying construct level:

$$\begin{aligned} \text{CCC}_{k,l}(\Lambda) &= P(Y_{ik}(t) = l | \Lambda_i(t) = \Lambda) \\ &= \Phi(\sigma_k^- (\eta_{kl+1} - \Lambda)) - \Phi(\sigma_k^- (\eta_{kl} - \Lambda)) \end{aligned} \quad (7)$$

- The item score expectation curve which can also be computed as a function of the underlying construct level:

$$\begin{aligned} \mathbb{E}(Y_{ik} | \Lambda_i(t) = \Lambda) &= \sum_{l=0}^{L_k-1} l P(Y_{ik}(t) = l | \Lambda_i(t) = \Lambda) \\ &= L_k - 1 - \sum_{l=1}^{L_k-1} P(Y_{ik}(t) \leq l | \Lambda_i(t) = \Lambda) \\ &= L_k - 1 - \sum_{l=1}^{L_k-1} \Phi(\sigma_k^- (\eta_{kl} - \Lambda)) \end{aligned} \quad (8)$$

These two curves allow representing the items and their properties. The items locations describe where the items function along the construct level while

220 the steepness of the item score expectation characterizes the items discrimina-  
 tions. For example, the steeper the curve for an item, the better it can discrim-  
 inate between two different construct levels.

### 2.3.3. Predicted trajectories of the items

The predicted trajectory over time of each item can be computed for a profile  
 225 of covariates  $\mathbf{x}(\mathbf{t})$  as follows:

$$\mathbb{E}(Y_{ik}(t) | \mathbf{X}_i(\mathbf{t}) = \mathbf{x}(\mathbf{t})) = \int_{\mathbb{R}^p} \mathbb{E}(Y_{ik}(t) | \Lambda_i(t) = \mathbf{x}(\mathbf{t})^\top \boldsymbol{\beta} + \mathbf{z}(\mathbf{t})^\top \mathbf{b}) f(\mathbf{b}) d\mathbf{b} \quad (9)$$

where  $\mathbb{E}(Y_{ik} | \Lambda_i(t) = \mathbf{x}(\mathbf{t})^\top \boldsymbol{\beta} + \mathbf{z}(\mathbf{t})^\top \mathbf{b})$  is computed as in equation (8) with  
 $\Lambda = \mathbf{x}(\mathbf{t})^\top \boldsymbol{\beta} + \mathbf{z}(\mathbf{t})^\top \mathbf{b}$ , and the integral over the random-effects  $b$  distribution  
 is obtained by Quasi Monte-Carlo approximation.

### 2.3.4. Fisher Information

230 The Fisher information provides a quantification of the level of information  
 brought by each item, and each item level. It is computed using the second  
 derivatives of the item level probability denoted  $P_{kl}(\Lambda) = P(Y_{ik}(t) = l | \Lambda_i(t) = \Lambda)$   
 for item  $k$  and level  $l$ . The item information function for category  $l$  is defined  
 as follows (calculations are detailed in Section 1 of the supplementary material):

$$\begin{aligned} I_{kl}(\Lambda) P_{ikl}(\Lambda) &= -\frac{\partial^2 \log(P_{ikl}(\Lambda))}{\partial \Lambda^2} P_{ikl}(\Lambda) \\ &= -\frac{1}{\sigma_k^2} \frac{\left( \phi\left(\frac{\Lambda - \eta_{kl+1}}{\sigma_k}\right) - \phi\left(\frac{\Lambda - \eta_{kl}}{\sigma_k}\right) \right)^2}{\Phi\left(\frac{\Lambda - \eta_{kl+1}}{\sigma_k}\right) - \Phi\left(\frac{\Lambda - \eta_{kl}}{\sigma_k}\right)} \\ &\quad - \left( \frac{\Lambda - \eta_{kl+1}}{\sigma_k} \phi\left(\frac{\Lambda - \eta_{kl+1}}{\sigma_k}\right) - \frac{\Lambda - \eta_{kl}}{\sigma_k} \phi\left(\frac{\Lambda - \eta_{kl}}{\sigma_k}\right) \right) \end{aligned} \quad (10)$$

The information curve provides the summary at the item level as follows:

$$\begin{aligned}
 I_k(\Lambda) &= \sum_{l=1}^{L_k-1} I_{kl}(\Lambda) P_{kl}(\Lambda) \\
 &= \sum_{l=1}^{L_k-1} -\frac{1}{\sigma_k^2} \frac{\left( \phi\left(\frac{\Lambda - \eta_{kl+1}}{\sigma_k}\right) - \phi\left(\frac{\Lambda - \eta_{kl}}{\sigma_k}\right) \right)^2}{\Phi\left(\frac{\Lambda - \eta_{kl+1}}{\sigma_k}\right) - \Phi\left(\frac{\Lambda - \eta_{kl}}{\sigma_k}\right)} \quad (11)
 \end{aligned}$$

### 3. Application to PREDIALA study

We applied the dynamic IRT model to analyze the repeated measures of depressive symptomatology in the PREDIALA study. The HADS (Hospital Anxiety and Depression scale) was used to measure anxiety and depression disorders [22]. The HADS consists of 14 items on a 4-point Likert scale, seven 240 of which are related to anxiety symptoms and seven to depression symptoms. Only the depression symptom domain is presented here. The 7 items rated from 0 (total agreement) to 3 (total disagreement) are as follows:

- Item 2 “I still enjoy the things I used to enjoy” (Enjoy)
- 245 • Item 4 “I can laugh and see the funny side of things” (Laugh)
- Item 6 “I feel cheerful” (Cheerful)
- Item 8 “I feel I am slowed down” (Slow)
- Item 10 “I have lost interest in my appearance” (Appearance)
- Item 12 “I look forward with enjoyment to things” (Looking forward)
- 250 • Item 14 “I can enjoy a good book or radio or TV programme” (Leisure)

Responses to items 8 and 10 are reversed so that higher levels systematically indicate more intense symptoms.

Our objective was to describe the trajectory of depressive symptomatology over time from registration on the waiting list for a renal transplant, and to describe the possible differences according to patients' renal replacement therapy at inclusion in the study, i.e. patients either dialyzed or not (preemptive). Indeed, being on the transplant waiting list may be experienced differently between dialyzed and preemptive patients. It can be hypothesized that, for example, the depressive symptoms experienced by dialyzed patients are more pronounced as compared to preemptive patients due to their experience with dialysis and expectations of associated complications. It is therefore possible that the need for clinical and psychological support is not the same for all patients.

A secondary objective was to assess whether the functioning of some items differed according to the group (preemptive or dialyzed) or shifted with time on the waiting list. Indeed, patients may perceive the items differently according to their renal replacement therapy and over time, despite having similar depression levels.

### *3.1. PREDIALA sample*

We included in the analysis all the patients from PREDIALA who entered the study within 48 months following their registration on the waiting list. They were either in the dialyzed or preemptive group at entry, and had at least one measure for each of the 7 items of the HADS before the end of the study. The end of the study was defined by either a switch in group (from preemptive to dialyzed status), a clinical event (mainly transplantation) or the administrative censoring. From the initial 577 patients included in PREDIALA study, this selection lead to a final sample of 561 patients and 1136 repeated visits. Among them 356 (63.5% ) were men and 288 (51.3%) were under dialysis. The median age at entry was 59 years (range 19-67 years), and the patients had been on the waiting list for a very variable time ranging from 0.1 to 43.1 months (median



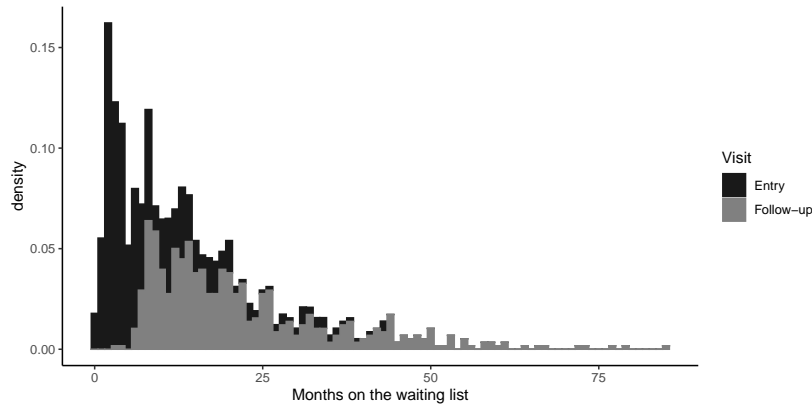


Figure 1: Distribution of the measurement times in the PREDIALA study according to the time since registration on the waiting list (at entry in black and during follow-up in grey)

280 5.1 months) at entry in the cohort. This leads to substantial variability in measurement timings across patients at entry and during follow-up as shown in Figure 1. This continuous distribution of the measurement times which would have been ignored using standard IRT methods is naturally handled in the dynamic IRT model thanks to the definition of the underlying construct as a  
 285 latent process in continuous time.

### 3.2. The dynamic IRT model specification

The dynamic model was defined following equation (1) for the trajectory of underlying depressive symptomatology and equation (2) for the 7 item-specific measurement model. As we did not have any assumption regarding the shape  
 290 of trajectory over time of depressive symptomatology, we used a basis of natural cubic splines with 2 internal knots placed at tertiles of the measurement time distribution, that is 7 and 15 months, and boundary knots placed at 0 and 60 months. Each of the four functions of time (intercept and 3 splines functions) was associated with fixed effects specific to the group (Preemptive/Dialyzed)  
 295 to assess the mean trajectories, and individual correlated random effects to account for the correlation within repeated measures of each individual. For the

measurement models, we assumed in the main analysis that all items functioned similarly, i.e., no DIF and no response shift occurred. We then explored in secondary analyses whether some items functioned differently by group (adding an item-specific contrast on group), and whether some items were affected by response shift over time (adding an item-specific contrast of the 3 time functions). The parameter estimates of these three models are provided in Table S1 of the supplementary material.

### 3.3. Predicted trajectories of depressive symptomatology

In the following description, 1 unit of depressive symptomatology, called 1 SD, corresponds to the inter-individual variability at registration in the dialyzed group.

The mean predicted trajectory over time of depressive symptomatology, displayed in Figure 2, varied over time and according to the group. In the preemptive group, the level of depressive symptomatology increased during the first year on the waiting list by 0.243 (-0.012,0.498) SD and then remained stable. The level of depressive symptomatology was higher in the dialyzed group compared to the preemptive group at the time of registration (difference of -0.482 (-0.814,-0.149) SD). It then slightly decreased during the first year to reach a similar level as in the preemptive group, and then increased again after approximately 2 years on the waiting list by a mean annual rate of 0.245 (0.053,0.438) SD (computed from 2 to 6 years).

### 3.4. Measurement scale structure

We exploited the dynamic IRT model to assess the HADS depressive symptomatology items characteristics. To help appreciating the item characteristics, we assessed the range of the distribution of the underlying depressive symptomatology based on the estimates of the dynamic IRT model and an hypothetical

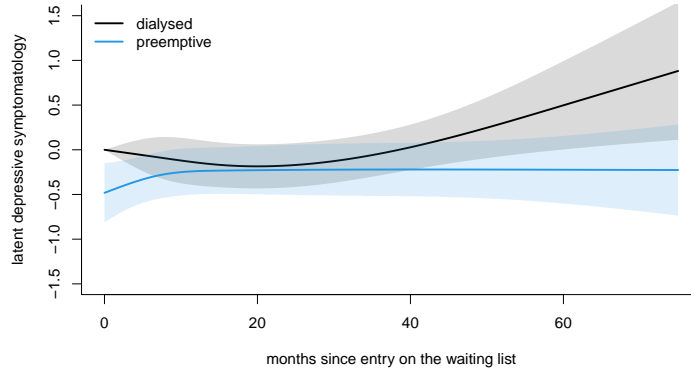


Figure 2: Mean trajectory (and 95% confidence interval in shades) of depressive symptomatology estimated by the dynamic IRT model from the 7 items of HADS repeatedly measured over time; represented for dialyzed and preemptive patients

population of 100000 preemptive patients and 100000 dialyzed patients with  
measures every month from registration up to 72 months. The resulting 95%  
325 prediction interval of the underlying depressive symptomatology was  $[-6.10, 5.90]$   
with the 10% and 90% percentiles of the distribution at  $-3.00$  and  $2.74$ , respec-  
tively.

Table 1 provides the estimated locations and discrimination while Figure  
3 shows the curves of item expectations (top) and curves of item information  
330 (bottom) according to the underlying depressive symptomatology. Figure S1  
and S2 of the supplementary material further display for each item category the  
probability curve and the information function, respectively.

The easiest items, in terms of their difficulty parameters, were Item 8 (Slow)  
and Item 2 (Enjoy) while the most difficult one was Item 14 (Leisure). This  
335 means that the level of depression required to respond to the most unfavorable  
response categories (i.e. indicative of higher depressive symptoms) of items 2 and  
8 (e.g., “Sometimes” for item 8 “I feel as if I am slowed down”) was lower than

the level of depression required to respond to the most unfavorable response categories of item 14 (e.g., “Not often” for item 14 “I can enjoy a good book or radio or TV program”). The most discriminant items with the steepest curves, representing their ability to discriminate patients with different levels of depression, were items 2, 4 and 12 concerning the ability to enjoy, laugh and look forward, respectively. The estimated curve of the Fisher information plotted in Figure 3 (bottom) also underlines the major role of these 3 items compared to the others. In contrast, item 14 about leisure does not bring much information in this population as it seems to measure much higher levels of depression than the other items.

Item	category 0 - 1		category 1 - 2		category 2 - 3		discrimination	
	est	SE	est	SE	est	SE	est	SE
2 - Enjoy	-0.46	0.13	0.77	0.14	1.52	0.18	1.29	0.13
4 - Laugh	-0.26	0.12	0.74	0.13	1.91	0.21	1.56	0.16
6 - Cheerful	-0.48	0.14	1.58	0.20	3.34	0.38	0.85	0.09
8 - Slow	-1.51	0.19	0.40	0.13	1.69	0.20	0.95	0.10
10 - Appearance	-0.05	0.12	1.00	0.16	2.27	0.26	0.88	0.10
12 - Looking Forward	-0.32	0.13	0.72	0.13	1.82	0.20	1.46	0.15
14 - Leisure	0.83	0.17	3.18	0.42	4.11	0.54	0.56	0.07

Table 1: Estimate (and associated standard error (SE) obtained by  $\Delta$ -method) of locations and discrimination of the 7 items of HADS measuring Depressive symptomatology

### 3.5. Differential Item Functioning and Response Shift

We first explored any differential item functioning on the group (dialyzed versus preemptive). Estimates are provided in Table 2 along with those of the main model that ignores DIF. Overall, the Chi-square test assessing simultaneously the 6 contrasts on group did not reject the null “no DIF on group” assumption ( $p=0.266$ ). However, taken individually, the difference between groups for item 2 (Enjoy) was significantly larger than for the other items (item-specific effect of preemptive group on the underlying level estimated at -0.150 (-0.269,-0.0314) SD). This suggests that this item is more difficult for preemptive patients (lo-

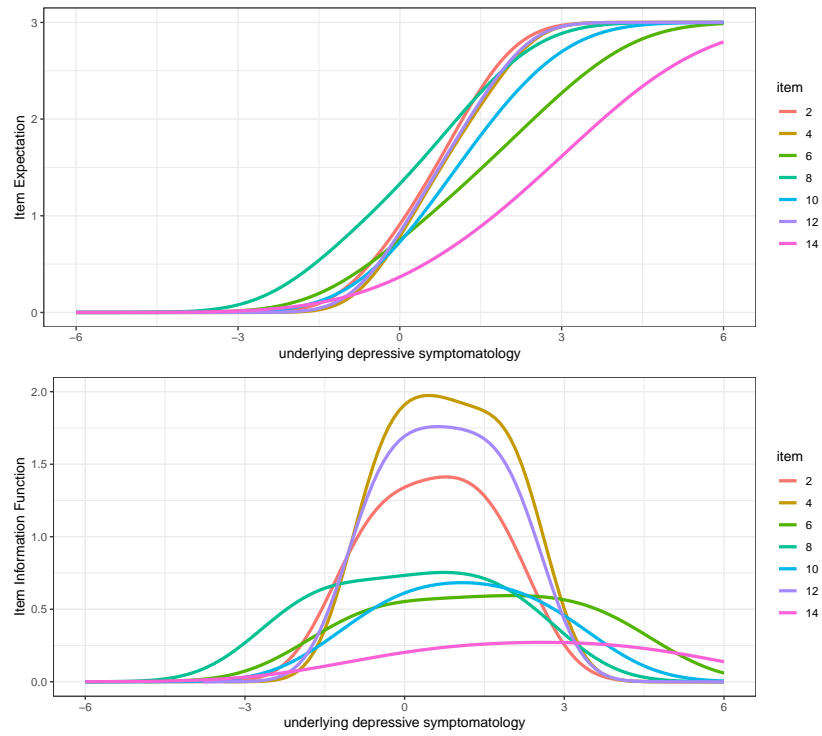


Figure 3: Estimated expectation of each item (top) and estimated Fisher information of each item (bottom) according to the underlying depressive symptomatology. Items are rated so that higher levels systematically indicate more intense depressive symptoms

cation parameter shifted to +0.150) at the same underlying level of depressive symptomatology; preemptive patients tend to respond more readily to more favorable response categories than patients under dialysis, despite having similar depression levels. In addition, accounting or not for DIF impacted the conclusions: the group effect on the underlying depressive symptomatology was not significant anymore when accounting for DIF suggesting that the difference between groups in the model without DIF was mainly carried by item 2.

	coef	SE	p	coef	SE	p
intercept	0.000	-	-	0.000	-	-
ns1	-0.305	0.172	0.075	-0.304	0.229	0.184
ns2	0.039	0.222	0.862	0.038	0.590	0.949
ns3	0.538	0.250	0.031	0.538	0.310	0.083
group	-0.482	0.170	0.004	-0.458	0.413	0.267
ns1:preemptive	0.477	0.223	0.032	0.473	0.265	0.075
ns2:preemptive	0.428	0.335	0.201	0.429	0.491	0.382
ns3:preemptive	-0.390	0.292	0.182	-0.391	0.338	0.247
<i>Contrasts on preemptive:</i>				<i>(global p=0.266)</i>		
	Item 2			-0.150	0.061	0.013
	Item 4			-0.031	0.059	0.600
	Item 6			0.020	0.077	0.800
	Item 8			-0.053	0.100	0.593
	Item 10			0.040	0.158	0.802
	Item 12			0.037	0.061	0.543
	** Item 14			0.138	0.124	0.266

\*\* coefficient not estimated but obtained as minus the sum of the others.

Table 2: Estimated fixed parameters in the dynamic IRT without (left) and with (right) differential item functioning on group. ns1,ns2,ns3 refer to the natural cubic splines functions.

We secondly explored any item responses shift over time by adding item-specific contrasts on the 3 natural cubic splines functions of time (ns1, ns2, ns3). The resulting predicted trajectories of each item in this model compared to the model assuming no response shift are given in Figure 4. No clear response shift was identified here although the overall test assessing simultaneously ns1, ns2 and ns3 contrasts in the model with RS suggested some potential for a different behavior of item 2 over time compared to the others (p=0.073). Each

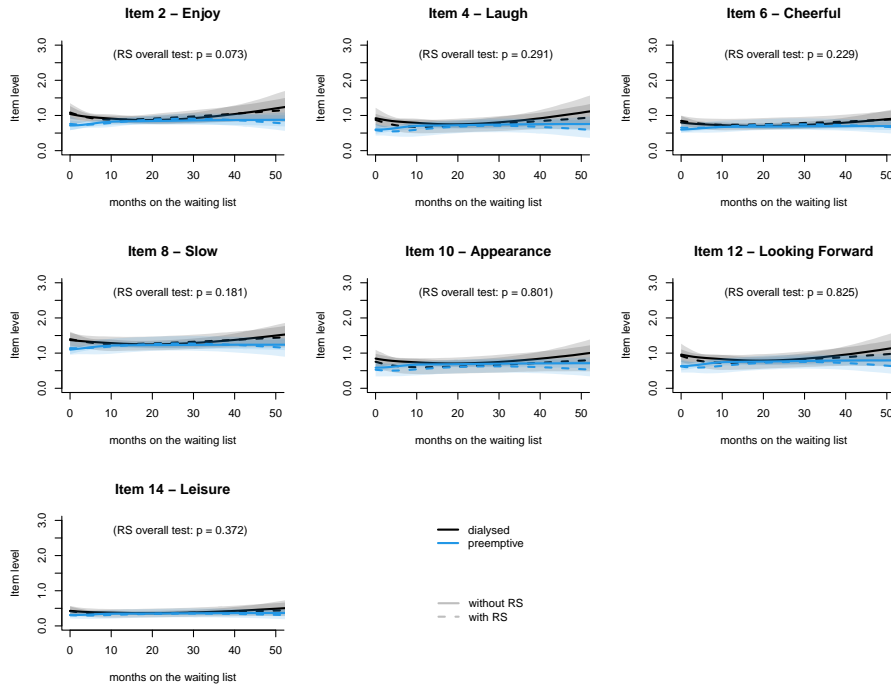


Figure 4: Predicted items trajectories according to time on the waiting list and group in the model neglecting potential response shift (plain) and in the one accounting for potential response shift (dashed) - with p-value of the overall  $\chi^2$  test for RS over the 3 splines functions in the model assuming RS.

item behavior over time was close to the one of the underlying construct showed in Figure 2 with some slight differences in the intensity of change in the first months after entry on the waiting list. Note that these slight differences do not reflect directly the statistical test which only focused on the model assuming

375 RS.

#### 4. Conclusions

We have described how to combine the item response theory with the linear mixed model theory for item-level analysis of longitudinal PRO or CRO data when measurement times may vary across individuals. Using a real case example

380 with the PREDIALA study, we have shown that this dynamic IRT model can

describe the latent construct trajectory over time and its determinants, while simultaneously assessing the item and scale properties, and exploring lack of measurement invariance between groups (DIF) and over time (response shift).

This analysis of PREDIALA data helped to better understand the experience  
385 of patients with end-stage renal disease on the renal transplant waiting list in terms of depressive symptoms. DIF was highlighted on item 2 (Enjoy) indicating that patients under dialysis had more difficulty in reporting having enjoyment than preemptive patients, despite having similar levels of depression. This may reveal that the need for clinical and psychological support may not be the same  
390 for all patients, according to their renal replacement therapy. Response shift was not significantly evidenced despite a trend on this same item 2. Adjusting for DIF and response shift, the level of depressive symptomatology of preemptive patients tended to slightly increase during the first year and to remain stable afterwards. The level of depressive symptomatology of patients under dialysis  
395 tended to be close to the one of preemptive patients after 2 years on the waiting list and to increase afterwards. Although it has been reported that the time on waiting list should be reduced to limit depressive symptoms [23] and improve health-related quality of life [24], the shortage of grafts unfortunately often makes this difficult to achieve.

400 The dynamic IRT model we described here unites the strengths of IRT and LMM theories. On the one hand, the use of a structural mixed model makes it possible to operationalize the latent construct as a latent process defined in continuous time and thus takes into account that most health phenomena intrinsically evolve in continuous time. On the other hand, the use of IRT  
405 methodology to define the measurement scale at each individual- and occasion-specific observation time enables a precise modeling of the items constituting the measurement scale, and their properties.



However, as for all methodologies, this method is not without some limitations and some further path of research may be put forward. First, we focused  
410 here on a specific measurement model, the GRM, which translates the discretization of the underlying latent process into ordinal categories as shown in Equation (3) [25]. However, coming from IRT models, this measurement model does not possess the specific objectivity property as Rasch Measurement Theory (RMT) models do [26]. It would thus be interesting to adapt the proposed methodology  
415 to RMT measurement models (see Blanchin et al. [REF] in this special issue and Barbieri et al. [18]). Of note, changing the measurement model does not impact the estimation procedure nor the structural part of the model, and the current version of the program already handles different measurement models for continuous items in addition to GRM for ordinal items.

420 Second, by relying on the mixed model theory and the maximum likelihood estimation, the dynamic IRT model relies on the missing at random assumption both for monotonic and for intermittent missingness [27]. In the presence of informative dropout, a joint model for the repeated item responses and the time to dropout, as described by Saulnier et al. in this special issue (REF),  
425 should be favored. This joint model combines the dynamic IRT model with a survival model that captures the association between the underlying latent construct and the dropout (or any other event of interest). Third, the methodology is fully parametric, so that analytic choices are systematically made. For instance, to simplify the application setting, we only included time-independent  
430 covariates even though time-dependent covariates could also be considered, such as a changing of group during the follow-up. We also globally tested the lack of measurement invariance over the items' parameters across the overall follow-up although a more precise assessment could be done regarding which function of time is affected or at which time the lack of measurement invariance occurs.

435 To conclude, by extending the IRT methodology to longitudinal data, and considering the time as continuous, our methodology provides a versatile and flexible approach for modeling item responses measured repeatedly over time as encountered in numerous longitudinal health studies.

**Acknowledgements** The authors gratefully thank the co-investigators of  
440 the study: Magali Giral, Aurélie Meurette, Emmanuel Morelon, and Laetitia  
Albano. The authors express sincere thanks to Elodie Faurel-Paul and  
Astrid Fleury for their assistance for the study, as well as all the partici-  
pants for their contribution to the study. The authors wish to thank mem-  
bers of the clinical research assistant team and DIVAT Consortium Collab-  
445 orators (Medical Doctors, Surgeons, HLA Biologists). Nantes: Gilles Blan-  
cho, Julien Branchereau, Diego Cantarovich, Agnès Chapelet, Jacques Dan-  
tal, Clément Deltombe, Lucile Figueres, Claire Garandeanu, Magali Giral, Caro-  
line Gourraud-Vercel, Maryvonne Hourmant, Georges Karam, Clarisse Kerleau,  
Aurélie Meurette, Simon Ville, Christine Kandell, Anne Moreau, Karine Re-  
450 naudin, Anne Cesbron, Florent Delbos, Alexandre Walencik, Anne Devis; Lyon  
E. Hériot : Lionel Badet, Maria Brunet, Fanny Buron, Rémi Cahen, Sameh  
Daoud, Coralie Fournie, Arnaud Grégoire, Alice Koenig, Charlène Lévi, Em-  
manuel Morelon, Claire Pouteil-Noble, Thomas Rimmelé, Olivier Thaunat.

**Funding** This work was funded by the French National Research Agency  
455 (Project DyMES - ANR-18-C36-0004-01) and the French Ministry of Health  
(PHRC-13-0224, 2013).

**Ethical approval** The PREDIALA study is part of the PreKit-QoL study  
which is registered on the ClinicalTrials.gov Registry (RC14.0078, NCT02154815).  
It has obtained approval from the ethical Committee for Persons' Protection  
460 (CPP, Tours, 2014-S8), and from the advisory committee on research data and  
information in health (CCTIRS, Paris, 14.314).

**Conflict of interest** All authors have no competing interest.

**Supplementary Material** Supplementary material can be found on the  
journal website.

465 **Data availability** The software is openly available in the R package lcmm

(on <https://github.com/CecileProust-Lima/lcmm> and on cran). The R script along with a dataset mimicking the PREDIALA study are also provided as a package vignette. The raw PREDIALA data can not be shared.

## References

- 470 [1] M. Cerou, S. Peigné, E. Comets, M. Chenel, Application of item response theory to model disease progression and agomelatine effect in patients with major depressive disorder, *The AAPS journal* 22 (1) (2019) 4, PMID: 31720897. doi:10.1208/s12248-019-0379-x.
- [2] R. Gortler, J.-P. Fox, J. W. R. Twisk, Why item response theory should  
475 be used for longitudinal questionnaire data analysis in medical research, *BMC Medical Research Methodology* 15 (1) (2015) 55. doi:10.1186/s12874-015-0050-x.  
URL <https://doi.org/10.1186/s12874-015-0050-x>
- [3] W. McCall, B. Porter, A. Pate, C. Bolstad, C. Drapeau, A. Krystal,  
480 R. Benca, M. Rumble, M. Nadorff, Examining suicide assessment measures for research use: Using item response theory to optimize psychometric assessment for research on suicidal ideation in major depressive disorder, *Suicide & Life-Threatening Behavior* PMID: 34237156. doi: 10.1111/sltb.12791.
- 485 [4] G. Abdelhamid, M. Bassiouni, J. Gómez-Benito, Assessing cognitive abilities using the wais-iv: An item response theory approach, *International Journal of Environmental Research and Public Health* 18 (13) (2021) 6835, PMID: 34202249 PMCID: PMC8297006. doi:10.3390/ijerph18136835.
- [5] S. E. Rakers, M. E. Timmerman, M. E. Scheenen, M. E. de Koning, H. J.  
490 van der Horn, J. van der Naalt, J. M. Spikman, Trajectories of Fatigue, Psy-

chological Distress, and Coping Styles After Mild Traumatic Brain Injury: A 6-Month Prospective Cohort Study, *Archives of Physical Medicine and Rehabilitation* (2021) S0003-9993(21)00462-7doi:10.1016/j.apmr.2021.06.004.

- 495 [6] I. Otto, C. Hilger, A. Magheli, G. Stadler, F. Kendel, Illness representations, coping and anxiety among men with localized prostate cancer over an 18-months period: A parallel vs. level-contrast mediation approach, *Psycho-Oncology*doi:10.1002/pon.5798.
- [7] A. Edjolo, C. Proust-Lima, F. Delva, J.-F. Dartigues, K. Pérès, Natu-  
500 ral History of Dependency in the Elderly: A 24-Year Population-Based Study Using a Longitudinal Item Response Theory Model, *American Journal of Epidemiology* 183 (4) (2016) 277-285, number: 4. doi:10.1093/aje/kwv223.
- [8] L. Cai, C. R. Houts, Longitudinal Analysis of Patient-Reported Outcomes  
505 in Clinical Trials: Applications of Multilevel and Multidimensional Item Response Theory, *Psychometrika* 86 (3) (2021) 754-777. doi:10.1007/s11336-021-09777-y.
- [9] N. M. Laird, J. H. Ware, Random-effects models for longitudinal data, *Biometrics* 38 (4) (1982) 963-74.
- 510 [10] F. Samejima, Graded Response Model, in: W. J. van der Linden, R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory*, Springer, New York, NY, 1997, pp. 85-100. doi:10.1007/978-1-4757-2691-6\_5. URL [https://doi.org/10.1007/978-1-4757-2691-6\\_5](https://doi.org/10.1007/978-1-4757-2691-6_5)
- [11] V. Sébille, J.-B. Hardouin, M. Giral, A. Bonnaud-Antignac, P. Tessier,  
515 E. Papuchon, A. Jobert, E. Faurel-Paul, S. Gentile, E. Cassuto, E. Morélon,

- L. Rostaing, D. Glotz, R. Sberro-Soussan, Y. Foucher, A. Meurette, Prospective, multicenter, controlled study of quality of life, psychological adjustment process and medical outcomes of patients receiving a preemptive kidney transplant compared to a similar population of recipients after a dialysis period of less than three years–The PreKit-QoL study protocol, BMC nephrology 17 (2016) 11. doi:10.1186/s12882-016-0225-7.
- [12] L. Auneau-Enjalbert, J.-B. Hardouin, M. Blanchin, M. Giral, E. Morelon, E. Cassuto, A. Meurette, V. Sébille, Comparison of longitudinal quality of life outcomes in preemptive and dialyzed patients on waiting list for kidney transplantation, Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation 29 (4) (2020) 959–970. doi:10.1007/s11136-019-02372-w.
- [13] A. Tong, C. S. Hanson, J. R. Chapman, F. Halleck, K. Budde, M. A. Josephson, J. C. Craig, 'Suspended in a paradox'-patient attitudes to wait-listing for kidney transplantation: systematic review and thematic synthesis of qualitative studies, Transplant International: Official Journal of the European Society for Organ Transplantation 28 (7) (2015) 771–787. doi:10.1111/tri.12575.
- [14] P. W. Holland, H. Wainer (Eds.), Differential Item Functioning, Routledge, New-York, NY, 2009.
- [15] M. A. Sprangers, C. E. Schwartz, Integrating response shift into health-related quality of life research: a theoretical model, Social Science & Medicine (1982) 48 (11) (1999) 1507–1515. doi:10.1016/s0277-9536(99)00045-3.
- [16] C. Proust-Lima, H. Amieva, H. Jacqmin-Gadda, Analysis of multivariate mixed longitudinal data: a flexible latent process approach, The British

journal of mathematical and statistical psychology 66 (3) (2013) 470–487.  
doi:10.1111/bmsp.12000.

[17] F. B. Baker, S. H. Kim, Item Response Theory. Parameter Estimation Tech-  
545 niques, 2nd Edition, Statistics: Textbooks & Monographs, Marcel Dekker,  
New York, 2004.

[18] A. Barbieri, J. Peyhardi, T. Conroy, S. Gourgou, C. Lavergne, C. Mollevi,  
Item response models for the longitudinal analysis of health-related quality  
of life in cancer clinical trials, BMC Medical Research Methodology 17 (1)  
550 (2017) 148. doi:10.1186/s12874-017-0410-9.  
URL <https://doi.org/10.1186/s12874-017-0410-9>

[19] P. Philipson, G. L. Hickey, M. J. Crowther, R. Kolamunnage-Dona, Faster  
Monte Carlo estimation of joint models for time-to-event and multivariate  
longitudinal data, Computational Statistics & Data Analysis 151 (2020)  
555 107010. doi:10.1016/j.csda.2020.107010.  
URL <https://www.sciencedirect.com/science/article/pii/S0167947320301018>

[20] V. Philipps, B. P. Hejblum, M. Prague, D. Commenges, C. Proust-Lima,  
Robust and efficient optimization using a Marquardt-Levenberg algorithm  
560 with R package marqLevAlg, R journal in press.

[21] C. Proust-Lima, V. Philipps, B. Liqueur, Estimation of Extended Mixed  
Models Using Latent Classes and Latent Processes: The R Package lmm,  
Journal of Statistical Software, Articles 78 (2) (2017) 1–56. doi:10.18637/  
jss.v078.i02.  
565 URL <https://www.jstatsoft.org/v078/i02>

[22] A. S. Zigmond, R. P. Snaith, The hospital anxiety and depression scale,

Acta Psychiatrica Scandinavica 67 (6) (1983) 361–370. doi:10.1111/j.1600-0447.1983.tb09716.x.

[23] E. Corruble, A. Durrbach, B. Charpentier, P. Lang, S. Amidi, A. Deza-  
570 mis, C. Barry, B. Falissard, Progressive Increase of Anxiety and  
Depression in Patients Waiting for a Kidney Transplantation, Be-  
havioral Medicine 36 (1) (2010) 32–36, publisher: Taylor & Fran-  
cis \_eprint: <https://doi.org/10.1080/08964280903521339>. doi:10.1080/  
08964280903521339.

575 URL <https://doi.org/10.1080/08964280903521339>

[24] S. C. Ong, W. L. Chow, S. van der Erf, V. D. Joshi, J. F. Lim, C. Lim,  
P. S. Tee, Y. M. Lu, T. Y. Kee, What factors really matter? Health-related  
quality of life for patients on kidney transplant waiting list, Annals of the  
Academy of Medicine, Singapore 42 (12) (2013) 657–666.

580 [25] D. Commenges, H. Jacqmin-Gadda, A. Alioum, P. Joly, B. Liqueur,  
C. Proust-Lima, V. Rondeau, R. Thiébaud, Chapter 5. Extensions of mixed  
models, in: Dynamical Biostatistical Models, CRC Press, 2015.

[26] D. Andrich, Rating scales and rasch measurement, Expert Review of  
Pharmacoeconomics & Outcomes Research 11 (5) (2011) 571–585, pMID:  
585 21958102. doi:10.1586/erp.11.59.

[27] R. J. A. Little, Modeling the Drop-Out Mechanism in Repeated-Measures  
Studies, Journal of the American Statistical Association 90 (431) (1995)  
1112–1121. doi:10.1080/01621459.1995.10476615.

590 URL [http://amstat.tandfonline.com/doi/abs/10.1080/01621459.](http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1995.10476615)  
1995.10476615