Chapter 6

Including auxiliary item information to handle missing questionnaire data in two longitudinal data examples

Under review: Eekhout, I., Enders, C.K., Twisk, J.W.R., de Boer, M.R., de Vet, H.C.W., & Heymans, M.W. Including auxiliary item information to handle missing questionnaire data in two longitudinal data examples. Journal of Clinical Epidemiology.

Abstract

Previous studies show that missing values in multi-item questionnaires can best be handled at item score level. The aim of this study is to demonstrate two novel methods for dealing with incomplete item scores in outcome variables in longitudinal studies. The performance of these methods was previously examined in a simulation study. The two methods incorporate item information at the background when simultaneously the study outcomes are estimated. The investigated methods include the item scores or a summary of a parcel of available item scores as auxiliary variables, while using the total score of the multi-item questionnaire as the main focus of the analysis in a latent growth model. That way the items help estimating the incomplete information of the total scores. The methods are demonstrated in two empirical datasets. Including the item information results in more precise outcomes in terms of regression coefficient estimates and standard errors, compared to not including item information in the analysis. The inclusion of a parcel summary is an efficient method that does not over-complicate longitudinal growth estimates. Therefore it is recommended in situations where multi-item questionnaires are used as outcome measure in longitudinal clinical studies with incomplete scores due to missing item scores.

Keywords: missing data, longitudinal data, multi-item questionnaire, auxiliary variables, full information maximum likelihood, methods, latent growth modeling, structural equation modeling

Introduction

Many medical and epidemiological longitudinal studies use patient-reported outcomes such as quality of life as the main focus of their analyses. These patient-reported outcomes are often repeatedly measured by a multi-item questionnaire. The item scores of the questionnaire are summed or averaged to a total score to represent the outcome of interest. In case respondents do not fill out all the questions in a multi-item questionnaire, the calculation of the total scores is impaired. As a solution, manuals of multi-item questionnaires often advise to average over the available items (e.g., (Bracken & Howell, 2004; Lambert, Lunnen, Umphress, Hansen, & Burlingame, 1994)), otherwise known as person mean imputation. Averaging over the available items is algebraically identical to substituting a person's mean item response. This solution can result in biased analysis results, especially when data are not missing completely at random (MCAR) (Eekhout et al., 2014; Gottschall, West, & Enders, 2012). Another option for handling missing data values is to apply a complete-case analysis. In that method only respondents that have all item scores observed are included in the analysis. This method only results in unbiased analyses when data are MCAR. A complete-case analysis always results in a decreased sample size, so power will be suboptimal in all situations. Nevertheless, this method is most often applied in epidemiological studies (Eekhout, de Boer, Twisk, de Vet, & Heymans, 2012).

More advanced methods to handle missing data are multiple imputation or full information maximum likelihood (FIML). Both methods use all observed data in the analyses. In multiple imputation, the missing values are replaced by imputed values. A regression model estimates predicted scores for the incomplete values and random error, drawn from a normal distribution around the estimated value, is added to the predicted score to account for uncertainty around the imputed values. This imputation process is repeated multiple times resulting in multiple imputed datasets. Subsequently, the data analysis is performed on each of these imputed datasets. The multiple results from these datasets are pooled into one final analysis result (Rubin, 1987; Schafer, 1997; van Buuren, 2012). In FIML, missing values are not replaced or imputed; instead all available data are used to estimate the population parameters with the highest likelihood of producing the sample data. Both multiple imputation and FIML perform well when the probability of missing data is related to other variables in the data, which is known as missing at random (MAR) (Rubin, 1976). Furthermore, with these techniques model estimations are generally unbiased and without loss of power.

In a multi-item questionnaire total scores may be missing because of missing item scores. In that case, there are two main approaches to handle the missing data. Missing data can be handled at the item level or at the total score level of the multi-item questionnaire. The missings are handled at the item level when a missing data method is applied to the incomplete item scores first and then the total scores are calculated (e.g., by summing imputed item responses) and used for the analysis. Handling the missings at the total score level means that the total scores will be incomplete when one or more item scores are missing. The missing data handling method is applied to these total scores directly. Previous studies have shown that it is most beneficial to handle the missing data in a multi-item questionnaire at the item level. Handling missing item scores at the item level improves precision (Eekhout et al., 2014; Gottschall et al., 2012). In the context of multiple imputation it is quite straightforward to handle the missings at the item level. The item scores are imputed in the imputation model, and after the imputation part, the item scores are summed to the total scores in each of the imputed datasets, which are used for the analysis. However, when the number of items is very large, for example in longitudinal studies where item scores from multiple time points are included in the analysis, multiple imputation of the item scores might cause complications. When the number of items in the study gets close to the sample size, there is not enough information in the data to estimate the imputation model parameters. For example in a study where a multi-item questionnaire with 20 items is measured at six time points, the total number of variables in an imputation model would be at least 120. Green (1991) described a rule of thumb where the sample size should be larger than 53+k to do a regression analysis for a medium effect size (i.e., 0.13), where k is the number of predictors. In the example we outline below with 120 variables, the minimum sample size should then be 173. Hence, the number of variables in an imputation model could easily exceed the maximum allowed number in a longitudinal study with many time-points and a multi-item questionnaire as outcome measure. Moreover, when outcomes are measured at multiple time-points in a longitudinal study it might be feasible to analyze the data with a longitudinal analysis method such as a latent growth model. Usually these models are estimated with FIML, which produces unbiased model estimates when missing outcomes are missing at random. Nevertheless, the item scores are generally not included in such an analysis, because mostly only the total scores are modeled. Ergo, growth models estimated by FIML encourage users to deal with missing data at the scale level rather than the item level. Since, as previously mentioned, it is better to handle missings in a multi-item questionnaire at the item level, it would be beneficial to include the item scores in the analysis as well.

In a previous simulation study, we investigated two novel methods for including item-level information in a latent growth model while still focusing on change at the scale score level; the purpose of that study was to outline a FIML analog to item-level imputation (Eekhout et al., in press). We showed that these methods yield valid and precise parameter estimates in a latent growth model when total scores were missing due to missing item scores. In that study, the item information was included in the model as auxiliary variables using well-established methods outlined by Graham (2003). Auxiliary variables are variables that are used to include extra information about the missingness of the data. They are related to the missingness in the data and/or are correlated with the incomplete variables. Including these variables in a missing data analysis will reduce bias and improve precision lost due to missing data (Collins, Schafer, & Kam, 2001). Auxiliary variable techniques are usually employed to incorporate predictors of missingness, thereby increasing the plausibility of the MAR assumption. We use these techniques to incorporate item-level information into a total score analysis. In this paper we will explain and demonstrate these two methods which include different item score information as auxiliary variables in a longitudinal study by using a latent growth model in two data examples.

Methods

Data examples

Data example 1 is a dataset from a study where the longitudinal effects of a randomized controlled trial were analyzed in which three treatments for neck pain were compared: manual therapy (specific mobilization techniques), physical therapy (exercise therapy), and usual care (analgesics, counseling and education) (Hoving et al., 2002). The main outcomes in the study were global perceived recovery, physical functioning, pain intensity, and neck disability. One of the secondary outcomes in the study was physical functioning. Physical functioning was measured by the physical functioning scale of the SF-36, which contains 10 items measured at a three point likert scale (Ware, Kosinski, & Keller, 1994). The item scores can be summed to obtain a total score for physical functioning. The outcome was measured at baseline and after 3, and 7 weeks. For this example we used the multi-item data from the physical function scale and included the treatment variable as central independent variable and age as a covariate in a latent growth analysis. In this dataset 170 out of 183 participants had completely observed data. We generated missing values in the items of the outcome measures in order to create situations for which we could compare the models that include the item information as auxiliary variables with the model without this auxiliary information. We used the 170 cases with complete data as a reference. In a copy of this dataset, missings were generated on the item-level of the SF-36 subscale for Physical Functioning by using the treatment variable and the age variable as predictors for missingness. That way the missing data on the items was missing at random. The baseline wave was complete. For the measurements at three and seven weeks about half of the items had about 15% missing data.

Data example 2 is from a randomized controlled trial about low back pain. The study population consisted of 299 workers that were listed as sick for a period of three weeks due to low back pain. Three treatment groups were compared in a randomized controlled trial. The treatments were high-intensity and low-intensity back schools compared to the usual treatment by the occupational physician. The outcomes were measured at baseline, after three and after six months and were days until return to work, days of sick-leave, pain, functional disability, kinesiophobia, and perceived recovery. The results for the treatment effects for the main outcomes were published previously (Heymans et al., 2006). For this example we used the data from the passive coping scale of the Perceived Coping Inventory as the outcome which was also measured at the three time points (Kraaimaat & Evers, 2003). This subscale contains 21 items measured on a four point Likert scale. Data example 2 is a dataset that already contained missing data. The missing data in this dataset was mostly due to participants that missed an entire wave. At baseline 4% of the participants didn't return the questionnaire, at wave 2 26% and at wave 3 30%. We generated additional missing values for the item scores to present a data situation with missing total scores due to item scores as well as missings caused by participants not returning the questionnaire. The resulting overall average percentage of missing item scores was 25%.

In summary, in data example 1 we only generated incidental missing item scores and in data example 2 we present a situation where missing data on the total scores were caused by both the item score missings and by participants missing entire measurement waves. Additionally, the data in data example 2 also contains missing data for the baseline measurement. Both missing data situations are realistic and common in epidemiological studies. Furthermore, the number of items per scale of data example 2 is twice as high as the number of items per scale in data example 1.

Full Information Maximum Likelihood analyses

The data for both examples were analyzed by a latent growth model estimated with FIML. In a latent growth model the change in total scores over time is modeled, where the individual growth of each case in the study can be treated as a random effect. That way the variance between persons is taken into account, because person A might have a different development over time than person B. So the intercept and slope coefficients may vary across individuals, and are therefore referred to as random effects, or latent growth factors (Kwok et al., 2008). In models that use questionnaire total scores as the outcome, the total scores are computed prior to including them in the analysis. The total score is only computed when all items are observed. When some or all items are missing, the total score is missing. So for each wave the observed item scores are ignored for the cases with incomplete item scores.

In order to examine the change in physical functioning over the three time-points for data example 1 we used the model of Figure 6.1. The factor loadings of the latent intercept were fixed at 1 and the factor loadings of the slope factor were set at the time-scores, which were 0, 3, and 7. The age and treatment covariates were included in the model. The three treatment categories were included as two dummy variables in order to distinguish between the effects of each treatment. The total scores are the sums of the item scores at each measurement wave. The estimates from the reference dataset with the 170 participants with complete observations were compared to the model estimates of the dataset with incomplete total scores due to the generated missing item scores.

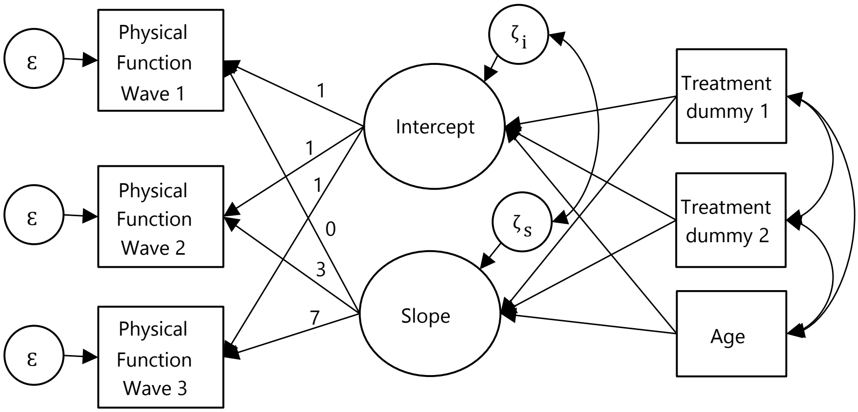

Figure 6.1. Latent growth model diagram for example data 1. Treatment dummy 1 denotes physical therapy versus continued care by a general practitioner; treatment dummy 2 denotes manual therapy versus continued care by a general practitioner.

For data example 2, the latent growth model presented in Figure 6.2 was fitted to measure the change in the passive coping score. The factor loadings for the growth factor were 0 for baseline and 3 and 6 for the follow-up waves. The loadings for the intercept factor were fixed at 1. The treatment variable was included as a dummy variable in the model in order to distinguish between the effects of the separate treatments. The total scores for passive coping are the sum of the item scores for each measurement wave and these were incomplete when one or more items were missing.

For each model we compared estimates for the average baseline score for the control group (intercept latent mean), the average difference at baseline for each treatment group relative to the control group (intercept on treatment), the average growth of the control group (slope latent mean) and the difference of linear growth for each treatment group relative to the control condition (slope on treatment).

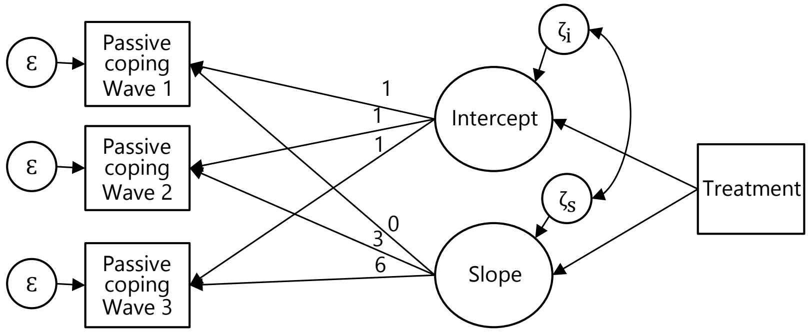

Figure 6.2. Latent growth model diagram for example data 2. Treatment dummy 1 denotes low-intensity treatment versus usual care; treatment dummy 2 denotes high-intensity care versus usual care.

Including the item information as auxiliary variables

Usually, for each wave only the cases with completely observed item data are included in the latent growth analyses (as in the analysis described above), since only for those cases the total score can be computed. This leads to a decreased precision of estimates, because less than an optimal amount of information is included. Furthermore, the scale scores at different waves could have different missingness rates. In order to improve estimates the item information was included as auxiliary variables in the models for data in the datasets with missing values. Graham (2003) described a method to include auxiliary variables that can be applied to structural equation models that use latent variables. The auxiliary variables should be (a) correlated to the manifest independent variables, (b) correlated to the residuals of all manifest endogenous variables (e.g., repeated measured scale scores); and (c) correlated with each other. The item scores would be ideal candidates as auxiliary variables, since the item scores are related to the scale scores and to the missingness on the scale scores as well. Accordingly we can include item scores by (a) correlating them in the model to the independent variables (e.g., treatment and age from data example 1), and (b) to the residuals of the scale scores from each wave and (c) correlating them with each other.

In Figure 6.3 an example of two auxiliary item scores included in the model from data example 1 is displayed. Item information can be included in the model by two methods: using the item scores as auxiliary variables or using a parcel summary score of the items as auxiliary variables (Eekhout et al., in press). For the method where the item scores were used, we included the observed item scores for each time-point in the auxiliary part of the model. That way, additional to the main model information, also the information from the items is used to estimate the most likely model parameters. It would be most ideal to include as many item scores as possible while still reaching convergence of the model. The process of obtaining the full information maximum likelihood estimates is called convergence. Convergence problems can be related to the fact that the auxiliary part of the model is too similar to the total score outcomes, i.e., collinearity. In that case some extra noise should be added by removing some items, minimally one item per wave with the most missing data. Another reason for a lack of convergence might be that the number of correlations that have to be estimated in the model exceeds the sample size. For example when number of included item scores is large in longer questionnaires measured at many time-points. Therefore a second method that includes a summary of the item scores or a summary of a parcel of the item scores as auxiliary variables can be used. An example of such a summary is the average over the available items. That way the item information is included in the model without over-complicating the model estimation process. Our rationale for using a parcel summary score, is that the average of available items can capture most of the available information in the observed items while dramatically reducing the number of parameters (Enders, 2008). The average over the available items proved to be a valid and efficient method to include the item information (Eekhout et al., in press). Both of these methods accomplish the same end-point, which is smaller standard errors and therefore increase the precision of model estimates.

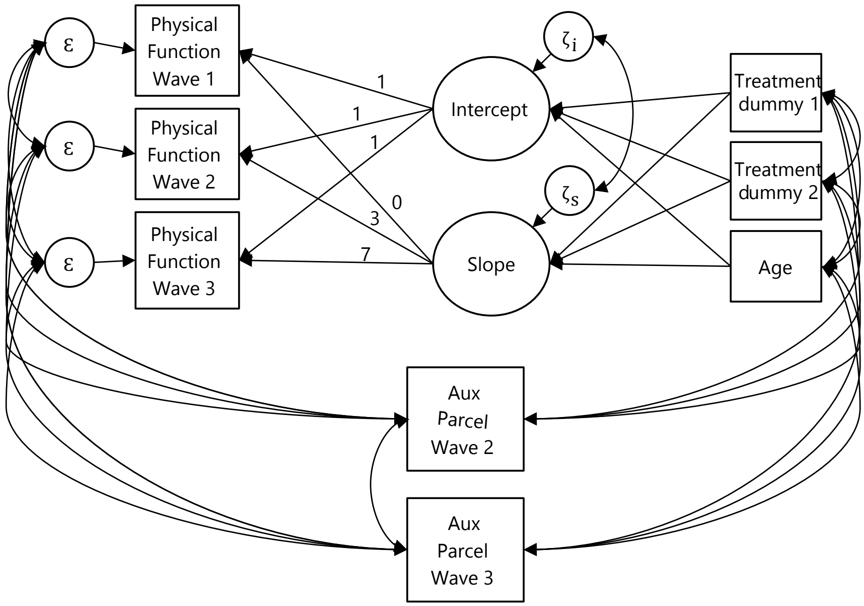

Figure 6.3. Auxiliary variables included in the latent growth model of data example 1. Treatment dummy 1 denotes physical therapy versus continued care by a general practitioner; treatment dummy 2 denotes manual therapy versus continued care by a general practitioner

For the first method that includes the item information it would be desirable to include 50% or more of the items as auxiliary variables. In data example 1, we included all but one items per wave with missing data. For data example 2, there was also some missing data on the baseline wave. So we also included the items from that wave in the auxiliary part of the model. Including all but one items per wave in the model could not be estimated, so we included 17 out of the 21 items per wave. We included the items with the lowest percentages of missing values.

For the second method, using a parcel summary of the items in order to include the item information, it is again desirable to include at least 50% of the items in the parcel summary. The parcel summary score was computed for each wave by taking the average value of the available items in the parcel. For data example 1, the parcel summary score of all but one item for wave 2 and 3 were included in the parcel. For the data example 2, the parcel summary scores for all but two items were used for all waves, because in this example also the baseline contained missing data. We excluded two items per wave, because the parcel summary scores of all but one item were too similar to the total score outcomes in the model and therefore caused computational problems. For each wave, we excluded the two items with the most missing data.

In summary, for each data example we compared two procedures that include the auxiliary item information. The first method is the inclusion of the item scores separately and the second is the summary scores of the items. In data example 1 these procedures were compared to the reference results from the complete data and to the results from a model on the incomplete data without auxiliary variables included. In data example 2 we compared results from the methods with auxiliary variables to the results from the model in the incomplete data without auxiliary variables included. All models were estimated by full information maximum likelihood in Mplus (Muthén, Asparouhov, Hunter, & Leuchter, 2010). A detailed manual on how to apply these methods in Mplus is available from the first author upon request; the Mplus syntaxes for the two example datasets are presented in the Appendix 6.1 and 6.2.

Results

For all models of data example 1 the parameter estimates are presented in Table 6.1. In the first column the results of the complete data analysis are presented as a reference. The estimates of the incomplete data from the model without auxiliary variables show that the standard errors for the slope parameters are increased compared to the results from the complete data. This is what was expected from the results of the simulation study that we previously performed (Eekhout et al., in press). When the item scores were used as auxiliary variables, the increase in standard errors relative to the complete data model results was minimal. The same can be observed in the estimates from the model where the parcel summary scores were included. By computing the ratio of the squared standard errors for the model without auxiliary variables relative to the model with the auxiliary item information, we can compare the precision different on the sample size metric. Accordingly, for the slope on manual therapy parameter this ratio is: $0.117^2/0.084^2 = 1.94$, which means that the model without auxiliary variables would require a 94% increase in the sample size to achieve the same precision as FIML with auxiliary variables.

Table 6.1.
Coefficient and standard error estimates for the compared methods for data example 1

Parameter	Complete data	No auxiliary	Item scores	Parcel summary
Intercept latent mean	25.678(1.108)*	24.684(1.115)*	24.846(1.110)*	24.897(1.116)*
Intercept physical therapy[a]	-0.131(0.597)	-0.029(0.600)	-0.183(0.598)	-0.225(0.602)
Intercept manual therapy[b]	-0.533(0.598)	-0.503(0.600)	-0.598(0.598)	-0.635(0.602)
Slope latent mean	0.281(0.153)	0.160(0.215)	0.272(0.156)	0.357(0.157)
Slope on physical therapy[a]	-0.007(0.083)	0.076(0.116)	-0.013(0.084)	-0.058(0.084)
Slope on manual therapy[b]	0.136(0.083)	0.191(0.117)	0.137(0.084)	0.084(0.084)

Note: [a] *The dummy that reflects physical therapy versus continued care by a general practitioner;* [b] *The dummy that reflects manual therapy versus continued care by a general practitioner; *P < 0.01.*

For data example 2 the results for the models on the original data of are presented in Table 6.2. In dataset, 20 subjects had missings on all repeated measurements and these were excluded for the model without auxiliary variables model. For the models with auxiliary variables, 16 cases had some observed items which were included in the auxiliary part of the model.

Table 6.2.
Coefficient and standard error estimates for the compared methods for data example 2

Parameter	No auxiliary	Item scores	Parcel summary
Intercept latent mean	44.214(0.982)*	44.054(0.927)*	44.069(0.922)*
Intercept low-intensity[a]	1.864(1.417)	2.091(1.338)	1.628(1.329)
Intercept high-intensity[b]	0.832(1.407)	0.692(1.341)	0.540(1.335)
Slope latent mean	-1.355(0.213)*	-1.020(0.170)*	-1.116(0.164)*
Slope on low-intensity[a]	-0.491(0.312)	-0.896(0.242)*	-0.674(0.233)*
Slope on high intensity[b]	-0.031(0.306)	-0.064(0.244)	0.047(0.236)

Note: [a] *The dummy that reflects low-intensity back school treatment versus usual care;* **[b]** *The dummy that reflects high-intensity back school treatment versus usual care by an occupational therapist; *P<0.01.*

In the results from Table 6.2 we can observe a gain in precision reflected in the smaller standard errors for the models with items scores and parcel summary scores included in as auxiliary variables, this gain was most apparent for the slope parameters. In addition, the regression coefficients show a stronger effect. The model without auxiliary variables did not present any significant treatment effect on passive coping. However, the models that included the auxiliary item information (i.e., item scores or parcel summary scores) showed a significant slope for low-intensity back school treatment. In this example we can see how improved methods can actually affect study conclusions. As for the previous example, we can calculate the effect of the precision gain also by putting the standard error differences on the sample size metric. For the slope on low-intensity treatment this ratio would be 0.3122/0.2422=1.66 and 0.3122/0.2332=1.79 for including item scores or a parcel summary of the items respectively. These ratios imply a required increase in sample size for a model without auxiliary variables of 66% and 79% to reach the same precision as in the models with auxiliary item information.

Discussion

In this paper we presented two examples of longitudinal data analyses with a growth model when total scores are missing due to missing item scores. The compared models that include auxiliary item information improve the precision of the growth estimates which is important to correctly estimate a treatment effect. The level of precision that was obtained in the models that include the auxiliary item information can only be obtained in a model without auxiliary variables by increasing the sample size substantively. As was shown in the examples the required increase in sample size to reach the same level of precision can be as high as 94% (i.e., doubling the sample size). Furthermore, in data example 2 we showed that smaller standard errors caused by the auxiliary item information resulted in a significant treatment effect for the low-intensity back school. Especially in such clinical research situation it is important to estimate a model with optimal precision.

We presented two different methods to include item information in the auxiliary part of a latent growth model. In the first method the item scores were included separately and the most optimal amount of information is included in the model. However, the amount of correlations that have to be estimated in such a model can become problematically large. For that reason we also presented a method where a parcel summary score of the items is included. Including the item scores separately or including a parcel summary score of the item scores both performed well and improve precision. However, the model with the parcel summary score is easier to estimate and is therefore advised.

A previous simulation study of our group showed the performance with respect to bias (i.e., better coefficient estimates) and precision (i.e., smaller standard errors) in many longitudinal data situations (Eekhout et al., in press). Thought the bias was minimal for all tested FIML models, the effect on precision was substantial; for the models that did not include auxiliary item information, sample sizes should nearly be doubled to achieve the same level of precision as in the models with a parcel summary score of the items. The results from the data example 1 in the current study were compared with a complete data situation, so in that case we have a true reference situation to show that the inclusion of the item score information does not change the model interpretation, but improves the growth estimates in the model. This example contained incomplete total scores due to missing item data. Data example 2 presents a situation where missing total scores result from incidental missing item scores but also from participants that did not return the questionnaire. Also in that missing data situation the inclusion of item information in the auxiliary part of the model is beneficial. Furthermore, when data are missing at the baseline wave, cases are excluded from analyses as the 20 subjects in data example 2. By including the auxiliary item information for the cases that have observed item scores available, more cases are part of the analysis.

Strengths and limitations

This study shows the performance of including item information in the auxiliary part of a latent growth model to improve precision of parameter estimates in an empirical longitudinal dataset. The applications of our methods to empirical data correspond to the results from a previous simulation study. The presentation of data example 2 showed that the improvement of precision and accuracy of parameter estimates can be crucial in some data situations.

For both example datasets we generated extra missing data at the item level. Many epidemiological studies encounter missing data problems and when multi-questionnaires are used the missing data often occurs at the item level. By generating situations with extra missing data at the item level we can present the robustness of the methods we propose in such realistic missing item data situations.

The parameter estimates presented for the data examples sometimes seem to vary a little across the methods. For example, in data example 1 the slope on manual therapy estimate varied between 0.136 for the complete data to 0.191 in the model without auxiliary variables and 0.137 and 0.084 in the models with auxiliary item information. However, in our simulation study we found that the data with missing total scores analyzed in a latent growth model estimated by FIML does not bias parameter estimates (Eekhout et al., in press). Furthermore, the parameter estimates for intercept and slope vary together. For example, the intercept and slope on manual therapy for complete data in data example 1 were -0.533 and 0.136 and for the incomplete data model without auxiliary variables the intercept and slope were -0.503 and 0.191, so in the complete data the intercept is a bit further from zero but the slope is closer to zero and the opposite is true for the model on the incomplete data.

The datasets we used in the examples were chosen to demonstrate our methods in different situations. Data example 1 includes a questionnaire of 10 items, while data example 2 includes a questionnaire with 21 items. For the method where the item scores are included separately in the auxiliary part of the model, all but one item is most optimal. However, when questionnaires contain many items, including all but one item score per wave can cause computational difficulties. In data example 2, we included 17 items per wave in the auxiliary part of the model. Nevertheless, this model still improved the precision of the estimates compared to not including auxiliary item information. This showed that even including a smaller part of the items can be very beneficial.

This paper aimed to explain how the inclusion of item information in the auxiliary part of a latent growth model works and to show the feasibility and the effects of the inclusion of item information in empirical longitudinal data. As previously mentioned, it is most feasible to include as much information as possible in the auxiliary part of the model. This is also applicable for the composition of the parcel summary score. The performance of different compositions of parcel scores should be further explored in a simulation study. In a small simulation previously conducted (data not shown) we found that using 50% of the available item scores already improves estimates. The item scores can then be included either separately or as a summary score. However, the most optimal number of items relative to scale length or number of repeated measures was not explored extensively yet, but can be studied in future research.

References

Bracken, B. A., & Howell, K. (2004). Clinical Assessment of Depression: Professional manual. Odessa, FL: Psychological Assessment Resources.

van Buuren, S. (2012). Flexible Imputation of Missing data. New York: Chapman & Hall/CRC.

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. Psychological Methods, 6(4), 330-351.

Eekhout, I., de Boer, R. M., Twisk, J. W., de Vet, H. C., & Heymans, M. W. (2012). Missing data: a systematic review of how they are reported and handled. Epidemiology, 23(5), 729-732.

Eekhout, I., De Vet, H. C. W., Twisk, J. W. R., Brand, J. P. L., De Boer, M. R., & Heymans, M. W. (2014). Missing data in a multi-item instrument were best handled by multiple imputation at the item score level. Journal of Clinical Epidemiology, 67(3), 335-342.

Eekhout, I., Enders, C. K., Twisk, J. W. R., De Boer, M. R., de Vet, H. C. W., & Heymans, M. W. (in press). Analyzing Incomplete Item Scores in Longitudinal Data by Including Item Score Information as Auxiliary Variables. Structural Equation Modeling: A Multidisciplinary Journal.

Enders, C. K. (2008). A Note on the Use of Missing Auxiliary Variables in Full Information Maximum Likelihood-Based Structural Equation Models. Structural Equation Modeling: A Multidisciplinary Journal, 15(3), 434-448.

Gottschall, A. C., West, S. G., & Enders, C. K. (2012). A Comparison of Item-Level and Scale-Level Multiple Imputation for Questionnaire Batteries. Multivariate Behavioral Research, 47(1), 1-25.

Graham, J. W. (2003). Adding Missing-Data-Relevant Variables to FIML-Based Structural Equation Models. Structural Equation Modeling: A Multidisciplinary Journal, 10(1), 80-100.

Green, S. B. (1991). How Many Subjects Does It Take To Do A Regression Analysis. Multivariate Behavioral Research, 26(3), 499-510.

Heymans, M. W., de Vet, H. C., Bongers, P. M., Knol, D. L., Koes, B. W., & van Mechelen, W. (2006). The effectiveness of high-intensity versus low-intensity back schools in an occupational setting: a pragmatic randomized controlled trial. Spine (Phila Pa 1976), 31(10), 1075-1082.

Hoving, J. L., Koes, B. W., de Vet, H. C., van der Windt, D. A., Assendelft, W. J., van Mameren, H., . . . Bouter, L. M. (2002). Manual therapy, physical therapy, or continued care by a general practitioner for patients with neck pain. A randomized, controlled trial. Ann Intern Med, 136(10), 713-722.

Kraaimaat, F. W., & Evers, A. W. (2003). Pain-coping strategies in chronic pain patients: psychometric characteristics of the pain-coping inventory (PCI). Int J Behav Med, 10(4), 343-363.

Kwok, O. M., Underhill, A. T., Berry, J. W., Luo, W., Elliott, T. R., & Yoon, M. (2008). Analyzing Longitudinal Data with Multilevel Models: An Example with Individuals Living with Lower Extremity Intra-articular Fractures. Rehabil Psychol, 53(3), 370-386.

Lambert, M. J., Lunnen, K., Umphress, V., Hansen, N., & Burlingame, G. M. (1994). Administration and scoring manual for the Outcome Questionnaire (OQ-45.1). Salt Lake City: IHC Center for Behavioral Healthcare Efficacy.

Muthén, B. O., Asparouhov, T., Hunter, A., & Leuchter, A. (2010). Growth Modeling with Non-Ignorable Dropout: alternative analysis of the STAR*D antidepressant trial. February version wide.

Rubin, D. B. (1976). Inference and missing data. Biometrika, 63(3), 581-592.

Rubin, D. B. (1987). Multiple Imputation for Nonresponse in Surveys. New York: J. Wiley & Sons.

Schafer, J. L. (1997). Analysis of Incomplete Multivariate Data. London, UK: Chapman & Hall.

Ware, J. E., Jr., Kosinski, M., & Keller, S. D. (1994). SF-36 Physical and Mental Health Summary Scales: A User's Manual. Boston, MA: Health Assessment Lab.

Appendix 6.1|Mplus syntaxes example data 1

Growth model without auxiliary variables

```
Mplus VERSION 7.11
MUTHEN & MUTHEN

INPUT INSTRUCTIONS
data:
!file =’C:\filelocation\dataset1_complete.dat’;
file =’C:\filelocation\dataset1_missing.dat’;
variable:
names =
y1_1 - y1_10
y2_1 - y2_10
y3_1 - y3_10
tx age;
usevariables = age dummy1 dummy2 scale1 scale2 scale3;
missing = all(-9);
define:
!define dummies for the treatment groups;
IF (tx==1) THEN dummy1=0;
IF (tx==2) THEN dummy1=1;
IF (tx==3) THEN dummy1=0;
IF (tx==1) THEN dummy2=1;
IF (tx==2) THEN dummy2=0;
IF (tx==3) THEN dummy2=0;
! calculation of scale scores;
scale1 = sum(y1_1-y1_10);
scale2 = sum(y2_1-y2_10);
scale3 = sum(y3_1-y3_10);
model:
i s | scale1@0 scale2@3 scale3@7;
i on age
dummy1
dummy2;
s on age
dummy1
dummy2;
i with s;
[i];
[s];
i;
s;
scale1 - scale3 (resvar);
```

Growth model with item scores as auxiliary variables

```
Mplus VERSION 7.11
MUTHEN & MUTHEN

INPUT INSTRUCTIONS
data:
file ='C:\filelocation\dataset1_missing.dat';
variable:
names =
y1_1 - y1_10
y2_1 - y2_10
y3_1 - y3_10
tx age;
usevariables = age dummy1 dummy2 scale1 - scale3;
missing = all(-9);
! including the items as auxiliary variables;
auxiliary = (m) y2_2-y2_10
y3_1-y3_3 y3_5-y3_10;
define:
!define dummies for the treatment groups;
IF (tx==1) THEN dummy1=0;
IF (tx==2) THEN dummy1=1;
IF (tx==3) THEN dummy1=0;
IF (tx==1) THEN dummy2=1;
IF (tx==2) THEN dummy2=0;
IF (tx==3) THEN dummy2=0;
! calculation of scale scores;
scale1 = sum(y1_1-y1_10);
scale2 = sum(y2_1-y2_10);
scale3 = sum(y3_1-y3_10);
model:
i s | scale1@0 scale2@3 scale3@7;
i on age
dummy1
dummy2;
s on age
dummy1
dummy2;
i with s;
[i];
[s];
i;
s;
scale1 - scale3(resvar);
```

Growth model with the parcel scores as auxiliary variables

```
Mplus VERSION 7.11
MUTHEN & MUTHEN

INPUT INSTRUCTIONS
data:
file ='C:\filelocation\dataset1_missing.dat';
variable:
names =
y1_1 - y1_10
y2_1 - y2_10
y3_1 - y3_10
tx age;
usevariables = age dummy1 dummy2 scale1-scale3 parcel2-parcel3;
missing = all(-9);
! including the parcel scores as auxiliary variables;
auxiliary = (m) parcel2 - parcel3;
define:
!define dummies for the treatment groups;
IF (tx==1) THEN dummy1=0;
IF (tx==2) THEN dummy1=1;
IF (tx==3) THEN dummy1=0;
IF (tx==1) THEN dummy2=1;
IF (tx==2) THEN dummy2=0;
IF (tx==3) THEN dummy2=0;
! calculation of scale scores;
scale1 = sum(y1_1-y1_10);
scale2 = sum(y2_1-y2_10);
scale3 = sum(y3_1-y3_10);
! calculation of the parcel summary scores;
parcel2 = mean(y2_2-y2_10);
parcel3 = mean(y3_1-y3_3 y3_5-y3_10);
model:
i s | scale1@0 scale2@3 scale3@7;
i on age
dummy1
dummy2;
s on age
dummy1
dummy2;
i with s;
[i];
[s];
i;
s;
scale1 - scale3 (resvar);
```

Appendix 6.2|Mplus syntaxes example data 2

Growth model without auxiliary variables

```
Mplus VERSION 7.11
MUTHEN & MUTHEN

INPUT INSTRUCTIONS
data:
file ='C:\filelocation\dataset2_missing.dat';
variable:
names =
y1_1 - y1_21
y2_1 - y2_21
y3_1 - y3_21
tx age;
usevariables = age dummy1 dummy2 scale1 scale2 scale3;
missing = all(-9);
define:
!define dummies for the treatment groups;
IF (tx==0) THEN dummy1=0;
IF (tx==1) THEN dummy1=1;
IF (tx==2) THEN dummy1=0;
IF (tx==0) THEN dummy2=0;
IF (tx==1) THEN dummy2=0;
IF (tx==2) THEN dummy2=1;
! calculation of scale scores;
scale1 = sum(y1_1-y1_21);
scale2 = sum(y2_1-y2_21);
scale3 = sum(y3_1-y3_21);
model:
i s | scale1@0 scale2@3 scale3@6;
i on dummy1
dummy2;
s on dummy1
dummy2;
i with s;
[i];
[s];
i;
s;
scale1 - scale3 (resvar);
```

Growth model with item scores as auxiliary variables

```
Mplus VERSION 7.11
MUTHEN & MUTHEN

INPUT INSTRUCTIONS
data:
file ='C:\filelocation\dataset2_missing.dat';
variable:
names =
y1_1 - y1_21
y2_1 - y2_21
y3_1 - y3_21
tx age;
usevariables = age dummy1 dummy2 scale1 - scale3;
missing = all(-9);
! including the items as auxiliary variables;
auxiliary = (m) y1_1-1_2 y1_6-y1_9 y1_11-y1_21
y2_1-y2_4 y2_6-y2_11 y2_13-y2_15 y2_17 y2_19-y2_21
y3_1-y3_3 y3_5-y3_12 y3_14 y3_16-y3_18 y3_20 y3_21;
define:
!define dummies for the treatment groups;
IF (tx==0) THEN dummy1=0;
IF (tx==1) THEN dummy1=1;
IF (tx==2) THEN dummy1=0;
IF (tx==0) THEN dummy2=0;
IF (tx==1) THEN dummy2=0;
IF (tx==2) THEN dummy2=1;
! calculation of scale scores;
scale1 = sum(y1_1-y1_21);
scale2 = sum(y2_1-y2_21);
scale3 = sum(y3_1-y3_21);
model:
i s | scale1@0 scale2@3 scale3@6;
i on dummy1
dummy2;
s on dummy1
dummy2;
i with s;
[i];
[s];
i;
s;
scale1 - scale3(resvar);
```

Growth model with the parcel scores as auxiliary variables

```
Mplus VERSION 7.11
MUTHEN & MUTHEN

INPUT INSTRUCTIONS
data:
file ='C:\filelocation\dataset2_missing.dat';
variable:
names =
y1_1 - y1_21
y2_1 - y2_21
y3_1 - y3_21
tx age;
usevariables = age dummy1 dummy2 scale1-scale3 parcel1-parcel3;
missing = all(-9);
! including the parcel scores as auxiliary variables;
auxiliary = (m) parcel1 - parcel3;
define:
!define dummies for the treatment groups;
IF (tx==0) THEN dummy1=0;
IF (tx==1) THEN dummy1=1;
IF (tx==2) THEN dummy1=0;
IF (tx==0) THEN dummy2=0;
IF (tx==1) THEN dummy2=0;
IF (tx==2) THEN dummy2=1;
! calculation of scale scores;
scale1 = sum(y1_1-y1_21);
scale2 = sum(y2_1-y2_21);
scale3 = sum(y3_1-y3_21);
! calculation of the parcel summary scores;
Parcel1 = mean(y1_1-1_2 y1_4-y1_9 y1_11-y1_21);
parcel2 = mean(y2_1-y2_11 y2_13-y2_15 y2_17-y2_21);
parcel3 = mean(y3_1-y3_12 y3_14-y3_18 y3_20 y3_21);
model:
i s | scale1@0 scale2@3 scale3@6;
i on dummy1
dummy2;
s on dummy1
dummy2;
i with s;
[i];
[s];
i;
s;
scale1 - scale3 (resvar);
```