Journal Title
XX(X):1–30
©The Author(s) 2017
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Missing item responses in latent growth analysis; IRT versus CTT

R. Gorter[1], J.-P. Fox[2], I. Eekhout[3,4], M.W. Heymans[4], J.W.R. Twisk[4]

Abstract

In medical research, repeated questionnaire data is often used to measure and model latent variables across time. Through a novel imputation method a direct comparison is made between latent growth analysis under Classical Test Theory (CTT) and Item Response Theory (IRT), while also including effects of missing item responses. For CTT and IRT, by means of a simulation study the effects of item missingness on latent growth parameter estimates are examined given longitudinal item response data. Several missing data mechanisms and conditions are evaluated in the simulation study. The additional effects of missingness on differences in CTT- and IRT-based latent growth analysis are directly assessed by rescaling the multiple imputations.The multiple imputation method is used to generate latent variable and item scores from the posterior predictive distributions to account for missing item responses in observed multilevel binary response data. It is shown that a multivariate probit model, as a novel imputation model, improves the latent growth analysis, when dealing with missing at random (MAR) in CTT. The study also shows that the parameter estimates for the latent growth model using IRT show less bias and have smaller MSE's compared to the estimates using CTT.

Keywords

missing data, longitudinal data, multilevel IRT, questionnaires, CTT, multiple imputation

[1] Military Mental Health Research Centre, Ministery of Defence, Utrecht, The Netherlands.
[2] Department of Research Methodology, Measurement, and Data Analysis, Faculty of Behavioural, Management & Social Sciences, University of Twente, Enschede, The Netherlands.
[3] TNO Child Health, Netherlands Organization for Applied Scientific Research, Leiden, The Netherlands.
[4] Department of Epidemiology and Biostatistics, Amsterdam Public Health research institute, Amsterdam Medical Centre, Amsterdam, The Netherlands.

1 Introduction

Multi-item questionnaires are often used for measuring latent variables, such as depression or quality of life. Item-level data is collected using questionnaires, scale scores are computed, and they are considered measurements of a construct. For example, the Beck Depression Inventory (BDI-II) is used to measure symptoms of depression [1], the NEO Personality Inventory (NEO-PI-3) to measure the Big Five personality traits [2], and the Brief Pain Inventory (BPI) to measure pain severity and interference [3]. The computation of construct scores requires a measurement model to relate the item responses to the underlying latent trait. Item Response Theory (IRT) or Classical Test Theory (CTT) is used to compute latent variable scores given observed item-level data [4].

When analyzing the development over time of the latent variable for instance by applying a latent growth model, the measurement model, i.e. either IRT or CTT, can highly influence the results [5]. For a longitudinal study on complete data, parameter estimates of a repeated measurements model were directly compared using multiple imputations for the latent variable from a posterior predictive distribution under an IRT and a CTT model. They showed that the posterior predictive distribution for the latent variable scores, constructed under an IRT model performed much better in retrieving the true parameter values than the predictive distribution constructed under a CTT model. It was shown that IRT utilizes all response pattern information, which leads to more heterogeneous latent variable scores than the sum scores constructed under CTT, reducing the bias in parameter estimates [5].

In this latent growth modeling comparison, the presence of missing item-level data is ignored. However, in practice missing data on item level can occur when participants refuse to answer sensitive items, inadvertently skip items, or skip items that do not apply to them. Item-level missing data can bias test results and requires careful handling to make correct statistical inferences [6]. The common approach to deal with this problem, averaging the available items to compute a scale score, results in bias [7, 8, 9]. The handling of item-level missing data is necessary to avoid a loss in power due to missing item scores and to utilize all observed item-level data in the analysis.

In the IRT-based analysis, multiple imputation to handle missing data is operationalized by the posterior predictive distribution of the item responses and by the posterior predictive distribution of the longitudinal latent variable. The IRT model is defined at the level of item scores, which facilitates the construction of a posterior predictive distribution for the missing item responses. As a result of the repeated measurements, the data has a nested structure; i.e. the measurements on multiple occasions are nested within participants. This multilevel information is included in the generation of the latent variable scores by including the latent growth model for the latent variable in the construction of the posterior predictive distribution. This is in line with the multiple imputation method for binary multilevel data of Quartagno [10] and Audigier [11]. The multilevel approach can also be used to deal with an unbalanced design, where participants have been measured on different occasions and the number of measurements per participant is different. For instance, when missing data appears if the participant skips a measurement occasion, there is often sufficient information to estimate the average effects [12].

Missing item responses cause a problem in CTT modeling, since sum scores computed from the available item responses result in bias [7, 8]. Furthermore, the CTT model is defined at an aggregate data level and not at the level of item responses, which makes the CTT model not useful as an imputation model for item missings. Without a sum score, it is also not possible to define a random effect in order to define a multilevel imputation model for the missing item scores. A new CTT-based imputation model (i.e, a multivariate probit model) is proposed for binary item responses that takes into account latent growth of the latent variable and the nested structure of responses within participants. This multivariate probit model represents a marginal CTT model, where the true score is integrated out. From this model, a posterior predictive distribution for the responses can be defined, since it is defined at the level of observations. The marginal CTT model also preserves the multilevel structure of the data (i.e., responses nested within subjects), which is represented by the covariance structure, where the mean term can include the growth modeling part for the latent variable score. Therefore, the corresponding posterior predictive distribution for the item scores takes into account the uncertainty that is associated with the missingness and the individual trajectory of the latent variable over time to prevent inconsistency of the trend due to missing values.

The multivariate probit model as an imputation model differs from recently developed approaches based on the generalized linear mixed effects model [10, 11, 13, 14], since dependencies between item responses, nested in the same response pattern, are directly modeled in the probit model and not through a random effects structure. The imputation model differs from the analysis model that is used for calculating the effects. Note that the local stochastic independence still holds for the imputation model. When a response pattern contains only a few item responses, the random effects cannot be accurately measured and random effect differences across persons will be small, leading to homogeneous predictions across persons. The multivariate probit model will allow for more heterogeneity across persons in possible response values by avoiding conditioning on an unreliable and poorly discriminating random effect measurement. We developed an MCMC algorithm to facilitate the imputation of item responses and latent variable scores under IRT and CTT from their posterior predictive distributions.

In the present study, the influence of missing item-level data on the growth model estimates are examined under different missing data mechanisms. A missing at random

(MAR) mechanism is investigated, which has been extensively studied in scale analysis [15, 16, 17], and is the utmost possibility under the presented model. An analysis based on the MAR assumption can produce consistent estimates under a MAR or missing completely at random (MCAR) mechanism and can increase power relative to an MCAR-based analysis (e.g., listwise or pairwise deletion). When assuming MAR in latent growth analysis, it is expected that the probability of missing responses is explained by the repeatedly assessed longitudinal latent variable, which constitutes the response data as well as the covariate information. When conditioning on the longitudinal latent variable, and given an adequate fit of the IRT model, it is reasonable to assume that the probability of missing a response is independent of the missing response itself. In the situation where the available item scores only measure a part of the latent variable, the MAR assumption will not hold. Then, without controlling for the latent variable, the probability of missing a response depends on the missing response and the missing data mechanism is missing not at random (MNAR).

The developed imputation method is used in the current study in different simulation studies to examine the effects of missing data on latent growth parameter estimates, given longitudinal latent variables as outcomes measured using IRT or CTT. It is shown that the CTT model will lead to bias in the latent growth parameter estimates, which describe the shape of estimated growth trajectories. Furthermore, the use of the CTT model for measuring the longitudinal outcome variables will also lead to less individual variance in growth trajectories of the latent variable. In the first study, under CTT, predictive mean matching is used to impute missing item scores, which ignores the trend in the longitudinal latent variable scores, and will lead to bias in the growth parameter estimates. Next, to be able to include a trend in the measurements, the multivariate probit model is proposed to impute missing values to impute CTT scores, when the missing data mechanism is MAR. Subsequently, a more extreme missing data situation is considered, where the differences between the missing data probabilities for participants with a lower latent variable level differ more severely from participants with a higher latent variable score, and up till 50-70% of the responses in a response pattern can be missing. The findings of the simulation studies are illustrated in an example data set on the longitudinal development of low back pain [18].

In the next section, the considered latent growth model is introduced. The posterior predictive distributions are given of the latent variable scores and item scores, under IRT and CTT, to construct a multiple imputation method for the missing data. In a simulation study, different MAR-based missing data mechanisms are considered. They are analyzed using the proposed multiple imputation method to obtain latent growth estimates on a common scale under the different measurement models. In the presence of missing data, the implications of using sum-scores as measurements of latent variables in latent growth modeling is shown. Furthermore, a direct comparison is made with IRT-based multiple imputations. Subsequently, findings of the simulation study are illustrated in a real data study on coping with low back pain. Finally, conclusions and a discussion is given of the findings.

2 Methods

The considered latent growth model is discussed. Then, a multiple imputation method is introduced for the latent growth model analysis to deal with the missing latent variable values and missing item responses. The posterior predictive distribution of the latent variable and the item responses are presented under an IRT-based and an CTT-based framework as components of the imputation method.

2.1 Latent Growth Model

The latent growth model describes the changes in the latent variable given predictor variables and occasion-specific measurement information. Consider the following growth model for repeated measurements of a latent variable θ_{ij} for subject i and measurement occasion j,

$$\begin{aligned} \theta_{ij} &= \beta_{0i} + \beta_{1i} t_{ij} + e_{ij} \\ \beta_{0i} &= \gamma_0 + u_{0i} \\ \beta_{1i} &= \gamma_1 + u_{1i}, \end{aligned} \tag{1}$$

where t_{ij} is the j^{th} value for the measurement occasion for patient i, and $e_{ij} \sim N(0, \sigma^2)$ and $\mathbf{u}_i \sim N(0, \mathbf{T})$, and where $\mathbf{T}$ is a matrix with diagonal elements τ_0^2 and τ_1^2 and covariance parameter τ_{01}. The random effects are assumed to be multivariate normally distributed with covariance matrix $\mathbf{T}$. Parameter γ_0 is the population intercept, which represents the average latent variable score across persons on the measurement occasion where time equals zero. In most cases, the first measurement occasion corresponds with time point zero such that the random intercept variance, τ_0^2, can be interpreted as the between-subject variation in latent variable scores across subjects. Parameter γ_1 is the linear trend in the population, and represents the linear change in the latent variable across time. Variance parameter τ_1^2 on level two represents the variation in subject-specific trends across time. The random intercept and random slope are allowed to correlate, where τ_{01} is the covariance between the two random effects. The common error variance at level one is denoted by σ^2 and it represents the deviation between the subject-specific linear trend and the latent variable measurements.

2.2 Multiple-Imputation Methods Given Incomplete Response Data

Multiple imputation is a common way to deal with missing values and is based on missing data principles [19]. The multiple imputations are generated from the posterior predictive distribution of the latent variable to obtain a complete data set, treating the latent variable as missing data. All available information is used to construct the posterior of the latent variable. Samples from the posterior predictive distribution are made for patients given background characteristics in order to obtain (plausible) latent variable scores. Multiple sets of imputations are drawn to address the uncertainty associated with the latent variable scores.

Without missing item responses, multiple imputations can directly be generated from the posterior predictive distributions defined under the CTT or IRT model. The multiple imputations, also referred to as plausible values [20, 21], have the advantage over single point estimates that uncertainty associated with the latent variable scores is included in the estimation of the latent growth parameters. The multiple imputations are used to obtain unbiased estimates of the latent growth parameter estimates. Standard methods can be used to estimate growth parameters (e.g. multilevel modeling), when multiple imputations are used for the latent variable. When the data are collected through a complex sampling design, multiple imputations can be used to obtain correct standard deviations.

The multiple imputations are not restricted to a specific scale. The multiple imputations generated for the latent variable can be linearly transformed to any particular scale, for which the mean and variance can be freely specified. Results of latent growth parameter estimates, given the imputations derived from different posterior predictive distributions, can be directly compared, when transforming the generated imputations to a common scale. Therefore, a linear scale transformation can be applied to obtain results of a latent growth analysis on a common scale, while using posterior predictive distributions defined under different models (e.g., CTT and IRT) [22].

2.3 IRT-based Posterior Predictive Distributions

The combination of the IRT model with a latent growth model to model the latent variable can be viewed as a generalized linear multilevel model. For dichotomous items for instance, a normal ogive model can be used to model the probability of patient i to give a positive response ($Y_{ijk} = 1$) to item k on measurement occasion j, which is given by

$$P\left(Y_{ijk} = 1 \mid \theta_{ij}, a_k, b_k\right) = \Phi\left(a_k \theta_{ij} - b_k\right), \tag{2}$$

where Φ is the normal cumulative distribution function and a_k and b_k are the discrimination and the location parameter of item k, respectively [23, 24]. Latent variable

θ_{ij} denotes the latent trait for patient i at measurement occasion j. The item response observations are modeled by the IRT model (level one), and the patient-specific latent variables are modeled by the latent growth model (level two (time) and level three (patients)). The IRT model utilizes all available response information to measure the latent variable, where the variability in response patterns across time and persons, is represented in the variability across measured latent variable scores.

In the Bayesian modeling approach, the priors and hyper priors are defined for the item parameters a_k and b_k. The log-transformation is used for the discrimination parameters to restrict their values to be positive, then the multivariate distribution for the item parameters a_k and b_k is given by,

$$p\left(\log(a_k), b_k\right) \sim N\left(\left(\mu_a, \mu_b\right)^t, \mathbf{\Sigma}_I\right),$$

where $\mathbf{\Sigma}_I$ is the covariance matrix of each item's parameters, and μ_a and μ_b are the average item discrimination and item location of the population of test items. Prior distributions are defined for the hyper prior parameters, μ_a,μ_b and $\mathbf{\Sigma}_{\boldsymbol{I}}$,

$$\begin{gathered}\mu_a, \mu_b \sim N\left(0, \sigma_I^2\right) \\ \mathbf{\Sigma}_I^{-1} \sim W\left(\nu_I, \mathbf{\Lambda}_I\right),\end{gathered}$$

where σ_I^2 can be specified to be large to represent an uninformative prior. The parameter ν_I is often small but greater or equal to two, and $\mathbf{\Lambda}_I$ is an identity matrix to define an uninformative prior.

The IRT model provides a posterior predictive distribution for each item response, when the model parameters are sampled from their respective posterior distributions. An MCMC algorithm provides sampled values for the model parameters in iteration m, denoted as $\theta_{ij}^{(m)}$, $a_k^{(m)}$, $b_k^{(m)}$. The posterior predictive distribution for the item responses is a Bernoulli distribution, where the success probability is determined by parameter values sampled from their posterior distributions;

$$\begin{aligned} Y_{ijk}^{(m)} \mid \theta_{ij}^{(m)}, a_k^{(m)}, b_k^{(m)} &\sim B\left(\pi_{ijk}^{(m)}\right), \\ \pi_{ijk}^{(m)} &= \Phi\left(a_k^{(m)}\theta_{ij}^{(m)} - b_k^{(m)}\right) \\ &= \Phi\left(a_k^{(m)}\left(\beta_{0i}^{(m)} + \beta_{1i}^{(m)}t_{ij} + e_{ij}^{(m)}\right) - b_k^{(m)}\right). \end{aligned} \tag{3}$$

When assuming missing item response data to be MAR, multiple imputations can be generated from the posterior predictive distribution to simulate plausible values for the missing responses. In the last expression, the growth model for the latent variable (Equation (1)) is integrated. This shows that the change in the latent variable is represented in the success probability, where the term $\beta_{1i}^{(m)} t_{ij}$ represents the change over time and the time-invariant term $\beta_{0i}^{(m)}$ represents the person-specific latent variable level at the start of the study $(t = 0)$.

When using an MCMC algorithm for the estimation of the IRT parameters, the latent response formulation is often used, which facilitates the sampling of the model parameters. However, the latent item response distribution can also be used to construct a posterior predictive distribution. The latent response data Z_{ijk} is introduced as augmented data, and is assumed to be truncated normally distributed,

$$\begin{aligned} Z_{ijk}^{(m)} \mid Y_{ijk}, \theta_{ij}^{(m)}, a_k^{(m)}, b_k^{(m)} \quad &\sim \quad N(a_k\theta_{ij} - b_k, 1) \\ & \quad\; \begin{cases} I(Z_{ijk} \leq 0) & \text{if} \quad Y_{ijk} = 0 \\ I(Z_{ijk} > 0) & \text{if} \quad Y_{ijk} = 1 \\ Y_{ijk} \text{ missing.} \end{cases} \end{aligned} \tag{4}$$

If the observation Y_{ijk} is missing, then the posterior simulated value $Y_{ijk}^{(m)}$ equals one if the sampled value $Z_{ijk}^{(m)}$ is greater than zero, and equals zero if $Z_{ijk}^{(m)}$ is smaller than zero. The simulated latent responses can often be used as imputations and values for the dichotomous missing responses, $Y_{ijk}^{(m)}$, do not need to be determined. For instance, the posterior predictive distribution of the latent variable can be determined from the latent responses, which is described next.

The posterior predictive distribution of the latent variable θ_{ij} is constructed from the latent growth model and the IRT model. Let $\mathbf{\Omega}_{ij}$ denote the set of latent growth parameters, $\mathbf{\Omega}_{ij} = \{\boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma^2, \mathbf{T}\}$ for patient i and occasion j. When conditioning on latent item response data, it can be shown that the posterior (predictive) distribution of the latent variable is normal. The posterior distribution of the latent variable can be expressed as

$$g\left(\theta_{ij} \mid \mathbf{z}_{ij}, \mathbf{\Omega}_{ij}, \boldsymbol{a}, \boldsymbol{b}\right) \quad \propto \quad p\left(\boldsymbol{z}_{ij} \mid \theta_{ij}, \boldsymbol{a}, \boldsymbol{b}\right) f\left(\theta_{ij} \mid \mathbf{\Omega}_{ij}\right) \tag{5}$$

It follows that the posterior distribution $g()$ is normal, since it is constructed from normally distributed latent responses and a normal prior distribution, where the mean is given by

$$E\left(\theta_{ij} \mid \mathbf{z}_{ij}^{(m)}, \mathbf{\Omega}_{ij}^{(m)}, \boldsymbol{a}^{(m)}, \boldsymbol{b}^{(m)}\right) \quad = \quad \frac{\mathbf{a}^t\left(\mathbf{z}_{ij} + \mathbf{b}\right) + \left(\beta_{0i} + \beta_{1i}t_{ij}\right)/\sigma^2}{(\mathbf{a}^t\mathbf{a})^{-1} + \sigma^{-2}}$$

and the variance by $(1/(\mathbf{a}^t\mathbf{a}) + 1/\sigma^2)^{-1}$. When considering sampled parameter values for the parameters conditioned on, posterior predictive values (i.e., plausible values) for the latent variable can be sampled from

$$\theta_{ij}^{PV} \sim g\left(\theta_{ij} \mid \mathbf{z}_{ij}^{(m)}, \mathbf{\Omega}_{ij}^{(m)}, \boldsymbol{a}^{(m)}, \boldsymbol{b}^{(m)}\right), \tag{6}$$

and they can be linearly transformed to a particular scale. The density function $f()$ represents the latent growth model, the $p()$ represents the likelihood function, and the posterior density function $g()$ represents the posterior predictive distribution from which missing latent variable scores can be drawn.

2.4 CTT-based Posterior Predictive Distributions

In CTT, the (aggregate) sum-score is used as a measurement of the latent variable score. A true score, ϑ_{ij}, is assumed, the theoretical construct value, which is never observed. The observed sum-score is assumed to be equal to the true score plus an error term. When also assuming normally distributed errors, the CTT model is given by

$$\overline{y}_{ij} \quad = \quad \vartheta_{ij} + e_{ij} \tag{7}$$

where $e_{ij} \sim N\left(0, \sigma^2_{\vartheta}\right)$, and $\overline{y}_{ij} = \sum_k y_{ijk}/K$ is the average score over K item responses. The measurement error variance is assumed to be equal across persons. One characteristic of the CTT model is that the model is defined at the level of latent variable scores, while the observations are defined at the item level. Therefore, differences between response patterns leading to the same sum score are ignored in measuring the latent variable score. This loss of information in determining the latent variable scores leads to less variability in scores, when comparing it to the variability in observed response patterns. Another characteristic is that missing item responses lead directly to missing sum scores, since sum scores cannot be determined from an incomplete response pattern. The handling of missing item responses is further complicated, since the measurement model is defined at the level of latent variable scores. Therefore, in contrast to the IRT model, a different model is needed to generate imputations for missing item responses. However, it is important to use an imputation model to preserve the original relationships between the variables in the data.

2.4.1 CTT-based Multiple Imputation Model Response patterns with missing observations complicate the computation of sum scores, $\overline{y}_{ij}$, which are needed to generate multiple imputations for the latent variable ϑ_{ij}. The sum score is assumed to be the sum (or average) over dichotomous responses. Thus, the CTT model is simply not defined at the level of dichotomous observations, and the posterior predictive distribution of item responses cannot be constructed from the CTT model. Now, consider the CTT model in Equation (7), with a latent growth model for the true score ϑ_{ij} with mean $\beta_{0i} + \beta_{1i}t_{ij}$ and variance σ^2_{ϑ}. An underlying latent variable is defined, in a similar way as the one defined under the IRT model in Equation (4). Let Z_{ijk} be normally distributed with mean ϑ_{ij} and variance 1, where Z_{ijk} is greater than zero for a positive response, and less equal zero otherwise. The CTT model for the latent (continuous) responses $(Z_{ij1}, \ldots, Z_{ijK})$ is a linear random effects model. A marginal CTT model can be derived, where the true score is integrated out. Integrate the mean expression for the true score, according to the latent growth model in Equation (1), in the CTT model. Then, merge the error term at the observation level with the error term defined at the person level. This leads to a multivariate normal distribution for the item responses:

$$\begin{aligned} Z_{ijk} &= \beta_{0i} + \beta_{i1} t_{ij} + e_{ij} + r_{ijk} \\ &= \beta_{0i} + \beta_{i1} t_{ij} + E_{ijk} \end{aligned} \tag{8}$$

where the r_{ijk} are independently normally distributed with mean 0 and variance 1 and the e_{ij} are normally distributed with mean 0 and variance σ^2_ϑ, according to Equation (7). The error terms are independently distributed and the sum of the terms, E_{ijk}, is again normally distributed with mean 0 and variance $1 + \sigma^2_\vartheta$. The covariance of latent responses of item k and l of person i is equal to σ^2_ϑ, which is shown by

$$Cov\left(Z_{ijk}, Z_{ijl}\right) = Cov\left(e_{ij} + r_{ijk}, e_{ij} + r_{ijl}\right) = Var\left(e_{ij}\right) = \sigma^2_\vartheta.$$

For dichotomous observed data, a marginal CTT model (i.e., multivariate probit model) can be defined for the observed responses, where the covariance matrix is equal to $\boldsymbol{\Sigma} = \boldsymbol{I}_K + \boldsymbol{J}\sigma^2_\vartheta$ with $\boldsymbol{J} = \mathbf{1}_K \mathbf{1}^t_K$, which is the implied covariance structure according to the CTT model. Then, the marginal CTT model is given by

$$P\left(\mathbf{Y}_{ij} = \mathbf{y}_{ij} \mid \mu_\vartheta, \boldsymbol{\Sigma}\right) = \int_{R_{ij1}} \cdots \int_{R_{ijK}} \Phi\left(\mathbf{z}_{ij} \mid \mu_\vartheta, \boldsymbol{\Sigma}\right) dz_{ij1} \ldots dz_{ijK},$$

where R_{ijk} is the interval $(0, \infty)$ if $Y_{ijk} = 1$ and the interval $(-\infty, 0)$ otherwise, and $\mu_\vartheta = \beta_{0i} + \beta_{1i} t_{ij}$.

The object is to define the posterior predictive distribution for the item responses under the CTT model to define a multiple imputation method. For the marginal CTT model, the posterior predictive distribution of the latent response Z_{ijk} given the remaining $(K-1)$ responses $\mathbf{Z}_{ij(-k)}$ can be derived from their multivariate normal distribution. The covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{I}_K + \boldsymbol{J}\sigma^2_\vartheta$ can be partitioned, where $\boldsymbol{\Sigma}_{12}$ defines the covariance between the Z_{ijk} and the $\mathbf{Z}_{ij(-k)}$, and $\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{22}$ is the variance and covariance matrix of Z_{ijk} and $\mathbf{Z}_{ij(-k)}$, respectively. Then, the posterior predictive distribution of the latent response is given by,

$$Z_{ijk} \mid \mathbf{Z}_{ij(-k)}, \mu_\vartheta, \boldsymbol{\Sigma} \sim N\left(\mu_{ijk}, \sigma^2_{ijk}\right) \tag{9}$$

where

$$\begin{aligned}
\mu_{ijk} &= \mu_\vartheta + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\left(\mathbf{Z}_{ij(-k)} - \mu_\vartheta\right) \\
&= \mu_\vartheta + \sigma_\vartheta^2 \mathbf{1}_{K-1}^t \left(\boldsymbol{I}_{K-1} - \frac{\mathbf{1}_{K-1}\mathbf{1}_{K-1}^t}{1/\sigma_\vartheta^2 + K - 1}\right)\left(\mathbf{Z}_{ij(-k)} - \mu_\vartheta\right) \\
&= \mu_\vartheta + \frac{\sum_{l \neq k} \sigma_\vartheta^2 \left(Z_{ijl} - \mu_\vartheta\right)}{1 + (K-1)\sigma_\vartheta^2},
\end{aligned}$$

where the inverse of $\boldsymbol{\Sigma}$ and a partition of it, $\boldsymbol{\Sigma}_{22}$, can be found in [25][pp. 152]. Note that the mean is represented by the latent growth component; $\mu_\vartheta = \beta_{0i} + \beta_{1i}t_{ij}$. Thus, the predicted response will take into account a linear trend in the latent variable measurements. Finally, the variance σ_{ijk}^2 is equal to

$$\begin{aligned}
\sigma_{ijk}^2 &= \Sigma_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}^t \\
&= 1 + \sigma_\vartheta^2 - \sigma_\vartheta^2 \mathbf{1}_{K-1}^t \left(\boldsymbol{I}_{K-1} - \frac{\mathbf{1}_{K-1}\mathbf{1}_{K-1}^t}{1/\sigma_\vartheta^2 + K - 1}\right)\sigma_\vartheta^2 \mathbf{1}_{K-1}^t \\
&= \frac{1 + K\sigma_\vartheta^2}{1 + (K-1)\sigma_\vartheta^2}.
\end{aligned}$$

The posterior predictive distribution of the latent response depends on the covariance parameter σ_ϑ^2. This parameter represents the covariance between item responses, when not conditioning on the true score. The σ_ϑ^2 can be sampled from its posterior distribution given the latent responses. To obtain the posterior distribution of the covariance parameter, σ_ϑ^2, consider the multivariate distribution of the latent response data, which is given by

$$\begin{aligned}
P\left(\mathbf{Z}_{ij} = \mathbf{z}_{ij} \mid \mu_\vartheta, \boldsymbol{\Sigma}\right) &= (2\pi)^{\frac{-K}{2}} \left|\boldsymbol{\Sigma}^{-1}\right|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\mathbf{z}_{ij} - \mu_\vartheta\right)^t \boldsymbol{\Sigma}^{-1}\left(\mathbf{z}_{ij} - \mu_\vartheta\right)\right) \\
&= (2\pi)^{\frac{-K}{2}} \left|\boldsymbol{\Sigma}^{-1}\right|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\mathbf{E}_{ij}^t \left(\mathbf{I}_K - \frac{\mathbf{1}_K\mathbf{1}_K^t}{1/\sigma_\vartheta^2 + K}\right)\mathbf{E}_{ij}\right)
\end{aligned}$$

where the expression for inverse of the covariance matrix $\boldsymbol{\Sigma}$ is used. The sum of squares in the exponent can be partitioned in two components, where one component is the sufficient statistic for the covariance parameter. To see this, the terms in the exponent are rearranged as follows,

$$\begin{aligned}
\mathbf{E}_{ij}^t \left(\mathbf{I}_K - \frac{\mathbf{1}_K\mathbf{1}_K^t}{1/\sigma_\vartheta^2 + K}\right)\mathbf{E}_{ij} &= \sum_k E_{ijk}^2 - \frac{\sum_k E_{ijk} \sum_k E_{ijk}}{1/\sigma_\vartheta^2 + K} \\
&= \sum_k \left(E_{ijk} - \bar{E}_{ij}\right)^2 + K\bar{E}_{ij}^2 - \frac{K^2\bar{E}_{ij}^2}{1/\sigma_\vartheta^2 + K} \\
&= \sum_k \left(E_{ijk} - \bar{E}_{ij}\right)^2 + K\bar{E}_{ij}^2 \left(1 - \frac{K}{1/\sigma_\vartheta^2 + K}\right)
\end{aligned}$$

$$= \sum_k \left(E_{ijk} - \bar{E}_{ij}\right)^2 + \frac{K\bar{E}_{ij}^2}{1+\sigma_\vartheta^2 K}.$$

The second component on the right-hand side, $S_{B_{ij}} = K\bar{E}_{ij}^2$, represents the sum of squares which contains the information about σ_ϑ^2. The term $\sum_{i,j} \bar{E}_{ij}^2$ is chi-square distributed, since the average error term $\bar{E}_{ij}$ is independently normally distributed across persons and occasions. It follows that posterior distribution of σ_ϑ^2 is given by

$$p\left(\sigma_\vartheta^2 \mid \mathbf{z}\right) \propto \left(1/K + \sigma_\vartheta^2\right)^{\frac{NJ}{2}-1} \exp\left(-\frac{\sum_{i,j} \bar{E}_{ij}^2}{2\left(1/K + \sigma_\vartheta^2\right)}\right), \qquad (10)$$

using an uninformative prior

$$p\left(\sigma_\vartheta^2\right) \propto \left(1/K + \sigma_\vartheta^2\right)^{-1}.$$

Following Fox et al.[26], the posterior distribution of σ_ϑ^2 in Equation (10) can be recognized as a shifted-inverse gamma with shape parameter NJ/2 and rate parameter $S_B/2 = \sum_{i,j} \bar{E}_{ij}^2/2$. Note that the covariance parameter, $\rho = \sigma_\vartheta^2$, is restricted to be greater than $-1/K$ and is allowed to be negative, since the term $1/K + \sigma_\vartheta^2$ is restricted to be positive.

This marginal CTT model is defined at the level of observations and represents the imputation model for the missing responses. An algorithm can be defined to simulate scores constructed from observed and imputed responses under the CTT model. The following steps are defined in iteration m of an MCMC algorithm to generate sum scores partly based on imputed responses:

1. Simulate $\sigma_\vartheta^{2(m)}$ given $\mathbf{z}^{(m-1)}, \mu_\varphi^{(m-1)}$ from a shifted-inverse gamma distribution in Equation (10).
2. Simulate values $Z_{ijk}^{(m)}$ for all missing responses k for person i given $\mathbf{Z}_{ij(-k)}^{(m-1)}, \mu_\vartheta^{(m-1)}$, and $\mathbf{\Sigma}^{(m)}$ according to Equation (9).Then, $Y_{ijk}^{(m)}$ equals 1 if $Z_{ijk}^{(m)} > 0$ and $Y_{ijk}^{(m)} = 0$ if $Z_{ijk}^{(m)} \leq 0$. Update the sum score $\overline{y}_{ij}^{(m)}$ given observed and simulated responses.

The posterior predictive distribution of the true score is constructed from the latent growth component and the CTT model. It follows that

$$g\left(\vartheta_{ij} \mid \overline{y}_{ij}, \mathbf{\Omega}_{ij}\right) \propto p\left(\overline{y}_{ij} \mid \vartheta_{ij}\right) f\left(\vartheta_{ij} \mid \mathbf{\Omega}_{ij}\right), \qquad (11)$$

represents a normal distribution with mean

$$E\left(\vartheta_{ij} \mid \overline{y}_{ij}, \sigma_\vartheta^2, \mathbf{\Omega}\right) = \frac{\overline{y}_{ij}/\sigma_\vartheta^2 + \left(\beta_{0i} + \beta_{1i}t_{ij}\right)/\sigma^2}{1/\sigma_\vartheta^2 + 1/\sigma^2}$$

and variance $(1/\sigma_\vartheta^2 + 1/\sigma^2)^{-1}$, see also [22]. Subsequently, the multiple imputations are drawn from

$$\vartheta_{ij}^{PV} \sim g\left(\vartheta_{ij} \mid \overline{y}_{ij}, \sigma_\vartheta^{2(m)}, \mathbf{\Omega}_{ij}^{(m)}\right), \tag{12}$$

and the drawn values can be transformed to a particular scale using a linear transformation.

Other methods for handling missing item scores in multilevel data are based on multilevel imputation strategies. In Mice, the functions ”2l.norm” and ”2l.pan” offer solutions for continuous data. A reasonable way of handling the missing data in the CTT framework would be to use a multilevel logistic regression imputation or multilevel predictive mean matching method. The binary multilevel imputation method by Jolani et al. [13] is designed for structural missing data, and therefore not applicable for the current study. The miceadds R-package [27] provides a function for multilevel logistic multiple imputation. In the simulation study, this procedure led to convergence issues for the considered data, when integrating it in the imputation algorithm for latent growth modeling. The multilevel logistic multiple imputation method in miceadds is build on the generalized linear mixed effects routine in lme4. The routine requires sufficient observed data to estimate random effect(s) and other parameters to simulate values. More research is needed to assess empirical differences between the multivariate probit imputation model and the generalized multilevel imputation model. However, this is beyond the scope of the current study.

3 Bayesian Multiple Imputation Method for Latent Variables and Missing Item Responses

The multiple imputation method is aimed at drawing values from the posterior predictive distribution given observed data and parameters. The parameters are drawn from their posterior distribution given the observed data and realizations of the missing values.

3.1 Multiple Imputation Algorithm Under IRT

Step (I) Draw parameters of the imputation model: Draw item parameters $\mathbf{a}$, $\mathbf{b}$ and latent growth parameters $\mathbf{\Omega}$ from their posterior distribution given the complete data $\mathbf{y}$, where the complete data is constructed from imputed missing data and observed data.

Step (II) Generate multiple imputations for the missing data and latent variable: Draw values for the missing data given sampled parameter values according to Equation (4) (also possible Equation (3)), and draw latent variable scores from the posterior predictive distribution defined in Equation (6).

In Step (I), drawing the parameter values of the imputation model can be achieved by MCMC. The MCMC scheme for complete data is described in [22], which is based on the augmentation of latent responses as described in Equation (4). In Step (II), values for the missing data can be generated from the posterior predictive distribution, from which latent item responses are generated. This is sufficient for the sampling of the model parameters and latent variable scores. When binary predicted values are needed, positive and negative binary predicted values are obtained by generated latent responses located in the interval $(0, \infty)$ and $(-\infty, 0)$, respectively.

3.2 Multiple Imputation Algorithm Under CTT

Step (I) Draw parameters of the imputation model: The parameter, ϑ_{ij}, can be sampled from its posterior distribution represented in Equation (11), by sampling $\lambda = 1/K + \sigma_{\vartheta}^2$ and subtracting $1/K$ to obtain a sample for σ_{ϑ}^2. The latent growth parameters $\mathbf{\Omega}$ can be sampled from their posterior distribution as described under IRT, while using the current value for $\boldsymbol{\vartheta}$ as the outcome variable given the complete data $\mathbf{y}$, where the complete data is constructed from sampled missing data and observed data.

Step (II) Generate multiple imputations for the missing data and latent variable: Draw values for the missing data given sampled parameter values from Equation (11) and draw latent variable scores from the posterior predictive distribution defined in Equation (12).

As for IRT, in Step (I), the parameter values of the imputation model can be drawn using MCMC. In Step (II), the generated latent continuous missing data lead to binary predicted data, where the sign of the generated imputed values determines whether the binary predicted value is one or zero. Note that these predicted binary (missing) values are not directly needed.

The algorithm has been implemented in a modified version of the R-Package mlirt [28]. The convergence of the algorithm can be investigated by observing trace plots of the sampled values. At convergence, the sequences of sampled values should mix well and not show any structural patterns. The convergence diagnostics in the R-package Coda [29] can also be used to investigate whether the chains of sampled values has converged.

A general procedure can be applied to obtain parameter estimates of the LCM model using multiple imputations for the latent variable under the IRT model (Equation (6)) and CTT model (Equation (12)), while accounting for missing item responses using multiple imputations for the missing item scores. The final estimates of the latent growth model parameters, given the IRT-based and CTT-based multiple imputations for the missing item scores, are on the same scale due to the scale transformation of the multiple imputations for the latent variable.

(i) Generate multiple imputations for the latent variables θ_{ij} and ϑ_{ij} and missing responses, according to Step (II) of the IRT-based and CTT-based multiple imputation method.

(ii) Transform each vector of multiple imputations for the latent variable to a common scale using a linear scale transformation.

(iii) For each set of multiple imputations for the latent variable, draws of all LCM parameters are obtained using an MCMC algorithm.

(iv) Repeat steps (i)-(iii) multiple times (usually five).

(v) Pool the LCM estimation results from the IRT- and the CTT-generated multiple imputations.

4 Simulation Study

In the simulation study, latent growth model parameter estimates using IRT-based and CTT-based measurements of the outcome variable were compared to address the effects of the measurement model and missing response data on the estimation results. Three simulation studies are presented to show the advantage of using IRT- over CTT-based multiple imputations in longitudinal data analysis in the presence of missing item responses. The three simulation studies investigated the retrieval of the true values of the latent growth parameters under different missing data situations for different missing data mechanisms.

A general procedure was used to enable a direct comparison between the estimation results, while using different measurement models and imputation methods. In Figure 1, the general multiple imputation procedure is illustrated including the simulation of item response data. Data were simulated under specific conditions. Under CTT, multiple imputations were generated. In Simulation Study I, Predictive Mean Matching (PMM) was used, which ignored the trend in latent variable measurements. Under IRT, the generation of multiple imputations for the missing item scores was facilitated via the algorithm described in Section 3.

In Simulation Study II and III, the proposed CTT-based multiple imputation method was used using the multivariate probit model, where the imputed values for the true scores also addressed a linear trend in measurements. Following the procedure of Gorter et al. [5], the multiple imputations for the latent variable under IRT and CTT were rescaled to a common scale, and the latent growth model was fitted with the generated latent variable scores as outcomes. The results of the latent growth model analysis under IRT and CTT were compared in terms of bias and mean squared error (MSE) of the parameter estimates. The structural model (LCM) parameters can be compared directly, since the CTT-based multiple imputations were re-scaled to the scale of the IRT-based multiple imputations. to assess measurement model differences and effects of missing data. Therefore, fixed effect parameter estimates from the latent growth model were averaged over the five multiple imputations to obtain the final results. The variance of the parameters were pooled by calculating the sum over the within-imputation variance (the average of the variance estimates), and the between-imputation variance (the variance of the point estimates) [25, pp. 168-169].

The following procedure was followed to simulate item response data. The person-specific scores at intake (β_{0i}) and the person-specific trends β_{1i} were generated from a multivariate normal distribution with covariance matrix $\mathbf{T}$, where the variation across persons at intake equaled $\tau_0^2 = 1$, the variation across person-specific trend values equaled $\tau_1^2 = .50$, and the population covariance between the intake score and trend equaled $\tau_{01}^2 = .20$. The latent variable scores were drawn from a normal distribution given the sampled β_{0i} and β_{1i}, with a population average intercept of $\gamma_0 = 0$, a population-average linear trend of $\gamma_1 = 1$, and measurement error variance of $\sigma^2 = .20$. The item difficulty parameters b_k were sampled from a normal distribution with mean zero and variance $.50$. Items with difficulty parameters above (below) zero were marked as difficult (easy) items. The discrimination parameters were equal to one. Finally, dichotomous item responses were generated using the IRT model (Equation (2)), given the generated latent variable scores.

4.1 Simulation Study I

In the first simulation study, the probability of a missing item response was equal across all subjects and measurement occasions. However, this corresponded to MAR, since a positive linear trend was simulated in the latent variable measurements. Subsequently, positive responses were more likely to be missing at a later stage, and the missingness was explained by the trend in latent variable measurements. Conditional on the measurement occasion, the values were missing completely at random (MCAR).The percentage of missing data was varied. Four situations were investigated, the complete data set, and data with 20%, 50%, and 70% missingness. The missing data percentage represented the percentage of subjects with missing data, as well as the percentage of missing item responses per measurement occasion. For instance, in a condition with 20% missing item responses, also 20% of the subjects had missing item responses. Per condition, 50 replications were made with $N = 100$ patients responding to $K = 20$ dichotomous items on $J = 5$ measurement occasions. For CTT, multiple imputations were generated using Predictive Mean Matching (PMM), which resulted in five complete data sets. The *mice* package [30] was used to generate imputations using PMM on the wide dataset. After imputing the missing data, multiple imputations were generated for the latent variable. This resulted in five times five complete data sets. The latent growth model parameter estimates were pooled first for the five PMM imputed data sets, and pooled second over the five draws of the latent variable score. For IRT, multiple imputations for the missing item responses and latent variable scores were generated, according to Equation (3) and (6), respectively. Subsequently, latent growth model parameter estimates were pooled to obtain the final estimates. For estimating the parameters of the latent growth model, a $20,000$ iterations long MCMC chain was run, with $5,000$ burnin iterations per replication

4.2 Simulation Study II

In the second simulation study, the missing data was generated under MAR, where the missingness depended on the latent variable, which followed a linear trend. It was assumed that the observed responses contained sufficient information regarding the latent variable for estimating the latent variable scores [31]. Subjects with a lower latent variable score (below the population average) had a higher probability of missingness on the more difficult items than subjects with a higher latent variable score (above the population average) who had a lower probability of missingness on the more difficult items. A normal distribution was assumed for the population model. This population model combined with the observed (incomplete) data contained sufficient information on the latent trait to assume MAR. In Table 1, the different probabilities of missing response data are listed, which were used to generate missing data according to MAR. For example, in the 50% missing data condition, for those with a below-average latent variable score, the probability of a missing response was 70% for the difficult items, and 30% for the easy items. Those with an above-average latent variable score, the missing response probability was 60% and 40% for the difficult and easy items, respectively. On average the probability of a missing response was 50%. Subjects with a lower latent variable score were more likely to have missing responses on the more difficult items compared to subjects with a higher score, while accounting for the trend in latent variable scores. Three different missing data conditions were simulated with three different percentages of missing data, 20%, 50%, and 70%. The simulated data sets consisted of $N = 100$ patients measured on five different occasions on a questionnaire with 20 dichotomous items. A total of 50 replications were made per condition.

	20% Missing		**50% Missing**		**70% Missing**	
	Easy	Difficult	Easy	Difficult	Easy	Difficult
Low θ	.12	.28	.30	.70	.42	.98
High θ	.16	.24	.40	.60	.56	.84

Table 1. Missing data probabilities for subjects with low versus high latent variable scores for a total of 20%, 50%, and 70% missing item responses used in Simulation Study 2.

Multiple imputations were generated using the proposed CTT-based multiple imputation method with the multivariate probit model, which resulted in five complete data sets.After generating the multiple imputations, the latent growth model parameters were estimated using the five draws of IRT-based and CTT-based multiple imputations, using a $20,000$ iterations long MCMC chain with $5,000$ burn-in iterations.

4.3 Simulation Study III

A more extreme missing data situation was investigated in order to test the robustness of the multiple imputation methods, and to examine differences between IRT and CTT under more extreme conditions. Missing data was simulated in such a way that it became less likely that the data contained enough information to estimate the latent variable accurately. The missingness was dependent on the latent variable and in the extreme situation the observed data only contained partial information about the latent variable, and consequently MNAR could occur in such a situation. In Table 2, the probabilities of missing responses are given for the high versus low scoring subjects. The probabilities in Table 2 were chosen in such a way that subjects with a lower latent variable score had a higher probability of missingness for the difficult items, whereas for patients with a higher latent variable score, the missingness was not dependent on the difficulty of the items. When the population distribution is assumed to be normally distributed and the observed (incomplete) data is sufficient we can assume MAR. However, with the increasing relative amount of missingness in the difficult items for patients with a lower theta we approach the situation in which the observed data combined with the population distribution no longer contain sufficient information (MNAR). The missing data conditions entailed data sets with 20%, 30%, 40%, and 50% missing observations. For instance, a subject scoring below the population average had a 90% probability of missing a response to a difficult item in the 50% missing data condition. Observed data for this subject mainly contained information about the performance on the easy items, which usually leads to inaccurate measurements of the latent variable scores. In this study, the same estimation procedures and parameter settings were used as in Simulation Study II.

	20% Missing		**30% Missing**		**40% Missing**		**50% Missing**	
	Easy	Difficult	Easy	Difficult	Easy	Difficult	Easy	Difficult
Low θ	.04	.36	.06	.54	.08	.72	.10	.90
High θ	.20	.20	.30	.30	.40	.40	.50	.50

Table 2. Probabilities of a missing response for subjects with low versus high latent variable scores with a total of 20%, 30%, 40% and 50% missing responses used in Simulation Study III.

5 Results

5.1 Results Simulation Study I

In Table 3 and Table 4, the results of Simulation Study I are given. The latent growth parameter estimates along with the bias and MSE using IRT-based and CTT-based multiple imputations are given.

		IRT				**CTT**			
Par.	True	EAP	SD	BIAS	MSE	EAP	SD	BIAS	MSE
0% Missing									
Fixed									
γ_0	0	-.02	.12	-.02	.01	.08	.14	.08	.02
γ_1	1	1.00	.12	.00	.02	.83	.16	-.17	.05
Random									
σ^2	.20	.20	.03	.00	.00	.35	.04	.15	.03
τ_0^2	1.00	1.06	.19	.06	.03	1.18	.20	.18	.07
τ_1^2	.50	.63	.21	.13	.07	.31	.15	-.19	.05
τ_{01}^2	.20	.08	.11	-.12	.02	-.01	.09	-.21	.05
20% Missing									
Fixed									
γ_0	0	.00	.12	.00	.02	.70	.23	.70	.79
γ_1	1	1.01	.12	.01	.02	-.15	.32	-1.15	2.08
Random									
σ^2	.20	.21	.03	.01	.00	.69	.09	.49	.27
τ_0^2	1.00	.99	.21	-.01	.04	1.08	.34	.08	.07
τ_1^2	.50	.55	.19	.05	.03	.40	.23	-.10	.02
τ_{01}	.20	.22	.14	.02	.02	-.19	.27	-.39	.19

par.: parameter estimate, True: simulated parameter values, EAP: expected a posteriori, MSE: mean squared error.

Table 3. Simulation Study 1: Latent growth parameter estimates across 50 replications under varying proportions of missing data (0% and 20%), using IRT-based multiple imputations and CTT-based multiple imputations for the latent variable scores. For CTT, missing data were generated according to PMM on the wide data set, and for IRT, they were generated from the posterior predictive distribution (Equation (3)).

In the condition without any missing data, a similar pattern was found as shown by Gorter et al. [5, 32]. When looking at the results in Table 3, it can be seen that the linear trend was underestimated under CTT, where the trend was correctly estimated under IRT. Furthermore, the measurement error variance at level one was overestimated under CTT, and also correctly estimated under IRT. The IRT-based multiple imputations provided more information about the linear trend in the measurements than the CTT-based imputations. This showed that the IRT-based trajectories more accurately described the change in measurements than the CTT-based trajectories. There was also more variability

		IRT				CTT			
Par.	True	EAP	SD	BIAS	MSE	EAP	SD	BIAS	MSE
50% Missing									
Fixed									
γ_0	0	.02	.12	.02	.01	.59	.25	.59	.50
γ_1	1	1.01	.13	.01	.02	.05	.37	-.95	1.26
Random									
σ^2	.20	.20	.03	.00	.00	.85	.10	.65	.47
τ_0^2	1.00	1.00	.22	.00	.05	.85	.33	-.15	.08
τ_1^2	.50	.56	.20	.06	.03	.44	.27	-.06	.01
τ_{01}^2	.20	.17	.15	-.03	.03	-.18	.27	-.38	.16
70% Missing									
Fixed									
γ_0	0	.01	.12	.01	.02	.50	.20	.50	.43
γ_1	1	1.01	.13	.01	.02	.20	.29	-.80	1.06
Random									
σ^2	.20	.20	.03	.00	.00	.91	.11	.71	.54
τ_0^2	1.00	1.05	.24	.05	.04	.86	.32	-.14	.07
τ_1^2	.50	.54	.20	.04	.03	.45	.29	-.05	.01
τ_{01}^2	.20	.18	.16	-.02	.03	-.19	.28	-.39	.17

par.: parameter estimate, True: simulated parameter values, EAP: expected a posteriori, MSE: mean squared error.

Table 4. Simulation Study I: Latent growth parameter estimates across 50 replications under varying proportions of missing data (50% and 70%), using IRT-based multiple imputations and CTT-based multiple imputations for the latent variable scores. For CTT, missing data were generated according to PMM on the wide data set, and for IRT, they were generated from the posterior predictive distribution (Equation (3)).

detected in the baseline measurements and less variability in the individual trends under CTT than under IRT. Under IRT, the variation in trends was slightly overestimated.

For the 20% missing data condition, under IRT the parameter estimates are close to the true values, and MSE values are even smaller than in the condition of no missing data. When the missing data are MAR, the IRT-based multiple imputations for the missing data provided accurate information leading to accurate multiple imputations for the latent variable. Under CTT, the PMM imputation model on the wide data set was time-invariant and did not take a trend effect in the repeated measurements into account. The corresponding CTT-based multiple imputations for the latent variable scores led to bias in the latent growth model parameter estimates. It can be seen that the linear trend was really underestimated, and only a third of the true trend value was identified, where the measurement error variance was estimated to be more than three times larger than the true variance. The underestimation of the trend also led to an overestimation of the average baseline score.

In Table 4, the results of the missing data conditions of 50% and 70% missing data are presented. It can be seen that the estimates under IRT are still close to the true values, and the multiple imputations correctly represent the trend in the latent variable scores. The multiple imputation model describes accurately the missing data and missing latent variable scores, while addressing the linear trend in the latent variable scores.

The CTT-based imputation method (PMM on the wide data set) ignored the trend in the latent variable scores. The estimated linear trend effect was no longer significantly different from zero, which can be seen from the 95% highest posterior density (HPD) interval $[-.27, .36]$ (50% missing data condition) and $[-.13, .52]$ (70% missing data condition). The intercept estimates γ_0 were overestimated in both conditions, and showed that the multiple imputed scores did not contain information about a trend in the latent variable measurements. The measurement error variance was severely overestimated in both conditions ($.85$ and $.91$), since the intercept did not explain much variance in the outcomes.

5.2 Results Simulation Study II

In Simulation Study II, the proposed multiple imputation (multivariate probit) model for the missing responses under CTT was used to address the linear trend in the latent variable measurements. This was ignored in Simulation Study I, which led to severe bias in the latent growth estimates. Under CTT and IRT, the missing responses were generated from the posterior predictive distribution according to Equation (9) and (3) respectively.

Under CTT, the trend in latent variable scores was taken into account in the generation of the imputations for the missing responses, which led to a substantial reduction in the bias of the latent growth parameter estimates. In Table 5, it can be seen that in the condition with 20% missing data, the CTT-based estimates are comparable to the estimates of Simulation Study I with no missing data. Under CTT, the linear trend is underestimated, the measurement error variance overestimated and the individual variation in trend effects was underestimated. However, this resembled the results under CTT without missing data, and this bias occurred due to the use of sum scores as measurements of the outcome variable. Under IRT, the latent growth parameters estimates were correctly estimated. The covariance of the random intercept and linear trend was underestimated, but the 95% highest posterior density (HPD) interval ($[-.12, .27]$) still contained the true value of $.20$. The IRT-based multiple imputations provided more information about the linear trend in the measurements than the CTT-based imputations.

When increasing the percentage of missing data, the parameter estimates under both measurement models were comparable to the results of the 20% missing data condition. The proposed posterior predictive distribution to generate missing responses takes into account the trend in latent variable scores. Under both measurement models, in the

			IRT				CTT			
	Par.	True	EAP	SD	BIAS	MSE	EAP	SD	BIAS	MSE
20% Missing										
Fixed	γ_0	0	-.01	.12	-.01	.01	.09	.15	.09	.02
	γ_1	1	1.00	.12	.00	.02	.84	.17	-.16	.05
Random	σ^2	.2	.20	.03	.00	.00	.36	.04	.16	.03
	τ_0^2	1	1.14	.21	.14	.07	1.25	.21	.25	.10
	τ_1^2	.5	.60	.21	.10	.08	.28	.14	-.22	.06
	τ_{01}^2	.2	.08	.11	-.12	.02	-.02	.09	-.22	.05
50% Missing										
Fixed	γ_0	0	.04	.13	.04	.02	.14	.15	.14	.04
	γ_1	1	.99	.14	-.01	.02	.81	.18	-.19	.05
Random	σ^2	.2	.20	.04	.00	.00	.37	.05	.17	.03
	τ_0^2	1	1.16	.22	.16	.08	1.27	.22	.27	.12
	τ_1^2	.5	.59	.22	.09	.06	.27	.15	-.23	.07
	τ_{01}^2	.2	.08	.12	-.12	.02	-.02	.09	-.22	.05
70% Missing										
Fixed	γ_0	0	.02	.13	.02	.01	.14	.15	.14	.03
	γ_1	1	1.01	.15	.01	.03	.79	.18	-.21	.06
Random	σ^2	.2	.21	.05	.01	.00	.40	.06	.20	.04
	τ_0^2	1	1.12	.23	.12	.07	1.20	.21	.20	.08
	τ_1^2	.5	.60	.27	.10	.08	.25	.14	-.25	.07
	τ_{01}^2	.2	.07	.12	-.13	.02	-.01	.08	-.21	.04

par.: parameter estimate, True: simulated parameter values, EAP: expected a posteriori, MSE: mean squared error.

Table 5. Simulation Study II: Latent growth parameter estimates across 50 replications under varying proportions of missing data (20%, 50% and 70%), using IRT-based multiple imputations and CTT-based multiple imputations for the latent variable scores. For CTT and IRT, missing data were generated from the posterior predictive distribution, Equation (9) and (3), respectively.

condition of 70% missing data, the bias in the latent growth parameter estimates was comparable to the bias of the parameter estimates in the condition of 50% and 20% missing data. When the missingness can be explained by the latent variable scores, under both measurement models stable results were obtained by imputing missing data from the posterior predictive distributions. Differences in latent growth estimates between IRT and CTT-generated latent variable scores were caused by using sum scores, which do not utilize all response information.

5.3 Results Simulation Study III

The probabilities of missing responses are displayed in Table 1. For the different missing data conditions, the latent growth estimates across 50 data replications are given in Table 6. Although part of the observed response patterns did not have sufficient information about the latent variable, the parameter estimates were quite close to the true values given

			IRT				CTT			
	Par.	True	EAP	SD	BIAS	MSE	EAP	SD	BIAS	MSE
20% Missing										
Fixed	γ_0	0	.01	.12	.01	.01	.10	.14	.10	.03
	γ_1	1	1.00	.13	.00	.02	.84	.14	-.16	.04
Random	σ^2	.2	.21	.03	.01	.00	.37	.04	.17	.03
	τ_0^2	1	1.11	.19	.11	.05	1.21	.18	.21	.08
	τ_1^2	.5	.58	.18	.08	.04	.27	.10	-.23	.06
	τ_{01}^2	.2	.07	.06	-.13	.02	-.02	.03	-.22	.05
30% Missing										
Fixed	γ_0	0	.01	.12	.01	.02	.09	.14	.09	.03
	γ_1	1	.97	.12	-.03	.01	.83	.12	-.17	.04
Random										
	σ^2	.2	.20	.04	.00	.00	.36	.04	.16	.03
	τ_0^2	1	1.16	.22	.16	.08	1.25	.20	.25	.10
	τ_1^2	.5	.56	.21	.06	.05	.27	.10	-.23	.06
	τ_{01}^2	.2	.08	.09	-.12	.02	-.01	.03	-.21	.04
40% Missing										
Fixed	γ_0	0	.01	.12	.01	.02	.09	.13	.09	.03
	γ_1	1	.99	.14	-.01	.02	.85	.15	-.15	.04
Random	σ^2	.2	.21	.04	.01	.00	.37	.05	.17	.03
	τ_0^2	1	1.14	.19	.14	.06	1.22	.21	.22	.09
	τ_1^2	.5	.56	.22	.06	.05	.26	.11	-.24	.07
	τ_{01}^2	.2	.06	.08	-.14	.03	-.02	.03	-.22	.05
50% Missing										
Fixed	γ_0	0	.00	.13	.00	.02	.11	.12	.11	.03
	γ_1	1	.99	.15	-.01	.02	.81	.14	-.19	.05
Random										
	σ^2	.2	.2	.04	.00	.00	.36	.05	.16	.03
	τ_0^2	1	1.16	.18	.16	.06	1.25	.19	.25	.10
	τ_1^2	.5	.55	.16	.05	.03	.27	.09	-.23	.06
	τ_{01}^2	.2	.07	.07	-.13	.02	-.02	.03	-.22	.05

par.: parameter estimate, True: simulated parameter values, EAP: expected a posteriori, MSE: mean squared error

Table 6. Simulation Study III: Latent growth parameter estimates across 50 replications under varying proportions of missing data.

IRT-based multiple imputations and assuming MAR. The missing response data were generated using the latent variable scores. When the latent scores contain bias, the imputed response data will also contain bias. The results based on the IRT scores were quite robust against violations of MAR, and the parameter estimates and posterior standard deviations are comparable across the different missing data conditions. The results are also comparable to the results presented in Table 5.

Under CTT, the bias is also stable across missing data conditions. The CTT-based estimates show more bias compared to the IRT-based estimates, but the difference is comparable across conditions. This difference was identified in the complete data situation, where it was shown that the sum scores ignore response information, which led to an underestimation of the linear trend and an overestimation of the measurement error variance. The estimated MSE did not show an increase, which also shows that the proposed missing data imputation method is robust against violations of MAR.

6 Application: Coping With Low Back Pain

To illustrate the differences in results between using IRT and CTT-generated latent variable scores as outcomes in the latent growth model for different missing data situations, a real data set was analyzed. The aim is to show that results obtained in the simulation studies also apply in real life data situations, where a longitudinal latent variable was measured on multiple occasions.

Data were obtained from a repeated measurement study on coping strategies in patients with low back pain [18, 33]. The development of the latent variable low back pain was measured using the Dutch version of the Pain Coping Inventory (PCI) [34]. The PCI subscale on passive coping was used. The questionnaire was administered during the four time-points of the original study, at baseline, after six months, and one and two years after baseline. Time was measured in months and divided by 24, which was the maximum number of study months. So, time equaled one for the final measurement occasion, and time equaled zero for the baseline measurement occasion. For the current illustration, responses of patients with at least two measurement occasions without missing item scores were taken into account. This resulted in a sample of 254 patients. The answering categories were collapsed for the analysis, resulting in dichotomous responses.

First, the complete data set was analyzed. The scores for the latent variable PCI-passive were constructed using the IRT and CTT measurement model, where the latent growth model from Equation (1) was considered, where θ_{ij} is the latent variable PCI-passive. MCMC was run using $50,000$ iterations with a burnin of $10,000$ to estimate all model parameters. A population intercept (γ_0) and a population effect for time (γ_1) were used for predicting the trend in the latent variable PCI-passive. The covariance between the random intercept and random slope (τ_{01}) was fixed to zero. Inspection of the plots of the MCMC chains for the parameters of the latent growth model showed adequate convergence.

Missing responses were generated using the conditions described in Table 2, which were also used in Simulation Study III. The missing data mechanism becomes MNAR, when more item responses were missing. In Table 2, the proportions of missing data are listed, where patients with a below-average PCI score were more likely to miss the above-average difficult items. After generating the missing data patterns, multiple imputations were generated using either the IRT-based posterior predictive distributions (Equations (6) and (3)) or the CTT-based posterior predictive distributions (Equations (12) and (9)).

In Table 7, the results for pain coping in low back pain patients with different percentages of missing data are listed under IRT and CTT-based multiple imputations.

		IRT			CTT		
	par.	EAP	SD	95%HPD	EAP	SD	95%HPD
0% Missing							
Fixed	γ_0	.44	.05	[.34;.53]	.35	.06	[.24;.45]
	γ_1	-.94	.08	[-1.07;-.81]	-.74	.07	[-.86;-.63]
Random	σ^2	.21	.02	[.18;.24]	.46	.05	[.41;.51]
	τ_0^2	.39	.06	[.30;.48]	.47	.07	[.36;.58]
	τ_1^2	.68	.14	[.50;.89]	.02	.04	[.00;.09]
20% Missing							
Fixed	γ_0	.45	.05	[.36;.53]	.35	.06	[.25;.46]
	γ_1	-.97	.08	[-1.11;-.84]	-.77	.08	[-.9;-.66]
Random	σ^2	.19	.04	[.16;.21]	.45	.05	[.40;.50]
	τ_0^2	.38	.05	[.29;.46]	.47	.08	[.37;.59]
	τ_1^2	.78	.19	[.58;.99]	.03	.04	[.00;.11]
30% Missing							
Fixed	γ_0	.43	.05	[.34;.52]	.34	.06	[.23;.45]
	γ_1	-.93	.08	[-1.06;-.80]	-.74	.10	[-.87;-.63]
Random	σ^2	.21	.02	[.18;.24]	.44	.05	[.39;.49]
	τ_0^2	.41	.06	[.32;.51]	.49	.08	[.38;.61]
	τ_1^2	.70	.15	[.51;.90]	.02	.03	[.00;.06]
40% Missing							
Fixed	γ_0	.43	.05	[.34;.52]	.33	.06	[.22;.44]
	γ_1	-.94	.07	[-1.07;-.81]	-.72	.08	[-.84;-.60]
Random	σ^2	.20	.03	[.17;.23]	.46	.04	[.41;.51]
	τ_0^2	.41	.08	[.32;.50]	.47	.06	[.36;.58]
	τ_1^2	.64	.17	[.47;.83]	.02	.04	[.00;.09]
50% Missing							
Fixed	γ_0	.38	.06	[.28;.47]	.31	.07	[.20;.42]
	γ_1	-.82	.09	[-.95;-.69]	-.68	.12	[-.81;-.56]
Random	σ^2	.23	.04	[.20;.26]	.46	.04	[.41;.52]
	τ_0^2	.45	.07	[.35;.55]	.47	.06	[.37;.58]
	τ_1^2	.61	.15	[.43;.80]	.02	.03	[.00;.07]

par.: parameter estimate, EAP: expected a posteriori

Table 7. Coping with low back pain: 254 patients were measured four times on Pain Coping using the PCI-Passive scale.

In general, the results from Table 7 are in concordance with the results from the presented simulation studies. The linear trend effect, γ_{10}, was estimated less strong, when CTT-generated latent variable scores were used. Also, the variance in trend effects, τ_{11}^2, was not even detected, and the measurement error variance was higher, when CTT scores were used. These findings are in line with the results from the presented simulation studies. The results confirm that the IRT-based imputations for missing items are preferable over the CTT-based imputations for missing items, when the missing data are MAR. Furthermore, when the missings become MNAR, reasonable estimates were found using the IRT-based imputations.

When it comes to handling missing data, the IRT model and CTT model showed to be robust against violations of the MAR assumption. The proposed posterior predictive distribution of missing responses under CTT enabled imputing missing responses, while accounting for a linear trend in the latent variable scores. With an increase in the percentage of missing data, the CTT-based results were stable and did not show an increase in bias. However, the difference in results between IRT and CTT-based multiple imputations also remained stable. When using IRT, it was shown that much more heterogeneity was detected across measurements and subjects, leading to a more steep decline of the PCI-passive measurements and more variation in individual trends, in comparison to using CTT.

7 Discussion

It was shown through a novel imputation method that the use of sum scores for measuring a latent variable, when item scores are missing, leads to severely biased results of a latent growth analysis. These differences are to be expected based on what we learned in previous studies on the differences in results of IRT and CTT based longitudinal modeling [5, 32]. In all three simulation studies it was found that trend effects were underestimated under CTT-based imputations and that the measurement error variance was overestimated as well as the individual variation in baseline measurements. This indicates that the CTT model attributes more variance to the residual variance and differences between participants are more difficult to identify, resulting in an underestimation of the overall trend effect. When using predictive mean matching to impute values in the wide data set the linear trend was not detected. This followed directly from the used imputation method, i.e. predictive mean matching, which ignored the trend in the latent variable. The CTT imputation method in the second and third simulation study performs comparable to the IRT imputation method. The differences between the parameter estimates of the LGM are stable with the increasing amounts of missing data.

Despite the fact that the CTT-based imputation method provided accurate results in comparison to the complete-data analysis, the IRT-based latent growth analysis were more accurate than the CTT-based analysis. The IRT-based analysis provided better results in all simulated missing data conditions in terms of bias and MSE. It also showed robust results in more extreme missing data conditions. The consequent underestimation of the individual variation in growth parameters indicates that the CTT model is less capable of detecting differences between individuals. The CTT model neglects information in response patterns that lead to the same score. This loss of information becomes visible in the estimated variances in individual differences in baseline measurements and trend effects. The differences are also directly comparable across models through the rescaling of the multiple imputations. This underestimation in individual variation is problematic when the focus of the study is to determine how patients are developing over time. Bias is introduced when sum-scores are used and conclusions on the patient's trajectories are to be interpreted bearing in mind this bias. The consequences for interpretation of CTT-based effect estimates should be taken into consideration and preferably IRT-based scores are used when analyzing repeatedly measured questionnaires with missing items.

The analysis of a longitudinal trial study on coping with low back pain showed that using CTT-based scores for the latent variable leads to inferior parameter estimates compared to using IRT-based scores. When using IRT-based scores, the results show more pronounced negative slope effects and more variation over intercepts and trends between patients. In the IRT-based analysis, more of the variance can be attributed to the differences between patients, which is preferable in medical and epidemiological research. Now, more variance can be explained and more precise solutions can be found when testing hypotheses on differences between patients. When missing data is present, more advanced methods are necessary to make comprehensive inferences in latent variable research with a repeated measures design. We showed that the IRT-based imputation method performs better than the CTT-based method for retrieving trend and variance estimates in repeated measurement data for different missing data conditions. These findings are in concordance with earlier studies where comparable results were obtained, when analyzing complete-data sets [5, 32]. The IRT-based imputation method proved to be an excellent procedure for handling missing data.

Based on the results from this study, we recommend the use of IRT-based multiple imputations over CTT-based multiple imputations when analyzing longitudinal questionnaire data with missing item scores.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

[1] Beck A, Ward C, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Archives of General Psychiatry. 1961;4:561–571.

[2] McCrae R, Paul TJ, Martin T. The NEO–PI–3: A More Readable Revised NEO Personality Inventory. Journal of Personality Assessment. 2005;84(3):261–270. Available from: http://dx.doi.org/10.1207/s15327752jpa8403_05.

[3] Cleeland C. Measurement of pain by subjective report. In: Chapman C, Loeser J, editors. Advances in Pain Research and Therapy. Vol. 12. Issues in Pain Measurement. New York: Raven Press; 1989. p. 391–403.

[4] Lord F, Novick M, Birnbaum A. Statistical theories of mental test scores. Addison-Wesley Publishing Company, Inc.; 1968.

[5] Gorter R, Fox JP, Twisk J. Why Item Response Theory should be used for longitudinal questionnaire data analysis in medical research. BMC Medical Research Methodology. 2015;15(55).

[6] Eekhout I, de Vet H, Twisk J, Brand J, de Boer M, Heymans M. Missing data in a multi-item instrument were best handled by multiple imputation at the item score level. Journal of clinical epidemiology. 2014 3;67(3):335–342.

[7] Mazza GL, Enders CK, Ruehlman LS. Addressing Item-Level Missing Data: A Comparison of Proration and Full Information Maximum Likelihood Estimation. Multivariate behavioral research. 2015;50(5):504–19. Available from: http://www.tandfonline.com/doi/abs/10.1080/00273171.2015.1068157#.Vtbmp-bl_O0.

[8] Schafer JL, Graham JW. Missing data: Our view of the state of the art. Psychological Methods. 2002;7(2):147–177. Available from: http://doi.apa.org/getdoi.cfm?doi=10.1037/1082-989X.7.2.147.

[9] Eekhout I, Enders C, Twisk J, de Boer M, de Vet H, Heymans M. Analyzing Incomplete Item Scores in Longitudinal Data by Including Item Score Information as Auxiliary Variables. Structural Equation Modeling: A Multidisciplinary Journal. 2015;22:588–602.

[10] Quartagno M, Carpenter JR. Multiple imputation for IPD meta-analysis: allowing for heterogeneity and studies with missing covariates. Statistics in Medicine. 2016;35(17):2938–2954.

[11] Audigier V, White IR, Jolani S, Thomas PA, Quartagno M, Carpenter J, et al. Multiple imputation for multilevel data with continuous and binary variables. ArXiv preprint. 2017;.

[12] Twisk J, De Boer M, De Vente W, Heymans M. Multiple imputation of missing values was not necessary before performing a longitudinal mixed-model analysis. Journal of Clinical Epidemiology. 2013;66(9):1022–1028. Available from: http://dx.doi.org/10.1016/j.jclinepi.2013.03.017.

[13] Jolani S, Debray T, Koffijberg H, van Buuren S, Moons K. Imputation of systematically missing predictors in an individual participant data meta-analysis: A generalized approach using MICE. Statistics in Medicine. 2015;34(11):1841–1863.

[14] Resche-Rigon M, White I. Multiple imputation by chained equations for systematically and sporadically missing multilevel data. Statistical Methods in Medical Research. 2016;.

[15] Sijtsma K, van der Ark L. Investigation and treatment of missing item scores in test and questionnaire data. Multivariate Behavioral Research. 2003;38(4):529–569. Available from: http://www.leaonline.com/doi/abs/10.1207/s15327906mbr3804_5.

[16] Enders CK. Using the expectation maximization algorithm to estimate coefficient alpha for scales with item-level missing data. Psychological Methods. 2003;8(3):322–337.

[17] Huisman M. Imputation of Missing Item Responses: Some Simple Techniques. Quality and Quantity. 2000;34:331–351.

[18] Heymans M, de Vet H, Bongers P, Knol D, Koes B, van Mechelen W. The effectiveness of high-intensity versus low-intensity back schools in an occupational setting: a pragmatic randomized controlled trial. Spine. 2006;31(10):1075–1082.

[19] Rubin D. The calculation of posterior distributions by data augmentation: Comment: A noniterative sampling/importance resampling alternative to the data augmentation. Journal of the American Statistical Association. 1987;82(398):543–546.

[20] Marsman M, Maris G, Bechger T, Glas C. What can we learn from Plausible Values? Psychometrika. 2016;Available from: http://link.springer.com/10.1007/s11336-016-9497-x.

[21] von Davier M, Gonzalez E, Mislevy R. What are plausible values and why are they useful? IERI monograph series. 2009;p. 9–36.

[22] Gorter R, Fox JP, Ter Riet G, Heymans M, Twisk J. Latent growth modeling of IRT versus CTT measured longitudinal latent variables. Statistical Methods in Medical Research. 2019;.

[23] Fox JP. Multilevel IRT using dichotomous and polytomous response data. The British journal of mathematical and statistical psychology. 2005 5;58(Pt 1):145–72.

[24] Embretson S, Reise S. Item response theory for psychologists. L.Erbaum Associates; 2000.

[25] Fox JP. Bayesian item response modeling. Theory and applications. New York: Springer; 2010.

[26] Fox JP, Mulder J, Sinharay S. Bayes factor covariance testing in item response models. psychometrika. 2017;82(4):979–1006.

[27] Robitzsch A, Grund S, Henke T. Miceadds: Some additional multiple imputation functions, especially for mice; 2016. Available from: https://cran.r-project.org/package=miceadds.

[28] Fox JP. Multilevel IRT Modeling in Practice with the package mlirt. Journal of Statistical Software. 2007;20(5).

[29] Plummer M, Best N, Cowles K, Vines K. Coda: Convergence Diagnosis and Output Analysis for MCMC. R News. 2006;6:7–11.

[30] Buuren Sv, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. Journal of Statistical Software. 2011;45(3).

[31] Huisman M. Missing data in behavioral science reserach: Investigation of a collection of data sets. Kwantitatieve methoden. 1998;57:69–93.

[32] Gorter R, Fox JP, Apeldoorn A, Twisk J. The influence of measurement model choice for randomized controlled trial results. Journal of Clinical Epidemiology. 2016;79:140–149. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0895435616301925.

[33] Heymans M, De Vet H, Knol D, Bongers P, Koes B, Mechelen WV. Workers' beliefs and expectations affect return to work over 12 months. Journal of Occupational Rehabilitation. 2006;16(4):685–695.

[34] Kraaimaat F, Evers A. Pain-coping strategies in chronic pain patients: psychometric characteristics of the pain-coping inventory (PCI). International journal of behavioral medicine. 2003;10(4):343–363.

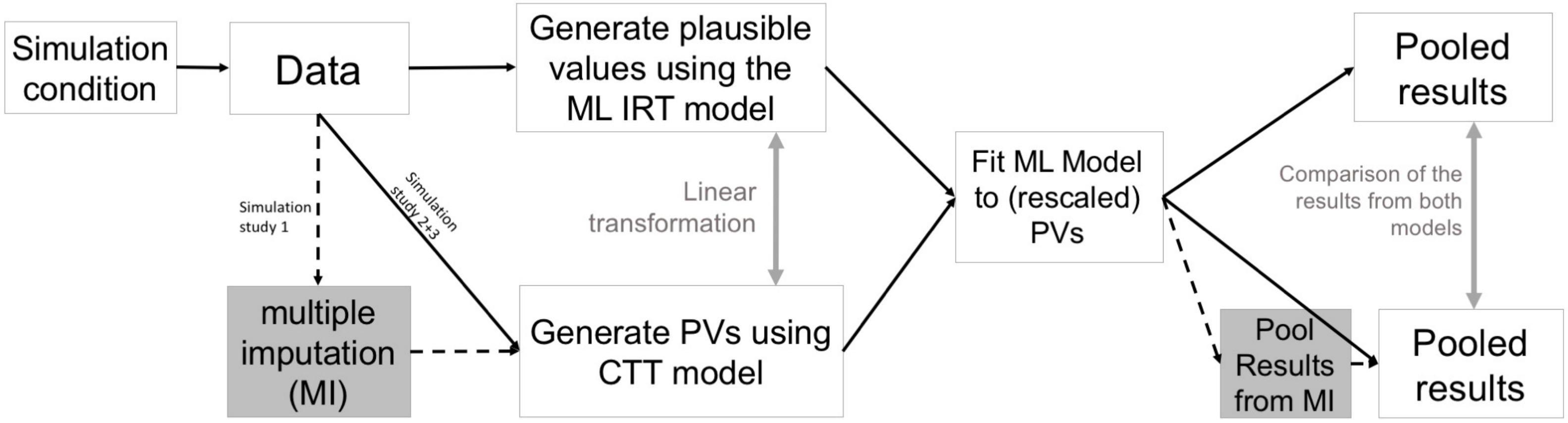

Figure 1. Procedure for the simulation studies. The dashed lines represent the route for Simulation Study I, and the solid lines represent the routes for Simulation Study II and III. PVs; Plausible Values.