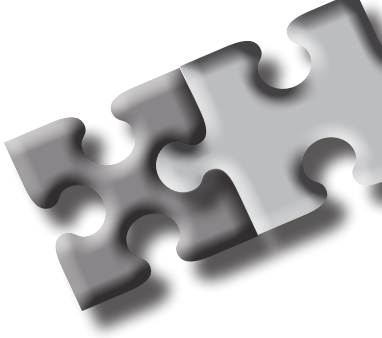

Chapter 8

General discussion

Under review as discussion section of a review article: Eekhout, I., de Vet, H.C.W., de Boer, M.R., Twisk, J.W.R., Heymans, M.W. Missing data in multi-item questionnaires: analyze carefully and don't waste available information. International Journal of Epidemiology.

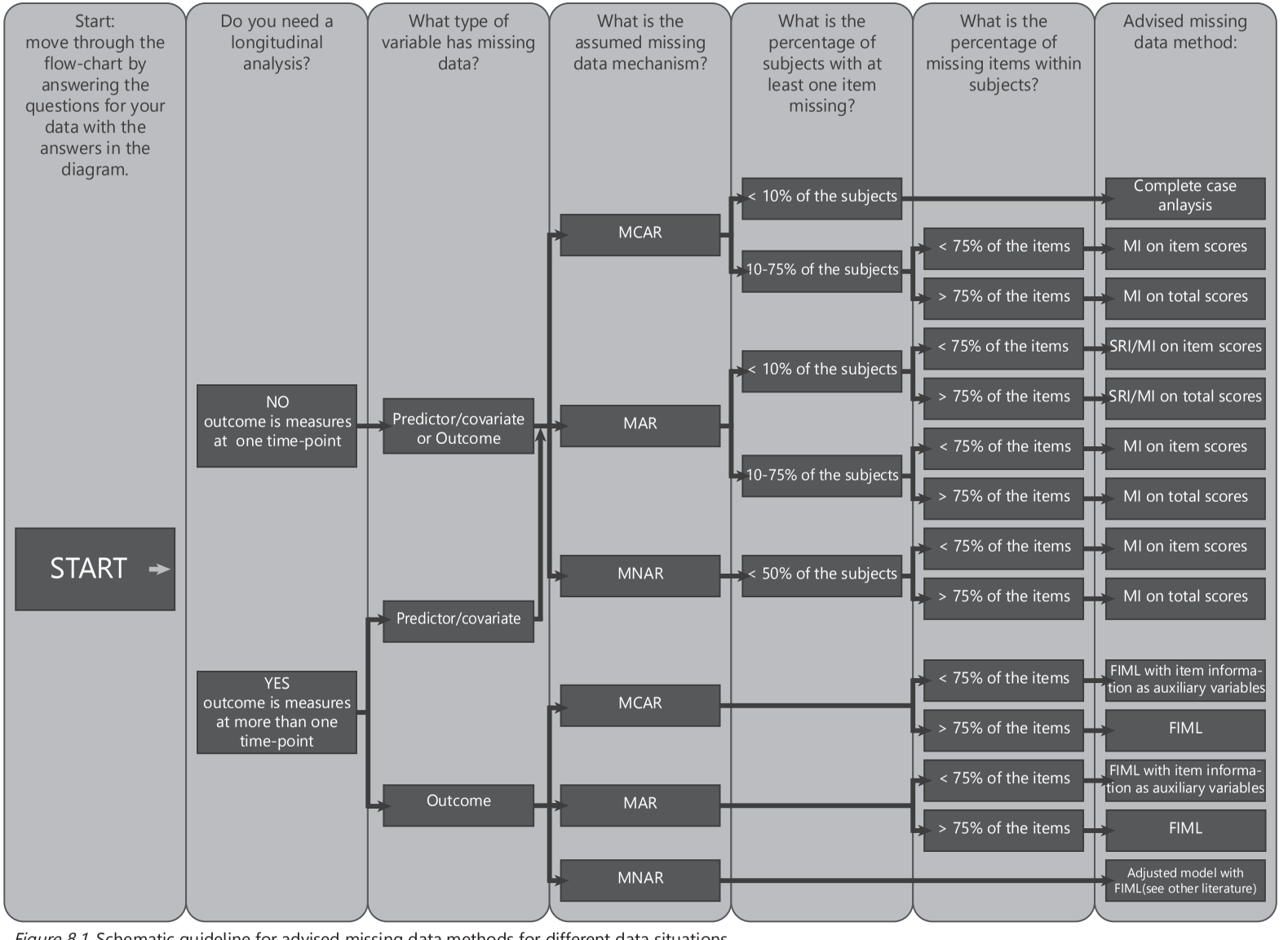

Figure 8.1. Schematic guideline for advised missing data methods for different data situations.

General advice

The best method for dealing with missing data in multi-item questionnaires depends on many aspects, for example the study design, the analysis method and type of the missing data. Missing data in a predictor in a cross-sectional study should be handled differently than missing data in the outcome of a longitudinal study. For that reason there is not one optimal solution for handling missing data in a multi-item questionnaire, but depending on the various facets about the missing data there are different missing data methods to use. An overview of data situations and corresponding missing data methods is given in Figure 8.1. The practical guide aims to help researchers find an optimal solution to their specific missing data problem.

In general there are several characteristics of the data analysis and missing data that have to be taken into account. These are the analysis method that is applied to analyze the dataset (i.e., longitudinal analysis or not), the type of variable with missing data (i.e., predictor/covariate or outcome), the missing data mechanism (i.e., MCAR, MAR or MNAR), the overall percentage of subjects with at least one item missing, and the level of missing data in the questionnaire (i.e., missing item scores or missing total scores). When less than 75% of the item scores are missing within subjects, the missing data is considered to be at the item score level and when more than 75% of the item scores are missing within subjects, the missing data is considered to be at the total score level.

For data analysis of studies with outcomes at only one time-point the type of variable in the analysis model that contains missing data is of minor importance. In general whether the missings are in the outcome, in the predictor or covariate of the analysis, the solution is similar. In these studies with the predictors and outcome data collected at one time-point, it is of primary importance whether the missing data are MCAR, MAR or MNAR.

If the missing data are MCAR, the percentage of missing data plays an important role. If the percentage of subjects that have missing data on at least one item is low (i.e., < 10%), then the missing data will have a minor effect on the study results and a complete-case analysis can be performed. However, when the percentage of subjects who have missing data on at least one item is larger (i.e., 10%-75%), it will be more important to sustain statistical power. In that case it is advised to perform an imputation method to be able to use all the available data. In that situation it is advised to apply multiple imputation to the missing item scores, if <75% of item scores are missing within subjects, or multiple imputation applied to the total scores if >75% of item scores are missing within subjects (Eekhout et al., 2014).

For studies with one outcome that have MAR data for a small percentage of data (i.e., < 10% of the subjects have at least one item missing) applying a single stochastic regression imputation could suffice. Although multiple imputation would be most

optimal in these situations, stochastic regression imputation results in unbiased results when data are MAR for a small percentage of subjects with missing data. In stochastic regression imputation, the uncertainty about the missing data is not incorporated in the missing data method. When the percentage of subjects with at least one item missing is small (i.e., <10%) the effect of not including this uncertainty is negligible (Eekhout et al., 2014). When the percentage subjects with missing data is larger (i.e., 10%-75%), multiple imputation is advised. When the items scores are missing (i.e., <75% of the item scores missing within subjects), the imputation (i.e., single stochastic regression imputation or multiple imputation) should be applied to the item scores first, prior to computing the total score for analysis. When large part of the questionnaire was not filled out by the study participants (i.e., >75% of the item scores missing within subjects), so total score level data are missing, the imputation method should be directly applied to the total scores. In datasets where some study participants have item scores missing and some have total score missings, the imputation method should first be applied to the item scores and subsequently to the total scores for the subjects who have the total score missing data (Eekhout et al., 2014).

When the missing data in studies with one outcome are MNAR and the percentage of subjects with missing data is not too large (i.e., <50%), the missings on the item scores (i.e., <75% of the item scores missing within subjects) should be handled by multiple imputation of the items and the missings on the total scores (i.e., >75% of the item scores missing within subjects) by multiple imputation of the total scores. When data are MNAR and more than 50% of subjects have missing data, multiple imputation is not a reliable solution (Eekhout et al., 2014).

In data that require a longitudinal analysis strategy, so when outcomes at more time-points are analyzed, missings in the predictors or covariates need to be handled differently than missings in the outcome. The missing data in the predictors or covariates should be handled the same way as the missing data in studies that have the outcome measured at one time-point.

The missing data in the outcome of a longitudinal study can be handled within the analysis, when a method based on full information maximum likelihood estimation (e.g., mixed model or structural equation model) is used. However, for participants that have all outcomes missing at all the time-points, the missing data cannot be handled and these participants will not be analyzed. When the missing outcome data are MCAR or MAR, the longitudinal methods based on full information maximum likelihood estimation are advised. In situations where the outcomes are missing due to missing item scores, the observed item information should be included in the model as auxiliary variables. When longitudinal data are MNAR, the missing data can be handled in a MNAR model. In a MNAR model there is a relation between the

probability of missing data and the outcome. Two common MNAR models are the selection model and the pattern mixture model. An explanation and evaluation of these models is beyond the scope of this thesis, but are described in other literature (Enders, 2011b; Molenberghs, Thijs, Kenward, & Verbeke, 2003).

In many empirical data situations, the missing data are neither only at the item score level nor only at the total score level. Mostly a dataset contains missing data at both levels. In the diagram (Figure 8.1) a solution is indicated for each situation. In practice, for the subjects who have missing data at the item level (i.e., <75% of the items missing), the multiple imputation procedure should be applied to the items. Simultaneously, the subjects who have missing total scores (i.e., >75% of the items missing) should have their total score imputed. In many software packages it is quite complicated to do this simultaneously.

As a practical solution, there are four steps that can be taken to do this in an appropriate manner in SPSS. (1) Two copies of the dataset can be used. (2) In one copy all data can be imputed at the item level and after the imputation procedure, the total scores should be calculated using the imputed items. (3) In the second copy, the total scores can be calculated prior to the imputation. These total scores are left incomplete when one or more items are missing. Subsequently, the multiple imputation can be applied to the total scores directly. (4) After the imputation procedures are performed, the total scores from the imputed items and the imputed total scores can be merged into one dataset. Then the total scores from the imputed items should be selected for the subjects who had less than 75% of the item scores missing, and the imputed total scores for the subjects with more than 75% of the items missing. After this procedure the regular MI analysis phase and pooling phase can be performed using the merged total scores in the analysis.

The practical guide presented in Figure 8.1 includes no condition where more than 75% of the subjects have missing data. Generally, when more than 75% of the data contains missing values it might not be wise to analyze the data. Although there are situations imaginable where the missing data can be handled adequately and valid analysis results can be obtained. For example in a dataset in which 85% of the subjects have item scores missings, but in this data only a small percentage of subjects have the total scores missing (i.e., <10% of the subjects have >75% of the item scores missing), the other subjects with missing data have less than 75% of the item scores missing. The total scores that will be calculated will be missing for 85% of the subjects, because the totals score will be missing when one or more item scores are missing. However, the fraction of missing information will be smaller than 85%, because the available item score information contains information about the missing data in the total score (Graham, Olchowski, & Gilreath, 2007). In that situation, multiple imputation applied to the item scores might result in valid study results. This

means that the performance of missing data methods as multiple imputation and full information maximum likelihood is not necessarily directly related to the percentage of subjects who contain missing data, but more so to the fraction of missing information (Schafer, 1997). The fraction of missing information (FMI) represents the amount of missing information available to estimate parameters (Rubin, 1987). Theoretically the FMI is as large as the total percentage of missing data, however, this value is reduced by auxiliary variables that can include additional information about the missing data into the analysis or missing data handling (Graham, 2012).

The guidelines on percentages of missing data in the practical guide are recommendations. However, depending on the missing data patterns and location of missing data on the multi-item questionnaire (i.e., missing item scores or missing total scores), the advanced missing data methods might also perform effectively in situations where more than 75% of the subjects have some missing data.

Methodological considerations

Multiple imputation versus full information maximum likelihood

Both multiple imputation and full information maximum likelihood are currently considered to be the state-of-the-art methods for missing data handling (Schafer & Graham, 2002). Both methods can handle missing data in studies with the outcome measured at one time-point or with the outcome measured more than once (Baraldi & Enders, 2010). However, the advice in this thesis is focused on the most optimal and practical methods for epidemiological researchers and therefore one of the two methods is recommended in each study design. In studies that assess information at one time-point, the analysis method might be preferred to be kept simple and straight forward and therefore handle the missing data with multiple imputation. However, if in these studies the missing data would be handled by full information maximum likelihood, the analysis should be specified in for example a structural equation model. This would require additional knowledge of structural equation modeling to accommodate the missing data, while in multiple imputation the analysis method that was planned at the design stage of the study can be applied after the multiple imputation procedure. Whereas, epidemiologists who perform longitudinal analysis might be familiar with methods based on full information maximum likelihood estimation (Twisk, 2013). In that case, it might be more practical for them to handle missing data in studies with outcomes at multiple time-points with full information maximum likelihood than with multiple imputation.

If the missing data are only in the outcome, handling the missing data by multiple

imputation or by full information maximum likelihood in a longitudinal model will yield similar results when the variables in the imputation model are the same as the variables in the longitudinal model (Collins, Schafer, & Kam, 2001; Schafer, 2003). Handling the missing data in the model directly, as is done in full information maximum likelihood, will be more feasible (Enders, 2011a). However, some researchers point out that multiple imputation is the preferred strategy to include auxiliary variables to make the MAR assumption more plausible and therefore prefer handling missing data via this approach when auxiliary information is available (Bell & Fairclough, 2013). Nevertheless, in structural equation models it is possible to include auxiliary variables, without changing model interpretations. Software programs as Mplus accommodate this method and in these programs the inclusion of auxiliary variables to handle missing data is relatively easy (B. O. Muthén, Asparouhov, Hunter, & Leuchter, 2010; L. K. Muthén & Muthén, 1998-2012).

Multiple imputation in practice

As previously mentioned, in multiple imputation many different methods are available to adapt the imputation algorithm to the assumed distribution of the data. Item scores in a multi-item questionnaire are frequently measured by a Likert scale. These are ordinal items which are not necessarily normally distributed, and incomplete ordinal data might best be imputed with the proportional odds model. However, in a simulation study was found that the distribution of the items did not limit the performance of the linear regression algorithm of multiple imputation (Eekhout et al., 2014). Furthermore, the predictive mean matching procedure is more robust against the deviations from the normal distribution and imputes more realistic values compared with linear regression imputation. For that reason, predictive mean matching might be attractive for imputing categorical item scores; however the linear regression algorithm performed just as well as predictive mean matching when final analysis results were evaluated (Eekhout et al., 2014).

Another example where the distribution of the data deviates from normality is cost-data. The total costs in a study are often the sum of several sub-costs. The relation between the sub-costs and the total costs is not reflective, however also in the cost data it was most feasible to apply the missing data method to the sub-costs instead of to the total costs directly. This is in concordance with the advice with respect to multi-item questionnaires, where the missing data need to be handled at the item level. In studies where costs are measured for economic evaluations, a part of the sample have zero costs, but the participants that actually have costs sometimes have very large cost values. Furthermore, costs cannot become negative. So, the distribution of cost data is not normal and might require different methods for multiple imputation. It is possible to transform the data with a log to obtain a (close

to) normal distribution. Alternatively one can use a method that imputes the missing cost data in two separate steps. In this method first a value for having costs versus not having costs is imputed by a logistic method and in the second step the people that are indicated to have costs will have their costs imputed by predictive mean matching. Another option is to use predictive mean matching as a method without the first step. It might be expected that a method that takes into account all aspects of the distribution would perform best. Nevertheless, in a study that was conducted on multiple imputation of cost data (MacNeil-Vroomen et al., under review) the two step method did not perform better than predictive mean matching without the first step. Moreover, simply log-transforming the data and imputing that distribution was the most stable solution in larger percentages of missing data.

The application of multiple imputation to the item scores can pose some problems in some study designs. The basic rule for the construction of the imputation model is to include all relevant information about the analysis model in the imputation model, together with the information relevant to the missing data handling. This includes all variables used in the main analysis and auxiliary variables. However, when many multi-item questionnaires are administered in one study the imputation model might become extremely large and even make model estimations impossible. As a solution it is possible to construct the imputation model in such a way that for each separate questionnaire the item scores are imputed using the total scores from the other questionnaires as predictors. These total scores from the other questionnaires are calculated from the imputed item scores of that questionnaire. This strategy, called passive imputation, is further explained in a simulation study that evaluated this method (Eekhout, De Vet, De Boer, Twisk, & Heymans, under review).

Missing Not At Random

The assumption about the missing data mechanism is very important in order to select a valid method to handle the missing data. Unfortunately, it is not possible to distinguish between MAR and MNAR mechanisms, because the missing values are unknown. Brand (1999) describes in Chapter 2 of his dissertation two examples that demonstrate how an initially MNAR missing data mechanism can change into MAR by including additional variables that are related to the probability of missing data. In practice, by including variables related to the probability of missing data a MNAR mechanism can get closer to MAR. Accordingly, the MAR assumption can be made more plausible by including auxiliary information in the missing data handling method (Baraldi & Enders, 2010). Furthermore, it is often advised to do sensitivity analysis by applying additional MNAR models (e.g., selection models or pattern mixture models (Enders, 2011b; Molenberghs et al., 2003)) and examine if the conclusions change (Enders, 2011a).

Future research

This thesis focuses on handling missing data in the total score that are caused by missing item scores. Several solutions are offered, however the solutions presented here might not be the only valid options to handle missing item score data. It might be interesting to compare the methods that were proposed here, multiple imputation applied to the items or including the item scores as auxiliary variables in a full information maximum likelihood analyses, to methods that use the item scores in the analysis model directly. This can be accomplished by including the item scores as indicators for a latent variable in a structural equation model or by using another latent variable model that is robust against missing data, such as item response techniques. It might be interesting to study in which situations it is preferred to use the item scores in the analysis. For example, the internal consistency of the multi-item questionnaire (i.e., coherence of the items) might be related to the performance of such methods. Also in this context, the possibilities of using the item scores in the analysis to handle missing data in epidemiological studies that measure at one time-point can be further explored and compared to applying multiple imputation.

The studies about missing data in multi-item questionnaires referred to in this thesis are about questionnaires with a reflective model. In a reflective model the change in the construct (e.g., better physical functioning) is reflected by changes in the items (e.g., higher scores on the items). In this setting the construct is measured indirectly by the items. In questionnaires with a formative model the construct is more like an index score, for example food intake measured by a food frequency questionnaire. In this case the items can be an extensive list of food items that together form tot total food intake. In a formative model the items are mostly not as highly correlated and a change in the construct (e.g., food intake) is not necessarily reflected by a change in all of the items (de Vet, Terwee, Mokkink, & Knol, 2011). The investigated methods for missing data have not been extensively evaluated in multi-item questionnaires with a formative model. However, it may be expected that the advice formulated for questionnaires with a reflective model (i.e., multiple imputation at the item level and full information maximum likelihood with auxiliary item information) will apply to these questionnaires as well. Nevertheless, the distributions of the items may be very skewed or zero-inflated in some of the formative questionnaires due to the count nature of the items. This might require different approaches that take account of this distribution as investigated in some studies (e.g., Fraser et al., 2009; Nevalainen, Kenward, & Virtanen, 2009; Parr et al., 2008). This needs to be investigated in future research.

The relation between the fraction of missing information and the performance of missing data methods has been previously studied in non-questionnaire data. Most studies about fraction of missing information were aimed at the required number of

imputations for multiple imputation (e.g., Bodner, 2008; Graham et al., 2007; Schafer, 1997). The fraction of missing information in multi-item questionnaire data and the consequences for the missing data methods might be a relevant tool to analyze the performance of missing data handling for item scores. This should be explored in future research.

In longitudinal studies, where the outcomes are measured more than once, the missing data in the predictors or covariates can be handled by multiple imputation. It might be interesting to study the possibilities of handling the missing data in the predictors in a structural equation model. It can also be useful to further study the applications of MNAR models when total scores are incomplete due to missing item scores. And further, to what extent the items can be included as auxiliary variables in these models and whether this improves model estimates, might be an interesting and useful focus for future studies.

Conclusion

This thesis provides a practical guide on how to handle missing data in multi-item questionnaires. In Box 8.1 an overview of what is known, what is new and future challenges are summarized as the key messages of this thesis. Overall, it is important to incorporate all available information from the item scores in order to obtain the optimal level of accuracy and precision.

What is known:

- Missing data can cause biased results when the missings are not missing completely at random.
- Missing data in multi-item questionnaires can be handled at the item level or at the total score level.
- Multiple imputation and Full information maximum likelihood estimation methods are the advanced state-of-the-art missing data methods.

What is new:

- Methods that are advised in manuals for multi-item questionnaires are often sub-optimal and should be ignored.
- Missing data in multi-item questionnaire should always be handled at the item level. Using the information from the observed item scores in the missing data handling method improves accuracy and precision of analysis results.
- Including item information as "auxiliary variables" to handle missing data in longitudinal models that analyze the total scores improves precision and power of coefficient estimates.
- Applying "passive imputation" to impute the item scores when the number of items is extremely large is a valid method to handle missing item scores in large survey studies.

Future challenges:

- To investigate the relation between the fraction of missing information and the amount of missing item scores and the performance of missing data methods.
- Comparing the use of missing data methods to handle the missing item scores, to advanced methods that include the item scores in the main analysis, for example when missings are in the predictor.
- Incorporating the observed item scores in longitudinal models that correct for MNAR data.

Box 8.1. Key messages of this thesis

References

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. Journal of School Psycholology, 48(1), 5-37.

Bell, M. L., & Fairclough, D. L. (2013). Practical and statistical issues in missing data for longitudinal patient reported outcomes. Stat Methods Med Res, 19, 19.

Bodner, T. E. (2008). What Improves with Increased Missing Data Imputations? Structural Equation Modeling: A Multidisciplinary Journal, 15(4), 651-675.

Brand, J. P. L. (1999). Development, implementation and evaluation of multiple imputation strategies for the statistical analysis of incomplete data sets. . Dissertation, Rotterdam: Erasmus University Rotterdam.

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 Vet, H. C., De Boer, M. R., Twisk, J. W., & Heymans, M. W. (under review). Passive imputation of missing values in studies with many multi-item questionnaire outcomes. American Journal of Epidemiology.

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.

Enders, C. K. (2011a). Analyzing longitudinal data with missing values. Rehabilitation Psychology, 56(4), 267-288.

Enders, C. K. (2011b). Missing not at random models for latent growth curve analyses. Psychol Methods, 16(1), 1-16.

Fraser, G. E., Yan, R., Butler, T. L., Jaceldo-Siegl, K., Beeson, W. L., & Chan, J. (2009). Missing data in a long food frequency questionnaire: are imputed zeroes correct? Epidemiology, 20(2), 289-294.

Graham, J. W. (2012). Missing data analysis and design. New York: Springer.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. Prev.Sci., 8(3), 206-213.

MacNeil-Vroomen, J., Eekhout, I., Dijkgraaf, M. G., Van Hout, H., De Rooij, S. E., Heymans, M. W., & Bosmans, J. E. (under review). Comparing multiple imputation strategies for zero-inflated cost data in economic evaluations: which method works best? European Journal of Health Economics.

Molenberghs, G., Thijs, H., Kenward, M. G., & Verbeke, G. (2003). Sensitivity Analysis of Continuous Incomplete Longitudinal Outcomes. Statistica Neerlandica, 57(1), 112-135.

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.

Muthén, L. K., & Muthén, B. O. (1998-2012). Mplus User's Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén.

Nevalainen, J., Kenward, M. G., & Virtanen, S. M. (2009). Missing values in longitudinal dietary data: a multiple imputation approach based on a fully conditional specification. Stat.Med., 28(29), 3657-3669.

Parr, C. L., Hjartaker, A., Scheel, I., Lund, E., Laake, P., & Veierod, M. B. (2008). Comparing methods for handling missing values in food-frequency questionnaires and proposing k nearest neighbours imputation: effects on dietary intake in the Norwegian Women and Cancer study (NOWAC). Public Health Nutr, 11(4), 361-370.

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.

Schafer, J. L. (2003). Multiple Imputation in Multivariate Problems When the Imputation and Analysis Models Differ. Statistica Neerlandica, 57(1), 19-35.

Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. Psychol. Methods., 7(2), 147-177.

Twisk, J. W. R. (2013). Applied Longitudinal Data Analysis for Epidemiology: A Practical Guide, Second Edition. New York: Cambridge University Press.

de Vet, H. C. W., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). Measurement in Medicine. Cambridge: Cambridge University Press.