Chapter 3

Missing data in a multi-item questionnaire are best handled by multiple imputation at the item score level

Published: 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.

Abstract

Regardless of the proportion of missing values, complete-case analysis is most frequently applied, although advanced techniques such as multiple imputation are available. The objective of this study is to explore the performance of simple and more advanced methods for handling missing data in case some, many, or all item scores are missing in a multi-item instrument. Real-life missing data situations were simulated in a multi-item variable used as a covariate in a linear regression model. Various missing data mechanisms were simulated with an increasing percentage of missing data. Subsequently, several techniques to handle missing data were applied to decide on the most optimal technique for each scenario. Fitted regression coefficients were compared using the bias and coverage as performance parameters. Mean imputation caused biased estimates in every missing data scenario when data are missing for more than 10% of the subjects. Furthermore, when a large percentage of subjects had missing item scores (>25%), multiple imputation methods applied to the items outperformed methods applied to the total score. We recommend applying multiple imputation to the item scores in order to get the most accurate regression model estimates. Moreover, we advise not to use any form of mean imputation to handle missing data.

Keywords: missing data, multiple imputation, multi-item questionnaire, item imputation, methods, bias, simulation

Introduction

Missing data on multi-item instruments is a frequently seen problem in epidemiological and medical studies. Multi-item instruments can be used to measure for example quality of life, coping ability or other psychological states. A multi-item instrument generally consists of several items that measure one construct (de Vet, Terwee, Mokkink, & Knol, 2011), for example the Pain Coping Inventory assesses active coping skills for people with pain complaints by 12 items (Kraaimaat & Evers, 2003). Missing data on these kinds of instruments can occur as missing item scores, when several items are not completed, or as missing data in total scores when the entire instrument is not filled out. Furthermore, missing item scores impair the calculation of the total score which can lead to missing total scores as well. For missing data in item and total scores, different missing data handling methods are available, with complete-case analysis (CCA) as the most frequently used method (Eekhout, de Boer, Twisk, de Vet, & Heymans, 2012). In general, CCA tends to perform well under the strict assumption that missing data are a completely random subsample of the data, in other words missing completely at random (MCAR) (Rubin, 1976). However, CCA reduces power caused by a decreased sample size. Single imputation methods, such as mean imputation of the total score and item mean imputation, may be used to preserve the sample size by replacing the missing values by the mean score, but these methods reduce the variability in the data. Single stochastic regression imputation (SRI) uses observed data to predict the missing value and adds residual error to the imputed data to restore the variability in the data, but this method does not take the uncertainty of the imputed values into account.

Mostly, the probability of missing data depends on other observed variables, indicated as missing at random (MAR) (Rubin, 1976). In contrast to traditional methods as CCA and mean imputation, more advanced methods, such as multiple imputation, produce reliable and unbiased results under the MAR mechanism and take missing data uncertainty into account (Janssen et al., 2010; Little & Rubin, 2002). Both traditional and advanced methods can be applied either to the missing item scores, or directly to the missing total scores.

The comparison between missing data methods for item level and total score level missingness in questionnaire data is seldom made in one study (Eekhout, et al., 2012). Other simulation studies have researched the performance of missing data methods applied to non-questionnaire data (Collins, Schafer, & Kam, 2001; Marshall, Altman, Royston, & Holder, 2010) or only studied methods applied to the item scores of a multi-item instrument (Burns et al., 2011; Hawthorne & Elliott, 2005; Roth, Switzer, & Switzer, 1999; van Buuren, 2010; van Ginkel, Sijtsma, van der Ark, & Vermunt, 2010). For example Burns et al. (2011) studied the performance of multiple imputation of missing item scores, but did not compare this to imputing at the total score level of their questionnaire. So far it is still unclear if it is better to apply a missing data handling method to the missing item scores or to the total scores when some or many items in a multi-item instrument are missing. Moreover, the impact on the study results of different missing data methods when multi-item data are missing on the covariate has not been researched extensively yet. The current study aims to explore the performance of different missing data handling methods designed for missing item scores and missing total scores in a multivariable regression model. This objective is considered in the following two aspects: (1) which missing data methods should be used to handle missing data; and (2) should this missing data method be applied to the item scores or to the total scores.

Methods

Simulation set up

In order to investigate the differences between several imputation methods, we used a simulation procedure comparable to the study performed by Marshall et al. (2010). We based our simulation on an empirical dataset, which was previously used in a prospective cohort study investigating the prognosis of low back pain (Heymans et al., 2006). In this study we used a cross-sectional part of these data that contained the multi-item variable active coping of the Pain Coping Inventory (PCI-active) (Kraaimaat & Evers, 2003). The PCI consists of 12 items with four ordered response categories that result in a total sum score, which we consider as a continuous scale. Additionally, five other covariates were selected to be included in this dataset: gender, health status, job demands, number of working years, absenteeism, and the outcome variable was low back pain intensity. Using the means and covariance matrix of these empirical data, 500 simulated data samples of 500 subjects were generated using the mvrnorm function in package MASS in R statistical software (Venables & Ripley, 2002). Subsequently, in each simulated sample, missing data were created in only the multi-item covariate PCI-active under several missing data mechanisms. After this step, several techniques were applied to handle the incomplete datasets. The implications of these different techniques were compared by fitting a multivariable regression model to the data. This model regressed PCI-active total score, gender, health status and job demands on pain intensity. In order to have an imputation model that differs from the regression model, number of working years and absenteeism were only included in the imputation models, but not in the final regression model. Model coefficients fitted to the 'handled data' were compared to the 'true' model coefficients fitted to the same samples without missings.

Step 1 Generating missing data

The generation of missing data in the multi-item covariate PCI-active was performed using a program that was translated from SAS software (Brand, van Buuren, Groothuis-Oudshoorn, & Gelsema, 2003) into R statistical software (R Core Development Team, 2012) by the first (IE) and last author (MWH). This program was used to generate multivariate missing data, according to the MCAR, MAR and MNAR mechanisms. In the MCAR situations the selection of item scores that were made missing in the PCI-active covariate was completely random. In the MAR situations item scores of the PCI-active covariate were made missing depending on the values of the observed items and the other covariates gender, health status, job demands number of working years, absenteeism and the outcome low back pain intensity. For the MNAR situations, the scores that were made missing also depended on the values of the PCI-active item scores themselves.

We generated missing item data in the PCI-active covariate according to the following four patterns: (1) a pattern where 25% of the item scores were made missing within subjects, (2) a pattern where 50% of the item scores were made missing, (3) a pattern where 75% of the item scores were made missing, (4) or a pattern where 100% of the items were considered missing. These missing item patterns were applied to 10%, 25%, 50%, or 75% of the subjects. This resulted for example in a situation where in 10% of the subjects, 25% of the item scores were missing.

Total scores of the PCI-active covariate can only be calculated if all item scores are available. Consequently, in the first three patterns mentioned above, where some of the item scores in the PCI-active covariate were made missing within a subject, the PCI-active total score for that subject was also missing. This situation was reflected by the fourth pattern. This made it possible to study separately, in each simulated sample, the influence of missing data methods when they were applied to missing values in item scores or total scores. By generating the incomplete data according to the above described scenarios, 48 different situations were investigated.

Step 2 Methods to handle missing data

The generated incomplete datasets were handled using different methods, summarized in Table 3.1. As previously mentioned, these methods can either be applied to the missing item scores, after which the total score can be calculated or to the missing total score directly. Both of these possibilities were explored in this study. After applying the method, a multivariable regression model was fitted and regression coefficients were estimated.

1. Complete-case analysis.

In a CCA, only the subjects with complete observations for the PCI-active covariate were included in the analysis. Accordingly, all subjects with missing item scores were removed from the data and the model was fitted to the remaining sample. Consequently this method would yield the same results when applied to the missing item data as when applied to the missing total scores data directly.

Table 3.1.
Summary of the application of different missing data methods to item scores and/or total scores.

Label	Method name	Applied to item scores	Applied to total scores
1 CCA	Complete-case analysis	X[a]	X
2a Mean	Mean imputation		X
2b Item mean	Item mean imputation	X	
2c Person mean	Person mean imputation	X	
2d Two-way	Two-way imputation	X	
3 SRI	Single stochastic regression imputation	X	X
4a MI-SR	Multiple Imputation by stochastic regression	X	X
4b MI-PMM	Multiple Imputation by predictive mean matching	X	X
4c MI-PO	Multiple Imputation by proportional odds model	X	

Note: [a] Equal to application to total scores.

2. Mean imputation.

In mean imputation the missing scores were imputed with the mean score of the non-missing data. The missing data on the PCI-active covariate total scores were imputed with the mean total score of all observed subjects in mean imputation applied to the total scores (2a). In item mean imputation (2b) a missing item score was imputed with the mean score for all complete data on that item. In person mean imputation (2c), the mean score of the items per subject was calculated, and for each subject missing item scores were imputed with this 'personal mean score'. Two-way imputation (2d) combines the person mean and the item mean to account for both the personal effect and the item effect. The person mean was added to the item mean, and then the overall mean was subtracted. Furthermore, a random error term was added to account for variability in the data (Bernaards & Sijtsma, 2000).

3. Stochastic regression imputation.

In SRI the missing values were imputed with the regression estimates from the observed variables when applied to the total score. For methods applied to the items, the regression estimate of the observed variables and the observed items was used. Regression assumes that the imputed values fall directly on the regression line, so it implies a correlation of 1 between the predictors and the incomplete outcome variable (PCI-active). Stochastic regression imputation aims to reduce the bias by an extra step of augmenting each predicted score with a normally distributed random error with a variance equal to the variance of the regression model.

4. Multiple imputation.

The multiple imputation (MI) method has three phases, the imputation phase, the analysis phase and the pooling phase. First, the incomplete data was completed by imputing a value for the missing scores in the imputation phase. When applied to the item scores, the missing item scores were imputed, but when applied to the missing total scores the PCI-active total score was imputed. The imputed values were estimated from the observed variables in the dataset by an imputation algorithm and a random residual term was added to each resulting estimate. The imputation algorithm is a regression equation specified in the imputation model using the observed variables to estimate the missing value. In case of the item score application, the observed item scores of the PCI-active covariate were also included in the imputation model. Multiple datasets were generated, each with different imputed values for the missing items or total scores. In this simulation study we generated 15 imputed datasets, which is higher than the minimum recommended number of 5 (van Buuren, 2012), and still computationally and practically possible in this simulation study. During the analysis phase, the analysis was carried out on each dataset using the same procedures that would have been used had the data been complete. So, when applied to the item scores, the item scores were added to form a total score and the regression analysis was performed. In case of the missing total scores, the imputed total scores were used in the analysis. Finally, in the pooling phase the multiple sets of results, or parameter estimates, were combined into one single set of results according to Rubin’s rules (Rubin, 2004). In this study three different imputation algorithms were applied and compared to impute the missing data in the imputation phase. These were stochastic regression (SR; 4a), predictive mean matching (PMM; 4b), and a proportional odds model (PO; 4c) (van Buuren, 2007, 2010). The latter model is recommended for missing ordinal categorical data (van Buuren, 2012). The multiple imputations were done using the mice function in package MICE (van Buuren & Groothuis-Oudshoorn, 2011).

Step 3 Comparing missing data handling methods

The model coefficients of the ‘handled datasets’ were compared to the coefficients based on the ‘true data’. The true data were generated by running the simulation process without missing data. Accordingly, the regression model was fitted to 500 simulated complete samples and the average regression coefficients formed the true values against which the missing data simulations were compared. Regression coefficients were considered ‘biased’ when the estimate was outside a limit of 0.5

standard error from the true coefficient (Schafer & Graham, 2002). We also looked at the estimates for the standard error (SE) of the regression estimates, which were required to be estimated somewhat larger than the true SE in order to incorporate the appropriate missing data uncertainty (Enders, 2010). Furthermore, the coverage of the true value of the regression coefficients within the confidence limits of the estimated coefficients was computed. This was calculated as the percentage of times that the true regression coefficient was within the confidence interval of the estimates from the datasets after the missing data handling methods were applied. Coverage of 95% is optimal whereas higher coverage indicates that the method might be too conservative and lower coverage value suggests a higher than expected type I error (Burton, Altman, Royston, & Holder, 2006). Decreased coverage can be caused by too narrow confidence intervals as a result from underestimated standard errors. Standard errors should incorporate uncertainty of missing data to overcome this problem (Siddique, Harel, & Crespi, 2012). All analyses and simulations were performed in R statistical software (R Core Development Team, 2014), scripts are available by the first author upon request.

Results

In Table 3.2, the regression coefficient and standard error estimates for the PCI-active total score under the three missing data mechanisms are presented. Not surprisingly, for the MCAR data the coefficient estimate was the same as the true coefficient value, but the standard error increased with higher data missing rates. A similar trend was seen in the MAR and MNAR missing data situations, however accompanied by much larger deviations in SE's.

Table 3.2.
Coefficient and standard error (SE) of the PCI-active total score covariate according to the three missing data mechanisms MCAR, MAR and MNAR after a complete-case analysis.

	'True' Coefficient (SE)	10%[a] Coefficient (SE)	25%[a] Coefficient (SE)	50%[a] Coefficient (SE)	75%[a] Coefficient (SE)
MCAR	0.1411 (0.0135)	0.1413 (0.0142)	0.1413 (0.0156)	0.1410 (0.0191)	0.1424 (0.0274)
MAR	0.1411 (0.0135)	0.1414 (0.0142)	0.1413 (0.0157)	0.1418 (0.0198)	0.1428 (0.0259)
MNAR	0.1411 (0.0135)	0.1412 (0.0143)	0.1416 (0.0159)	0.1498 (0.0207)	0.1729 (0.0285)

Note: [a] Percentage of cases that had a missing PCI-active total score.

Figures 3.1 and 3.2 present the effect of the missing data handling methods on the estimates of regression coefficients for an increasing amount of missing total scores and an increasing amount of missing item scores when 25% of the subjects have missing data, respectively. A full tabulation of the results is presented in the online Appendices (www.jclinepi.com).

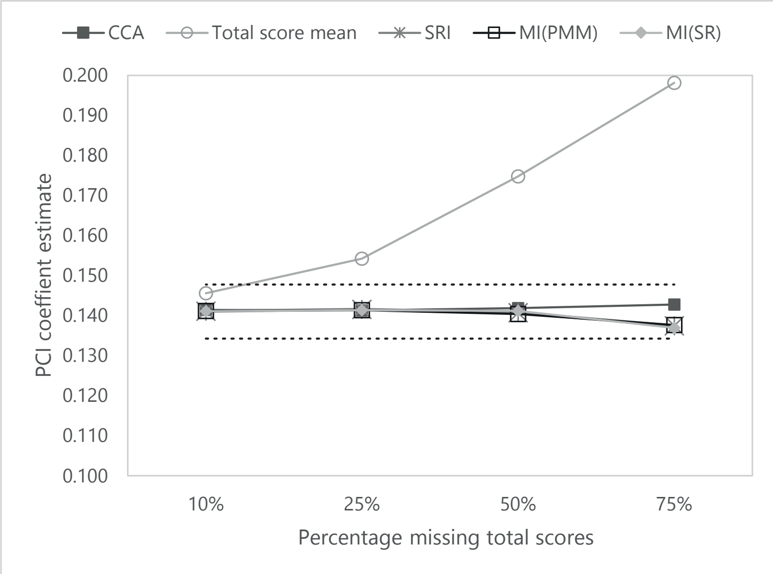

Figure 3.1. Regression coefficient estimates of the PCI-active covariate for different missing data methods applied to the total score for when an increasing percentage of subjects had missing at random data in total score. The black dashed lines depict the thresholds for bias at 0.5 SE from the true coefficient.

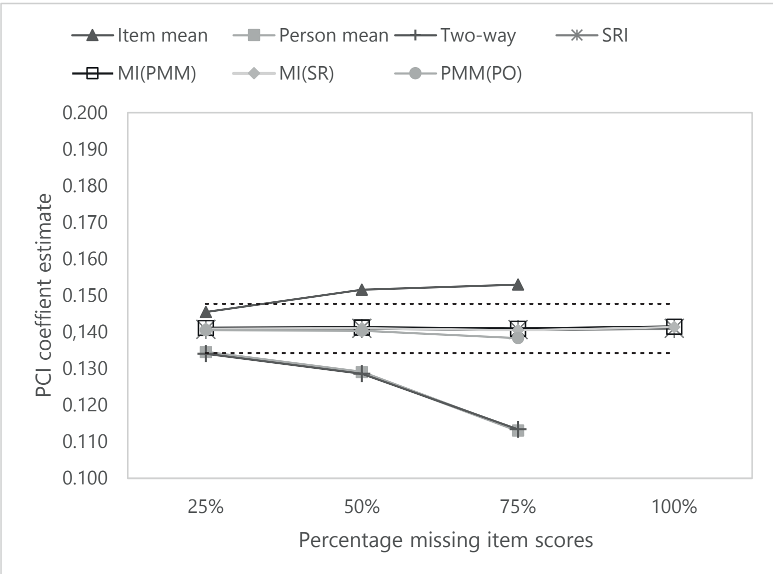

Figure 3.2. Regression coefficient estimates of the PCI-active covariate for different missing data methods applied to the items when 25% of the subjects had missing at random data in an increasing percentage of missing item scores. The black dashed lines depict the thresholds for bias at 0.5 SE from the true coefficient.

Regression coefficients

In the MAR data situation multiple imputation as well as SRI and CCA gave unbiased regression coefficient estimates of the PCI-active variable, regardless of the number of missing total scores and items (Figure 3.1 and 3.2). The worst method applied to the total scores was mean imputation and the worst methods applied to item scores were item and person mean imputation and two-way imputation which yielded biased coefficients. For example, item mean imputation had values in the range of 0.150 and 0.172 compared to the true value of 0.141 for the PCI-active variable, when 50% of the subjects had missing item scores (online Appendix, www.jclinepi.com). Mean imputation of the total score yielded biased regression estimates when more than 10% of the cases had a missing total score. Results were the same in MCAR and MNAR data; mean imputation methods were the worst solutions as well. For example item mean imputation had estimates ranging from 0.152 to 0.185 compared to the true value of 0.141 when 50% of the subjects had MNAR data. With 50% or more MNAR data in item or total scores, also advanced MI methods gave biased coefficient estimates.

Standard errors and coverage

In MAR data the methods that yielded the largest bias in the regression coefficient estimates as reported above, also showed biased SE's (online Appendix, www.jclinepi.com). This was most evident for the mean imputation methods in all situations of missing item and total scores. Figure 3.3 and 3.4 display the SE estimates in MAR data for the methods that gave the best results with respect to the regression coefficients.

Despite the unbiased coefficient estimates for PCI-active in a CCA, the standard error was highly overestimated for this method in MAR data (Figure 3.3). Overall multiple imputation showed slightly smaller bias and better SE estimates when applied to the items than when this method was applied to the total scores when less than 50% of the item scores were missing. Similar SE results for the different methods were seen in MCAR and MNAR data (online Appendix, www.jclinepi.com). Standard error estimates in MI methods correctly incorporated missing data uncertainty both when applied to item scores and to total scores. Coverage rates, which measure the combined performance of the coefficient estimate and SE by evaluating the confidence interval, confirm this. Coverage rates were worst for single imputation methods and best for multiple imputation of the item scores (Table 3.3).

Other covariates

Other covariates in the model were also affected by the missing data in the PCI-active variable (online Appendix, www.jclinepi.com). In general, mean imputation methods resulted in biased estimates on the other covariates when an increasing

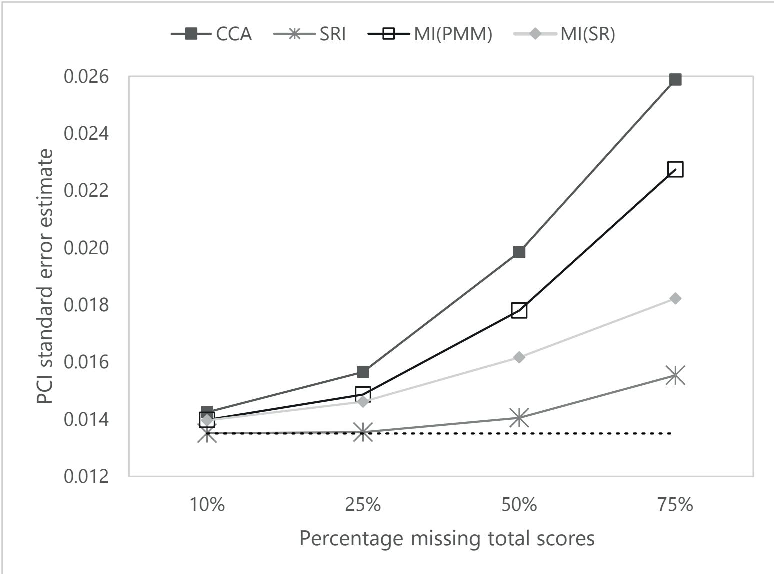

Figure 3.3. Standard error estimates for the missing data methods that have unbiased PCI-active coefficient estimates when an increasing percentage of subjects had missing at random data in total scores. The black dashed line depicts the SE estimate of the true coefficient (0.0135).

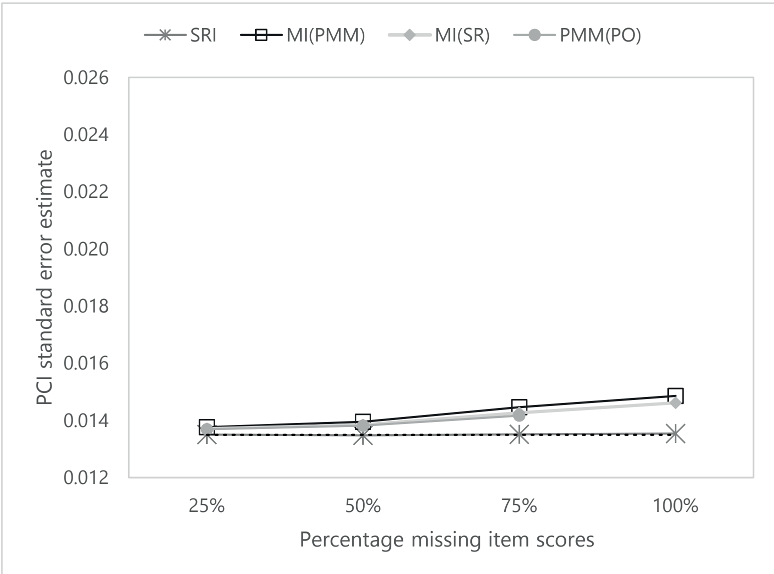

Figure 3.4. Standard error estimates for the missing data methods that have unbiased PCI-active coefficient estimates when 25% of the subjects had missing at random data in an increasing percentage of missing item scores. The black dashed line depicts the SE estimate of the true coefficient (0.0135).

amount of data were missing on the PCI-active covariate. MI methods applied to the item scores mostly resulted in unbiased coefficient estimates for all other covariates in all missing data situations. The SE estimates in the other covariates only showed large deviations for the CCA. The coverage was mostly acceptable, round 95%, except for the mean imputation methods, especially when more items were missing in a large amount of data (>50%).

Table 3.3.
Coverage rates for all missing data methods when 25% of the subjects had missing at random data in an increasing percentage of missing item scores.

	Percentage missing item scores			
Method	25%	50%	75%	100%
Item mean	94%	87%	85%	-
Person mean	94%	83%	41%	-
Two-way	92%	83%	40%	-
Total score mean	-	-	-	85%
SRI	94%	95%	93%	91%
MI(SR)	96%	96%	95%	95%
MI(PMM)	95%	96%	95%	95%
MI(PO)	96%	95%	93%	-

Note: All empty cells are not-applicable conditions.

Discussion

The results of our study are that missing item data are best handled by applying MI based on PMM or SR to the item scores regardless of how many subject scores and item scores are missing. Furthermore, SRI also seems to yield acceptable results, and mean imputation of the total scores performs worst. Additionally, we showed that the underlying mechanism influences the performance of the missing data handling method, especially when large amounts of data are missing. This is not only of concern when working with total scores but also when working with missing item scores in multi-item instruments.

Moreover, our results showed that complete-case analysis performed satisfactory with respect to the regression coefficient estimates when data are MAR on the covariate, which was also found in the simulation study by Marshall et al. (2010). However, standard errors are largely overestimated and hence power is reduced. For that reason it is not recommended to perform a complete-case analysis, especially when more than 10% of the subjects have missing data. Nevertheless, in about 80% of epidemiological studies a complete-case analysis is still used (Eekhout, et al., 2012).

Item mean imputation, which is advised in user-manuals of widely used multi-item questionnaires as the SF-36 (Ware, Kosinski, & Keller, 1994) and the PCI (Kraaimaat & Evers, 2003), results in highly biased estimates in all of the missing item data patterns when more than 10% of the subjects have missing data. Therefore we would not recommend item mean imputation in any situation of missing data. Furthermore, person mean and two-way imputation result in underestimated coefficient and SE estimates. Of the single imputation methods, SRI performs best by far. Other studies with missing covariate data also found this method to exceed the performance of other single imputation methods (Enders, 2010; Pastor, 2003). Even though SRI produced valid regression coefficient estimates at first sight, the imputed values are treated as real data and imputation uncertainty is ignored. This leads to narrow confidence intervals resulting in decreased coverage (Enders, 2010). In studies that investigated more complicated missing data situations, SRI proved to underestimate the standard error (Gold & Bentler, 2000; Newman, 2003). Repeating the imputation process multiple times to account for the missing data uncertainty, as is done in multiple imputation, is therefore recommended. Multiple imputation outperforms the ad-hoc methods and produces minimal bias in model estimates. Other studies have found these same results when comparing multiple imputation to ad-hoc methods applied to continuous variables (Donders, van der Heijden, Stijnen, & Moons, 2006; Kneipp & McIntosh, 2001; Marshall, et al., 2010; Schafer & Graham, 2002). We found that this works for missings in total scores as well as missing values in item scores.

When missing data occurred in 50% or more of the cases, all methods applied to the total score yield largely overestimated standard errors. However, when comparable missing data methods are applied to the item scores of this same data, resulting coefficient estimates and standard errors are much more accurate, with multiple imputation methods showing the best results. Therefore, multiple imputation applied to the item scores is preferred over imputation methods applied to the total scores in these situations.

In addition, we found that the estimation of regression coefficients of other covariates was disturbed by the missing data in the PCI-active covariate. Hence missing data on one variable in the model has an effect on the estimates of all other covariates in the model even though these covariates do not contain any missing data. This effect is larger in MAR and MNAR data than in MCAR data. It would be expected that this influence is typically seen in highly correlated data. However, in our simulated data correlations between the variables were low to moderate ranging from 0.10 to 0.45.

In this simulation study, an example dataset was used as a template to simulate missing data. This is a beneficial point because the simulated scenarios reflect real-life research situations and therefore give a realistic view of the magnitude of the effects on the results. Moreover, the simulated MAR data depended on all covariates and on the outcome which reflects a probable missing data situation. However, the PCI-active items in the example data had a Cronbach's alpha of 0.74, which reflects an acceptable but not excellent internal consistency (George & Mallery, 2009). This might be an explanation for the disappointing results for person mean, two-way imputation and MI with a proportional odds model, because these methods are based on the internal consistency of a multi-item instrument (Bernaards & Sijtsma, 2000). Also, previous simulation studies that reported optimistic results for two-way imputation only investigated small amounts of missing data (≤20%), in which situations we found unbiased estimates as well, or included repeating two-way imputation multiple times (van Ginkel, et al., 2010; van Ginkel, van der Ark, & Sijtsma, 2007a, 2007b). Furthermore, missing data were only generated on one covariate. Results might be different when missing data occurs in the outcome or both in the covariate(s) and the outcome, which can be explored in a future study.

The comparison between missing data methods applied to item scores and methods applied to total scores as investigated in our study has never been made before. Previous studies investigated solely single item imputation methods applied to missing item scores (Hawthorne & Elliott, 2005; Roth, et al., 1999). These studies did not compare these methods to multiple imputation methods, or to methods applied to total scores. Our results are important for all researchers working with multi-item instruments. Manuals of most multi-item questionnaires have been developed prior to the development and exploration of most advanced methods applied in this study. Therefore it is important to follow-up on current literature to make a fair assessment of the missing data solutions available aside from the methods described in the questionnaire guidelines. Moreover, many major statistical packages such as SAS (SAS Institute Inc., 2011), R statistical software (R Core Development Team, 2012), Stata (StataCorp., 2011), and also SPSS (SPSS Inc., 2008), presently offer applications to multiply impute missing data. Nevertheless, a previously conducted review showed that only 8% of epidemiological studies currently use multiple imputation to handle missing data (Eekhout, et al., 2012).

To sum up, as expected the more advanced methods, such as multiple imputation, perform better than the traditional methods. Furthermore, multiple imputation applied to the item scores performs better than methods applied to the total scores and is therefore advised. However, when only a small amount of item scores are missing (<25%) in only a small amount of data (<10%), single imputation such as SRI or CCA might be preferred with MAR data in the covariate over multiple imputation purely for practical reasons (van Ginkel, et al., 2010; van Ginkel, et al., 2007a). We advise not to use mean imputation applied to missings in items or total scores.

References

Bernaards, C. A., & Sijtsma, K. (2000). Influence of Imputation and EM Methods on Factor Analysis when Item Nonresponse in Questionnaire Data is Nonignorable. Multivariate Behavioral Research, 35(3), 321-364.

Brand, J. P. L., van Buuren, S., Groothuis-Oudshoorn, K., & Gelsema, E. S. (2003). A toolkit in SAS for the evaluation of multiple imputation methods. Statistica Neerlandica, 57(1), 36-45.

Burns, R. A., Butterworth, P., Kiely, K. M., Bielak, A. A., Luszcz, M. A., Mitchell, P., . . . Anstey, K. J. (2011). Multiple imputation was an efficient method for harmonizing the Mini-Mental State Examination with missing item-level data. J Clin Epidemiol.

Burton, A., Altman, D. G., Royston, P., & Holder, R. L. (2006). The design of simulation studies in medical statistics. [Research Support, Non-U.S. Gov't]. Stat Med., 25(24), 4279-4292.

van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. Stat.Methods Med.Res., 16(3), 219-242.

van Buuren, S. (2010). Item Imputation Without Specifying Scale Structure. Methodology, 6(1), 31-36.

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

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). MICE: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, 45(3), 1-67.

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.

Donders, A. R., van der Heijden, G. J., Stijnen, T., & Moons, K. G. (2006). Review: a gentle introduction to imputation of missing values. Journal of Clinical Epidemiology, 59(10), 1087-1091.

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.

Enders, C. K. (2010). Applied Missing Data Analysis. New York, NY: The Guilford Press.

George, D., & Mallery, P. (2009). SPSS for Windows Step by Step: A Simple Guide and Reference, 17.0 Update: Allyn & Bacon.

van Ginkel, J. R., Sijtsma, K., van der Ark, L. A., & Vermunt, J. K. (2010). Incidence of missing item scores in personality measurement, and simple item-score imputation. Methodology, 6(1), 17-30.

van Ginkel, J. R., van der Ark, L. A., & Sijtsma, K. (2007a). Multiple imputation for item scores when test data are factorially complex. Br.J.Math.Stat.Psychol., 60(Pt 2), 315-337.

van Ginkel, J. R., Van der Ark, L. A., & Sijtsma, K. (2007b). Multiple Imputation of Item

Scores in Test and Questionnaire Data, and Influence on Psychometric Results. Multivariate Behavioral Research, 42(2), 387-414.

Gold, M. S., & Bentler, P. M. (2000). Treatments of Missing Data: A Monte Carlo Comparison of RBHDI, Iterative Stochastic Regression Imputation, and Expectation-Maximization. Structural Equation Modeling: A Multidisciplinary Journal, 7(3), 319-355.

Hawthorne, G., & Elliott, P. (2005). Imputing cross-sectional missing data: comparison of common techniques. Aust.N.Z.J.Psychiatry, 39(7), 583-590.

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.

SPSS Inc. (2008). SPSS Statistics for Windows (Version Version 17.0). Chicago: SPSS Inc.

SAS Institute Inc. (2011). SAS/IML software, Version 9.2 of the SAS System for Windows. Cary, NC, USA.

Janssen, K. J., Donders, A. R., Harrell, F. E., Jr., Vergouwe, Y., Chen, Q., Grobbee, D. E., & Moons, K. G. (2010). Missing covariate data in medical research: to impute is better than to ignore. J Clin Epidemiol, 63(7), 721-727.

Kneipp, S. M., & McIntosh, M. (2001). Handling missing data in nursing research with multiple imputation. [Review]. Nurs Res., 50(6), 384-389.

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.

Little, R. J., & Rubin, D. B. (2002). Statistical Analysis with Missing Data (Second Edition ed.). Hoboken, NJ: John Wiley & Sons.

Marshall, A., Altman, D. G., Royston, P., & Holder, R. L. (2010). Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study. [Comparative Study Research Support, Non-U.S. Gov't]. BMC Med Res Methodol., 10, 7.

Newman, D. A. (2003). Longitudinal Modeling with Randomly and Systematically Missing Data: A Simulation of Ad Hoc, Maximum Likelihood, and Multiple Imputation Techniques. Organizational Research Methods, 6(3), 328-362.

Pastor, J. B. N. (2003). Methods for the Analysis of Explanatory Linear Regression Models with Missing Data Not at Random. Quality & Quantity: International Journal of Methodology, 37(4), 363-376.

Roth, P. L., Switzer, F. S., & Switzer, D. M. (1999). Missing Data in Multiple Item Scales: A Monte Carlo Analysis of Missing Data Techniques. Organizational Research Methods, 2(3), 211-232.

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

Rubin, D. B. (2004). Multiple Imputation for Nonresponse in Surveys. New York: John

Wiley and Sons.

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

Siddique, J., Harel, O., & Crespi, C. M. (2012). Addressing Missing Data Mechanism Uncertainty using Multiple-Model Multiple Imputation: Application to a Longitudinal Clinical Trial. Ann Appl Stat, 6(4), 1814-1837.

StataCorp. (2011). Stata Statistical Software (Version Release 12). College Station, TX: StataCorp LP.

R Core Development Team (2012). R: A language and environment for statistical computing. Vienna, Austria.

Venables, W. N., & Ripley, B. D. (2002). Modern Applied Statistics with S. Fourth Edition. New York: Springer.

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

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.