European Journal of Developmental Psychology

ISSN: 1740-5629 (Print) 1740-5610 (Online) Journal homepage: http://www.tandfonline.com/loi/pedp20

How to handle missing data: A comparison of different approaches

Margot Peeters, Mariëlle Zondervan-Zwijnenburg, Gerko Vink & Rens van de Schoot

To cite this article: Margot Peeters, Mariëlle Zondervan-Zwijnenburg, Gerko Vink & Rens van de Schoot (2015) How to handle missing data: A comparison of different approaches, European Journal of Developmental Psychology, 12:4, 377-394, DOI: 10.1080/17405629.2015.1049526

To link to this article: http://dx.doi.org/10.1080/17405629.2015.1049526

EUROPEAN JOURNAL OF DEVELOPMENTAL PSYCHOLOGY, 2015
Vol. 12, No. 4, 377–394, http://dx.doi.org/10.1080/17405629.2015.1049526

How to handle missing data: A comparison of different approaches

Margot Peeters[1], Mariëlle Zondervan-Zwijnenburg[2], Gerko Vink[2], and Rens van de Schoot[2,3]

[1]Department of Child and Adolescent Studies, Utrecht University, Utrecht, The Netherlands
[2]Department of Method and Statistics, Utrecht University, Utrecht, The Netherlands
[3]Optentia Research Program, Faculty of Humanities, North-West University, Vanderbijlpark, South Africa

Many researchers face the problem of missing data in longitudinal research. Especially, high risk samples are characterized by missing data which can complicate analyses and the interpretation of results. In the current study, our aim was to find the most optimal and best method to deal with the missing data in a specific study with many missing data on the outcome variable. Therefore, different techniques to handle missing data were evaluated, and a solution to efficiently handle substantial amounts of missing data was provided. A simulation study was conducted to determine the most optimal method to deal with the missing data. Results revealed that multiple imputation (MI) using predictive mean matching was the most optimal method with respect to lowest bias and the smallest confidence interval (CI) while maintaining power. Listwise deletion and last observation carried backward also scored acceptable with respect to bias; however, CIs were much larger and sample size almost halved using these methods. Longitudinal research in high risk samples could benefit from using MI in future research to handle missing data. The paper ends with a checklist for handling missing data.

Keywords: missing data; multiple imputation; high risk sample; longitudinal research.

Longitudinal studies are essential for studying development in psychology, as they can be used to track individual change across the lifespan and predict

Correspondence should be addressed to Rens van de Schoot, Department of Methodology and Statistics, Utrecht University, P.O. Box 80.140, 3508 TC Utrecht, The Netherlands, Tel: +31 302534468, Fax: +31 2535797. Email: a.g.j.vandeschoot@uu.nl
No potential conflict of interest was reported by the authors.

individual differences in such changes (Biederman et al., 1997; Jackson, Sher, Cooper, & Wood, 2002; Maggs, Patrick, & Feinstein, 2008). However, following the same people for a particular period of time does not always result in valid conclusions, due to missing data. The issue of missing data occurs when longitudinal surveys fail to observe the complete set of measurements of respondents. Respondents may have (consciously or unconsciously) skipped some items or they may have missed entire waves of the study. These different types of missing data can have different missing data mechanisms (e.g., missing at random (MAR), missing not at random (MNAR, Schafer & Graham, 2002). Missing data are increasingly problematic when researchers are interested in event occurrence, or levels of change over time (Van Buuren, Boshuizen, & Knook, 1999) and can result in systematic bias (Asendorpf, van de Schoot, Denissen, & Hutteman, 2014).

There are already many textbooks and tutorial papers on how to deal with missingness (Enders, 2010; Hallgren & Witkiewitz, 2013; Jeličić, Phelps, & Lerner, 2009; Schafer & Graham, 2002). Most researchers know that not dealing with missing data may lead to invalid results. Despite all this, many researchers still ignore the issue of missing data. Therefore, the aim of the current paper is twofold: (1) provide an understandable and comprehensive overview of how and why it is important to handle missing data carefully; and (2) provide an easy-to-use checklist.

To illustrate the issue of ignoring recent developments in the field of missing data, out of the 85 prospective studies (190 papers in total) published in *Addiction* in 2012, which is the most important journal in the field of addiction, only 55 (65%) reported on how the researchers had handled missing data. Among these papers, two reported to have no missing data, four reported to used ad hoc imputation methods (e.g., replace missing values with non-use or continued use), 33 papers used listwise or casewise deletion to deal with missings (23 of these studies carried out an attrition analysis), four reported to have used the default settings of a software package, and six reported to have used full information maximum likelihood (FIML, explained in more details later). Lastly, six papers reported to have used multiple imputation (MI) techniques to deal with missing data. The other 35% of the studies ignored the topic of missing data altogether. In the current paper, we will evaluate the afore mentioned methods (i.e. with the exception of FIML) to handle missing data and highlight the pitfalls associated with each of these methods using an empirical illustration of a high relevance to addiction research: the relation between drinking onset and undercontrolled personality styles in adolescence.

Our intended audience are all researchers who use longitudinal data and encounter missingness, in particular researchers who examine high risk populations. High risk population characterize themselves by an increased vulnerability for specific behavioural outcomes. In this study, our risk population reveals an increased risk for problematic alcohol use. Specifically, when studying high risk populations, missing data are more common than in community samples and loss of participants during the study is almost unavoidable (Windle, 1990; Wolke et al., 2009). Background variables associated with missingness, often are also of interest in studying risk behaviour, and many times these variables are correlated with each other (MAR). Socioeconomic status, for instance, has been linked to substance use (Fergusson, Lynskey, & Horwood, 1996; McGee, Williams, Poulton, & Moffitt, 2000), as well as to missingness (Van Beijsterveldt et al., 2002). In addition, factors such as single-parent households and disruptive behaviour in children have been associated both with attrition (Kazdin, Mazurick, & Bass, 1993; Wolke et al., 2009) and substance use in adolescents (Barrett & Turner, 2006; Kepper, Monshouwer, van Dorsselear, & Vollebergh, 2011; Wong et al., 2006). The difficulties of having missing data that characterize these high risk populations, strikingly often reflect the reasons why particularly these groups are at risk and should receive attention from researchers. This observation is supported by several studies that highlight the importance of studying risk groups (Moffitt & Caspi, 2001) and risk factors (Hawkins, Catalano, & Miller, 1992) in relation to substance use in adolescents, but at the same time struggle with missing data (Bonomo, Bowes, Coffey, Carlin, & Patton, 2004; Carroll et al., 2012; Fergusson, Horwood, & Ridder, 2007; Grant, Stinson, & Harford, 2001; Peeters, Monshouwer, Janssen, Wiers, & Vollebergh, 2014). Van Buuren mentioned in his book "*the occurrence of missing data has long been considered as a sign of sloppy science"* (p. 5, 24). Directly related, studies with significant amount of missingness have more often been criticized by reviewers and are more difficult to publish than papers with few missing data (Van Buuren, 2012). This preoccupation will definitely be justified in certain cases, but should not lead to undervaluation of important results when missing data are handled with appropriate imputation techniques.

Present study

In what follows, we first introduce the empirical data about drinking behaviour in high risk adolescents (Peeters et al., 2015). For these data, we wanted to know what technique was the most suitable to handle the missing data. This specific question was the rationale for (1) conducting a simulation study to evaluate different techniques for handling missing data (the full syntax is provided in the online appendices) and (2) applying the various missing data techniques on the empirical data and comparing the results in a robustness investigation. We will end this study by providing some concluding remarks and a checklist to handle missing data, specifically relevant for studies with high risk populations and many dropouts.

METHOD

Sample

The methodological issue of missing data is demonstrated, using an exemplary longitudinal study carried out in The Netherlands. The sample included adolescents at-risk for heavy alcohol use and problem drinking. Adolescents were selected from 17 secondary special education (SE) schools in the Netherlands. Previous findings have revealed that particularly adolescents from SE schools are at risk for alcohol-related problems (Kepper et al., 2011). Adolescents are referred to these schools when it is no longer possible for them to attend mainstream education because of behavioural and attention problems such as seen in adolescents with attention–deficit hyperactivity disorder or conduct disorder. At baseline, 374 adolescents participated in the study (original sample), of whom 330 (88%) were boys and 44 (12%) were girls. Adolescents were between 12 and 14 years of age at the beginning of the study (mean age = 13.6). Assessments were completed at schools under the guidance of a trained research assistant. Respondents were asked to fill out a questionnaire and completed a set of cognitive tasks on the computer. In accordance with Dutch ethical standards, the voluntary nature of participation was explained, anonymity was ensured and passive parental permission was obtained through an informative letter about the purpose of the study. In total, 15 parents (3.8%) did not give their permission. During data collection, seven adolescents (1.7%) declared that they did not want to participate in the study.

The assessment took place at school, every six to eight months over a period of two years (i.e. four waves). Table 1 includes information about the sample size and missingness for each wave separately and percentage of missing data on the outcome and predictor variables in this study.

Measures

Alcohol use. The variable of interest was the onset of drinking during study follow-up (Peeters et al., 2015). Previous studies indicate that in Europe, the

TABLE 1
Missing data pattern original sample including drinkers at baseline ($N = 374$)

	Wave 1	*Wave 2*	*Wave 3*	*Wave 4*
N	374	279	230	196
Missing participants	2%	25%	39%	48%
Missing data outcome variable[a]	7.8%	29.9%	39.9%	52.0%
Missing data predictors[b]	9.4%			

[a] Quantity by frequency of alcohol (QF).
[b] sensation seeking, impulsivity.

average age of drinking onset is 13–14 (Hibell et al., 2012). In America, the age of onset is around 14–15 (Kosterman, Hawkins, Guo, Catalano, & Abbott, 2000). In Europe and America, approximately 50% of the adolescents start drinking before the age of 15. An early onset of drinking is associated with rapid escalation of alcohol use and later problem drinking (Monshouwer, Smit, de Zwart, Spruit, & van Ameijden, 2003). The adolescents in the current high risk sample reveal an increased risk for later problem drinking because of the early and heavy use of alcohol (Kepper et al., 2011).

For the onset of drinking, we used a quantity–frequency measure (QF) that assesses the number of glasses consumed during the week and weekend. Participants were asked to indicate on how many days during the week and during the weekend they regularly consumed alcohol. In addition, the number of glasses on such regular day was reported (Engels, Knibbe, & Drop, 1999). Drinking onset was based on drinking one or more glasses on either a weekend or weekday resulting in a dichotomous variable. Adolescents who already consumed alcohol at baseline were not of interest for the analysis because the onset could not be determined. This resulted in a sample of 213 adolescents for the simulation study. See Table 2 for descriptive statistics for the drinking and non-drinking sample at baseline separately.

Two undercontrolled personality styles, Impulsivity and Sensation seeking, were assessed with the Substance Use Risk Profile Scale (SURPS, Woicik, Stewart, Pihl, & Conrod, 2009). This scale includes four subscales, Sensation seeking, Impulsivity, Anxiety Sensitivty and Hopelessness. For the current study, we only used the Impulsivity and Sensation Seeking subscale. Participants could indicate how strongly they agree or disagree with 23 statements (e.g., four response categories). An example item of the Impulsivity scale is: "I usually act without stopping to think". We used a translated Dutch version of this scale (Malmberg et al., 2010). Internal reliability of the four subscales in the present

TABLE 2
Descriptive statistics for drinkers at baseline and onsetters during study follow-up separately

	Drinkers baseline Mean (SD)	*Onsetters Mean (SD)*
1. Gender (% boys)	83,5%	90.6%
2. Age	13.95 *(0.78)*	13.70 *(0.86)*
3. QF	8.24 *(10.72)*	0 *(0.00)*
3. Impulsivity	2.30 *(0.73)*	2.18 *(0.72)*
4. Sensation seeking	2.90 *(0.70)*	2.49 *(0.72)*
5. Smokers (last month) %	57%	19%
6. Cannabis users (last six months) %	38%	5%
7. Illicit drug users (lifetime) %	7%	1%

Note: QF = quantity by frequency of alcohol use; SD = standard deviation.

sample was acceptable and comparable with Woicik et al. (2009). Sensation Seeking: .68, Impulsivity: .74, Hopelessness: .81, Anxiety sensitivity: .70.

Missing data in the empirical example

This study included four waves with data collected every six to eight months. Missingness in this longitudinal study was induced by absence during a particular assessment the study. Some participants ($N = 7$; 2%) missed the first wave; however, they were allowed to participate in the following waves. Only participants who completed at least one wave were included in the analysis. We carried out an attrition analyses (original sample, $N = 374$) including several predictors of missingness besides the variables used in this particular study: gender, age, alcohol, smoking, illicit drug use, family composition (e.g., divorced), externalizing behavioural problems (assessed with the Strength and Difficulties Scale, Goodman, 1997). Attrition analyses revealed that smoking, age and alcohol use were significant predictors of missingness. Additional analysis revealed that only age remained a significant unique predictor of missingness; particularly older participants had missing data The Dutch compulsory education law mandates youth up to 16 years to attend school. This law can explain the missing data of older adolescents in our sample. It is possible that older adolescents more likely come in contact with the justice system (court visit, serve a sentence), skip school more often and therefore are more likely absent during assessment. Therefore, it seems reasonable to assume that missing data on entire waves is not related to our outcome variable (i.e., alcohol) but rather a result of drop out of older participants (e.g., MAR).

We analysed two models; first, we were interested in how well the missing data handling techniques could replicate a prespecified number of onsetters in a simulation study. Second, to predict onset, we evaluated the effect of two undercontrolled personality styles, commonly associated with alcohol use, namely sensation seeking and impulsivity assessed with the SURPS (Woicik et al., 2009). The rationale for this last model lies in the complexity of real life models. As researchers, we do not simply want to know the number/percent of drinkers in a sample, but many times want to predict this outcome by other variables.

Methods of dealing with missing data

There are many single and MI methods for handling missing data. We selected five methods, which are frequently used in psychology and were observed during our literature search in *Addiction*. The methods that will be discussed and evaluated are: listwise deletion (*method 1*), last observation carried backward (*method 2*), conservative imputation (with missing values replaced by non-drinking) (*method 3*), MI using logistic regression (*method 4*) and MI using predictive mean matching (*method 5*).

The first method, *listwise deletion*, is the easiest way to deal with missing data: an entire record is excluded if any single value is missing on one of the variables that are of interest in the analysis. Since such a procedure limits the amount of usable data, it puts a restriction on the sample size and reduces power. Moreover, it can waste a great deal of multivariate information, resulting in inefficient parameter estimates (Muthén, Kaplan, & Hollis, 1987; Wothke, 2000). When the assumption of missing completely at random is not met, this method can result in serious bias (McPherson, Barbosa-Leiker, Burns, Howell, & Roll, 2012).

The second method, *last observation carried backward*, is a manual single imputation method. A missing value is replaced with the last observed value. If data of Wave 3 were missing for participant X, but Wave 4 was observed, then the value of this wave was used to replace the missing value. Missing values at Wave 4 were not replaced. Difficulties occur when there are no observed values after the missing data, and when more than one wave is missing. Therefore, we restricted this method to impute values for cases with only one missing. Similar to listwise deletion, this method results in loss of information and decrease of power (Graham, 2009; MacCallum, Browne, & Sugawara, 1996).

The third method is conservative imputation, a common single imputation method used in dropout after treatment (Croucher et al., 2012; McPherson et al., 2012; Walker et al., 2012). Missing values are replaced with use or non-use depending on the research question. This method assumes to avoid over-estimation of observed effects, by replacing missing values in a conservative way. For instance, for studies examining success of smoking treatment, dropout before follow-up analyses is often rated as treatment failure (i.e., participant is still smoking). However, this method results in biased treatment effects (McPherson et al., 2012).

In general, it is known that single imputation procedures should be avoided, unless one specifically corrects for its variance-related problems (Rubin, 1996). Even though single imputation procedures may yield unbiased estimates, not including the uncertainty about the missing data can have a serious impact on the variance properties of the estimates. As concluded by Rubin, single imputation will lead to standard errors that are too small, which in turn may lead to finding effects that should not have been found.

The last two methods are MI techniques and have been identified as adequate methods to handle missing data (Rubin, 1996; Sterne et al., 2009). Each missing value is imputed several times to reflect uncertainty, creating several imputed data-sets (Schafer & Graham, 2002; Sterne et al., 2009). An important factor associated with reliability and accuracy of MI is selecting the appropriate MI method (e.g., logistic, linear, Bayesian linear) (Royston & White, 2011; Rubin, 1996). Fortunately, many statistical packages allow researchers to select suitable

MI methods to handle missing data. One such method is logistic regression, which is particularly suitable for incomplete binary variables.

A slightly different approach is MI using predictive mean matching (developed by Little, 1988). Missing values are imputed based on the assumption that closely similar cases provide the best information for incomplete data (Royston & White, 2011). Missing data values are replaced by values of that variable coming from cases that are most similar in multivariate space. These cases can be similar to other cases, but of course also values of previous waves of the same case. An advantage of this method is that missing data are replaced by values that actually appear in the data for that specific variable (e.g., a binary variable is replaced by either one of the two values) (Little, 1988). Previous research has shown that predictive mean matching is able to handle many types of data and does not directly depend on any distributional assumptions (Vink, Frank, Pannekoek, & Van Buuren, 2014) and can be used to prevent bias due to attrition in longitudinal designs (Asendorpf et al., 2014).

In the present study, we used the R-package (R Development Core Team, 2013) MICE (Multivariate Imputation by Chained Equations; Van Buuren & Groothuis-Oudshoorn, 2011; R package Plyr, Wickham, 2011; R package Zoo, Zeileis & Grothendieck, 2005). We chose MICE for its flexibility and customizability in, amongst others, specifying the imputation model and evaluating the imputations.

Other methods. FIML is another method frequently used to handle missing data. This method uses all information available without imputing exact values. Parameter estimation for missing values is based on ‘borrowed information’ of other observed values. Observed values contribute more to the log-likelihood than missing values (Enders & Bandalos, 2001). In case of missingness on a binary dependent variable, in this case on onset of drinking, FIML is not an available method to handle the missing data and therefore was not examined in the current study. Besides MI, FIML has proven to be successful in handling missing data, as demonstrated by a recent simulation study (Hallgren & Witkiewitz, 2013).

Simulation study

A short summary of the simulation is provided in the next paragraph and a comprehensive description can be found in the online appendix A. The simulation study consisted of four steps:

(1) Based on the original sample (e.g., using sample statistics such as means and standard deviations, see appendix A for detailed description), 500 new data-sets, without missing data, were generated (i.e., “true data-set”).

(2) Based on missing data patterns (e.g., proportion of missingness for each wave, distribution of missingness) and information from the original sample, missingness was generated in new data-sets ($N = 500$).
(3) The missing data handling methods were applied on all of the data-sets (either imputation or deletion).
(4) As a last step, the models of interest were analysed (percentage of onsetters, and the regression of sensation seeking or impulsivity on onset of drinking).

We evaluated the model using criteria derived from general suggestions by Van Buuren (Van Buuren, 2012), namely

- Remaining missing data expressed in sample size (e.g., power of study, see Table 2)
- *Percentage* of onsetters (see Table 2)
- *Confidence interval (CI) width*: 95% CIs lengths for each data-set were calculated (mean over 500 data-sets is reported), with respect to the number of onsetters. Small length CIs indicate more certainty (see Table 3)
- *Coverage percentage*: As a measure of success, we calculated what proportion (average over 500 data-sets expressed in percentage) of these CIs indeed contained the true number of onsetters that we would have found if there were no missing data (see Table 3).
- Bias indicated in mean, median, minimum and maximum bias over 500 data-sets (see Figure 1).
- Predictors' sensation seeking and impulsivity: Bias in terms of mean, median, minimum and maximum (see Figure 2), percentage of coverage and CIs width that contain the true regression coefficients of the two predictors (see Table 4).

TABLE 3
Percentage of onsetters (mean over 500 generated data-sets) and sample size of the five methods

	Percentage of onsetters	*Sample size*	*CI width*	*Coverage percentage*
"True" data-set[a]	43.36	213		
1. Listwise deletion	41.31	130	16.88	99.0
2. Last observation carried backward	41.40	145	15.99	99.8
3. Conservative	33.63	213	12.66	34.0
4. MI logistic regression	45.33	213	13.35	98.8
5. MI predictive mean matching	44.79	213	13.31	98.4

[a] New generated complete data-set based on the original sample statistics.

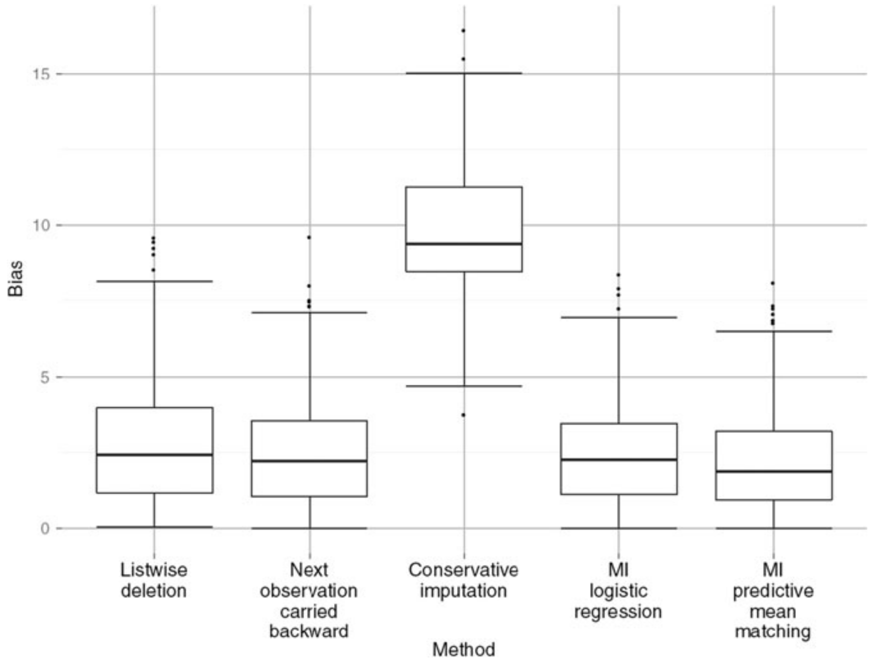

Figure 1. Graphical presentation of the mean, median, maximum and minimum bias for onset of drinking (over 500 data-sets) for each of the five methods.

Results simulation study

Imputing drinking onset. Results for the outcome variable, drinking onset, are discussed first, starting with remaining missing data, followed by percentage of onsetters, 95% CI width and coverage with respect to number of onsetters (with smaller width and higher coverage indicating more accuracy and better precision) and bias (mean, median, maximum and minimum) in onset of drinking over 500 data-sets (with lower bias indicating better approximation of the "true data set").

For listwise deletion (*method 1*), the full 39.0% of missing data remained, leaving only 130 cases in the analysis. Last observation carried backward (*method 2*) resulted in 31.9% of missing data ($N = 149$). The application of both MI methods and conservative imputation left no missing data (see Table 3).

Regarding the percentage of onsetters, the following methods hardly differed in outcome with respect to identified percentage of onsetters, namely listwise deletion (2.03% deviation), last observation carried backward (1.96% deviation) and MI (logistic regression; 1.97% and predictive mean matching; 1.42% deviation).

The 95% CI width for the number of onsetters and the coverage proportion with respect to the "true" number of onsetters were calculated. When balancing

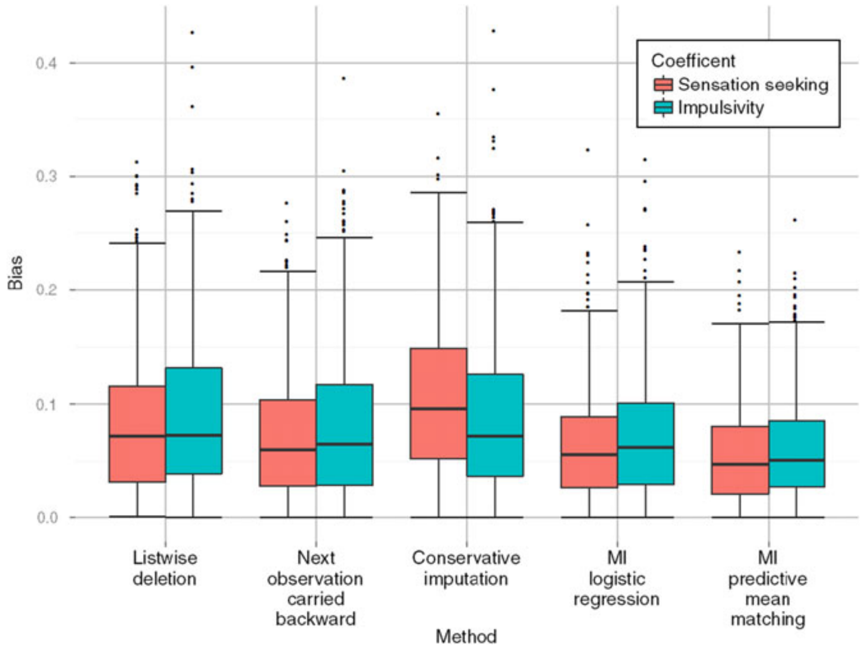

Figure 2. Graphical presentation of the mean, median, maximum and minimum bias for the regression coefficients (sensation seeking and impulsivity) over 500 data-sets, for each of the five methods.

both the width and the percentage of coverage, both MI methods (*method 4 and 5*) appeared to be the best methods (see Table 3). The conservative method revealed the smallest CI width (12.66); however, this method had the lowest coverage percentage (3%), resulting in only 3% of the CI containing the "true" percentage of onsetters. This indicates that the conservative imputation method does not properly account for the uncertainty associated with missing data and, in

TABLE 4
Regression coefficient coverage and CI width of the five methods

	CI width		*% Coverage*	
	β_1	β_2	β_1	β_2
1. Listwise deletion	0.66	0.74	100.0	99.6
2. Last observation carried backward	0.62	0.69	100.0	99.8
3. Conservative	0.53	0.59	98.4	99.4
4. MI logistic regression	0.71	0.79	100.0	100.0
5. MI predictive mean matching	0.60	0.66	100.0	100.0

Note: β_1 = sensation seeking, β_2 = impulsivity.

fact, considers its results more "certain" about the true value than it actually is. The highest coverage was obtained with last observation carried backward (*method 2*, 99.8%); however, the width of the CI interval for this method was relatively large (15.99), indicating more uncertainty about the estimate than is actually needed.

With respect to bias, the distribution of bias over all 500 sets was evaluated. The absolute lowest bias (0.0), lowest mean (2.32), lowest median (1.89) and the lowest maximum bias (8.10) on onset were obtained with MI using predictive mean matching (*method 5*, for more details, see Figure 1).

Results predictors: Sensation seeking and impulsivity. The lowest minimum and mean bias for the first regression coefficient (i.e., sensation seeking) were obtained with MI using predictive mean matching (*method 5*). Conservative imputation (*method 3*) scored worst on all distributional aspects as can be seen clearly in Figure 2. For the second regression coefficient (i.e., impulsivity), differences between methods were smaller. The best overall scores were obtained by predictive mean matching (*method 5*). MI using logistic regression (*method 4*) had scores close to that of predictive mean matching. A graphical presentation of the bias for each method and for both regression coefficients can be found in Figure 2. With respect to coverage and CI width, both the length and the width combined were most optimal for MI using predictive mean matching (see Table 4).

Results robustness investigation

The original data-set with missing data was imputed using either one of the five missing data methods. Subsequently, the percentage of onsetters was calculated, and onset of alcohol use was regressed on sensation seeking and impulsivity in a logistic regression. Table 5 summarizes the results of this robustness check. The percentage of onsetters as indicated by conservative imputation clearly deviates from that indicated by the other missing data handling methods. Irrespective of the missing data handling method, the slope of onset on sensation seeking does not significantly deviate from zero: The outcome is robust. With respect to the slope of onset on impulsivity, the significance does change as a consequence of the chosen missing data handling method: conservative imputation indicates that the slope significantly deviates from zero, while the other methods do not. The simulation study, however, indicated that PMM was the most optimal method to handle the missing data in this specific study. Hence, we adopt the results of the PMM imputation. The results using PMM as imputation method indicate that 35.3% of the adolescents start drinking during study follow-up, but indicate no significant predictive effect of sensation seeking or impulsivity on the onset of drinking.

TABLE 5
Results of the robustness investigation for each method separately

	Percentage of onsetters	*B (SE) Sensation seeking*	*B (SE) Impulsivity*
1. Listwise deletion	35.48	0.13 (0.25)	−0.36 (0.29)
2. Last observation carried backward	40.74	0.26 (0.22)	−0.44 (0.26)
3. Conservative	29.11	0.11 (0.17)	−0.46 (0.20)*
4. MI logistic regression	37.18	0.09 (0.20)	−0.30 (0.22)
5. MI predictive mean matching	35.31	0.07 (0.24)	−0.32 (0.29)

Note: $^{*}p < 0.05$.

CONCLUSION

This simulation study provided us the opportunity to choose the most optimal method to handle missing data in a real life example of missing data. Five methods were tested on several criteria, including sample size, percentage of onsetters (i.e., percentage of adolescents who started drinking during study follow-up), CI width and coverage with respect to number of onsetters and bias across 500 data-sets. The results indicate that MI is the most efficient approach to deal with missing data, yielding accurate, consistent and plausible results with excellent coverage. Performance of listwise deletion and last observation carried backward is worse than the performance of MI. Moreover, listwise deletion and last observation carried backward reduce sample size, which results in loss of power, and the use of such methods is strongly discouraged. With some caution, we can translate the results of the simulation analysis to other models, so that other researchers only need to replicate the robustness investigation for their particular study to examine whether there are differences in their conclusions caused by different imputation methods. If there are differences, then we recommend to trust the results PMM more compared to the other methods we discussed.

The current results also indicate that with a binary outcome, last observation carried backward and other single imputation methods are particularly risky because these methods impute a value that is either “yes” or “no” without reflecting any uncertainty. MI methods properly account for this uncertainty and take into consideration that the true value remains unknown. In the present example, MI, especially PMM, identifies the most optimal amount of onsetters, produces the best CI width and coverage, reveals the least bias and remains the original sample size, even when a substantial amount of data is missing (under conditions of MAR). The difference in performance between MI en PMM was small, probably resulting from the fact that the difference lies solely in the implementation of the numerical approximation of the imputation model. The flexibility in PMM lies in the way the imputations are obtained, which is different from the (often used as a standard) Bayesian normal linear regression imputation approach denoted by MI. The benefit of PMM over other MI routines lies in its distributional and robustness properties. It should be mentioned that we focused on ignorable missing data in the current study; however, in some circumstances, the missing data might be non-ignorable (MNAR). There are several methods and statistical package available to handle MNAR missing data (see Schafer & Graham, 2002 for an overview); however, it goes beyond the scope of this paper to discuss them all. With respect to the MI method used in the current study, it has been argued that when enough predictors for missingness are included in the imputation model, the assumption of MAR becomes more likely and the need for specific models under MNAR conditions is less prominent (Van Buuren et al., 1999). In addition, because we had to make a tradeoff between simplicity of the missing data model and accuracy of the model, we specified the predictors of missingness as having a normal distribution, while some of these variables realistically might be skewed and zero inflated (e.g., cannabis use, drug use). Only our outcome variable, QF, was specified as having a zero-inflated distribution. Nevertheless, the results of this simulation study show that none of the methods tested are flawless. As Orchard and Woodbury (1972, p. 697, 43) remarked: "*the best way to treat missing data, is having no missing data*". But as we have argued in the introduction, this is often not feasible in longitudinal studies, particularly not when it involves a high risk group. In such cases, using MI results in the least bias and maintains the most optimal sample size.

Checklist:

We advise always to report the answers to the following five questions:

- *Step 1.* What did you do to prevent missing data? Obviously, it is best to have no missing data. Can you recover data, for example, by approaching participants? Did you offer a reward for responding? Therefore, we stress the importance of reporting the procedure you used to prevent missing data.
- *Step 2.* How much missing data do you have? Be aware that there is a difference between the number of participants with missing data and the number of missing data points in the dataset. Report the percentage of missingness, and possibly missing data patterns.
- *Step 3.* What is the missing data mechanism? For example, create a dummy variable indicating missingness and see which variables in the data predict this missingness in a logistic regression or using other statistical tests. Report your strategy and which variables you related to missingness including statistics. Also, think about the possible mechanisms, and what could be the cause of missing data. Report your reasoning.
- *Step 4.* How did you hand the missing data? Was an imputation method used? If so, describe in detail this procedure and ideally upload the syntax used as supplementary materials.

- *Step 5.* Are your conclusions robust for different imputation methods? You might want to try several options to find out whether the main conclusions differ if different methods of dealing with missing data are used. This is called analysis robustness investigation.

Conclusion

If you are faced with missing data, be open and transparent how you dealt with missing data. Also, do not use old fashioned methods to deal with missing data because of severe influences on your conclusion.

Manuscript received 31 October 2014
Revised manuscript accepted 30 April 2015
First published online 18 June 2015

REFERENCES

Asendorpf, Jens B., van de Schoot, Rens van de, Denissen, Jaap J. A., & Hutteman, Roos (2014). Reducing bias due to systematic attrition in longitudinal studies: The benefits of multiple imputation. *International Journal of Behavioral Development*, *38*, 453–460. doi:10.1177/0165025414542713

Barrett, A. E., & Turner, R. J. (2006). Family structure and substance use problems in adolescence and early adulthood: Examining explanations for the relationship. *Addiction*, *101*, 109–120. doi:10.1111/j.1360-0443.2005.01296.x

Biederman, J., Wilens, T., Mick, E., Faraone, S. V., Weber, W., Curtis, S., … Soriano, J. (1997). Is ADHD a risk factor for psychoactive substance use disorders? findings from a four-year prospective follow-up study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*, 21–29.

Bonomo, Y. A., Bowes, G., Coffey, C., Carlin, J. B., & Patton, G. C. (2004). Teenage drinking and the onset of alcohol dependence: A cohort study over seven years. *Addiction*, *99*, 1520–1528. doi:10.1111/j.1360-0443.2004.00846.x

Carroll, K. M., Nich, C., LaPaglia, D. M., Peters, E. N., Easton, C. J., & Petry, N. M. (2012). Combining cognitive behavioral therapy and contingency management to enhance their effects in treating cannabis dependence: Less can be more, more or less. *Addiction*, *107*, 1650–1659. doi:10.1111/j.1360-0443.2012.03877.x

Croucher, R., Shanbhag, S., Dahiya, M., Kassim, S., Csikar, J., & Ross, L. (2012). Smokeless tobacco cessation in south Asian communities: A multi-centre prospective cohort study. *Addiction*, *107*, 45–52. doi:10.1111/j.1360-0443.2012.04085.x

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, *8*, 430–457. doi:10.1207/S15328007SEM0803_5

Engels, R. C. M. E., Knibbe, R. A., & Drop, M. J. (1999). Why do late adolescents drink at home? A study on psychological well-being, social integration and drinking context. *Addict Res Theory*, *7*, 31–46. doi:10.3109/16066359909004373

Fergusson, D. M., Horwood, L. J., & Ridder, E. M. (2007). Conduct and attentional problems in childhood and adolescence and later substance use, abuse and dependence: Results of a 25-year longitudinal study. *Drug and Alcohol Dependence*, *88*, S14–S26. doi:10.1016/j.drugalcdep.2006.12.011

Fergusson, D. M., Lynskey, M. T., & Horwood, L. (1996). Alcohol misuse and juvenile offending in adolescence. *Addiction*, *91*, 483–494. doi:10.1111/j.1360-0443.1996.tb02302.x

Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology Psychiatry*, *38*, 581–586. doi:10.1111/j.1469-7610.1997.tb01545.x

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549–576. doi:10.1146/annurev.psych.58.110405.085530

Grant, B. F., Stinson, F. S., & Harford, T. C. (2001). Age at onset of alcohol use and DSM-IV alcohol abuse and dependence: A 12-year follow-up. *Journal of Substance Abuse*, *13*, 493–504. doi:10.1016/S0899-3289(01)00096-7

Hallgren, K. A., & Witkiewitz, K. (2013). Missing data in alcohol clinical trials: A comparison of methods. *Alcoholism: Clinical and Experimental Research*, *37*, 2152–2160. doi:10.1111/acer.12205

Hawkins, J. D., Catalano, R. F., & Miller, J. Y. (1992). Risk and protective factors for alcohol and other drug problems in adolescence and early adulthood: Implications for substance abuse prevention. *Psychological Bulletin*, *112*, 64–105. doi:10.1037/0033-2909.112.1.64

Hibell, B., Guttormsson, U., Ahlström, S., Balakireva, O., Bjarnason, T., Kokkevi, A., & Kraus, L. (2012). The 2011 ESPAD report. *Substance use among Students In 2012*, 36.

Jackson, K. M., Sher, K. J., Cooper, M. L., & Wood, P. K. (2002). Adolescent alcohol and tobacco use: Onset, persistence and trajectories of use across two samples. *Addiction*, *97*, 517–531. doi:10.1046/j.1360-0443.2002.00082.x

Jeličić, H., Phelps, E., & Lerner, R. M. (2009). Use of missing data methods in longitudinal studies: The persistence of bad practices in developmental psychology. *Developmental Psychology*, *45*, 1195–1199. doi:10.1037/a0015665

Kazdin, A. E., Mazurick, J. L., & Bass, D. (1993). Risk for attrition in treatment of antisocial children and families. *Journal of Clinical Child Psychology*, *22*, 2–16. doi:10.1207/s15374424jccp2201_1

Kepper, A., Monshouwer, K., van Dorsselear, S., & Vollebergh, W. A. M. (2011). Substance use by adolescents in special education and residential youth care institutions. *European Child & Adolescent Psychiatry*, *20*, 311–319. doi:10.1007/s00787-011-0176-2

Kosterman, R., Hawkins, J. D., Guo, J., Catalano, R. F., & Abbott, R. D. (2000). The dynamics of alcohol and marijuana initiation: Patterns and predictors of first use in adolescence. *American Journal of Public Health*, *90*, 360–366. doi:10.2105/AJPH.90.3.360

Little, R. J. (1988). Missing-data adjustments in large surveys. *Journal of Business & Economic Statistics*, *6*, 287–296.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*, 130–149. doi:10.1037/1082-989X.1.2.130

Maggs, J. L., Patrick, M. E., & Feinstein, L. (2008). Childhood and adolescent predictors of alcohol use and problems in adolescence and adulthood in the national child development study. *Addiction*, *103*, 7–22. doi:10.1111/j.1360-0443.2008.02173.x

Malmberg, M., Overbeek, G., Monshouwer, K., Lammers, J., Vollebergh, W. A., & Engels, R. C. (2010). Substance use risk profiles and associations with early substance use in adolescence. *Journal of Behavioral Medicine*, *33*, 474–485. doi:10.1007/s10865-010-9278-4

McGee, R., Williams, S., Poulton, R., & Moffitt, T. (2000). A longitudinal study of cannabis use and mental health from adolescence to early adulthood. *Addiction*, *95*, 491–503. doi:10.1046/j.1360-0443.2000.9544912.x

McPherson, S., Barbosa-Leiker, C., Burns, G. L., Howell, D., & Roll, J. (2012). Missing data in substance abuse treatment research: Current methods and modern approaches. *Experimental and Clinical Psychopharmacology*, *20*, 243–250. doi:10.1037/a0027146

Moffitt, T. E., & Caspi, A. (2001). Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. *Development and Psychopathology*, *13*, 355–375. doi:10.1017/S0954579401002097

Monshouwer, K., Smit, F., de Zwart, W. M., Spruit, I., & van Ameijden, E. J. (2003). Progress from a first drink to first intoxication: Age of onset, time-windows and risk factors in a dutch national sample of secondary school students. *Journal of Substance Use*, *8*, 155–163. doi:10.1080/14659890310001600133

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, *52*, 431–462.

Orchard, T., & Woodbury, M. A (1972). A missing information principle: theory and applications. In *Proceedings of the 6th Berkeley Symposium on mathematical statistics and probability* (Vol. 1, pp. 697–715). Berkeley, CA: University of California Press.

Peeters, M., Janssen, T., Monshouwer, K., Boendemaker, W., Pronk, T., Wiers, R., & Vollebergh, W. (2015). Weaknesses in executive functioning predict the initiating of adolescents' alcohol use. *Developmental Cognitive Neuroscience*, in press.

Peeters, M., Monshouwer, K., Janssen, T., Wiers, R. W., & Vollebergh, W. A. (2014). Working memory and alcohol use in At-Risk adolescents: A 2-Year Follow-Up. *Alcoholism: Clinical and Experimental Research*, *38*, 1176–1183. doi:10.1111/acer.12339

R Development Core Team. (2013). A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, Retrieved from http://www.R-project.org

R Package *Plyr*: Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statiscal Software*, *40*, 1–29.

R Package Zoo: Zeileis, A., & Grothendieck, G. (2005). Zoo: S3 infrastructure for regular and irregular time series. *Journal of Statiscal Software*, *14*, 1–29.

Royston, P., & White, I. R. (2011). Multiple imputation by chained Equations (MICE): Implementation in stata. *Journal of Statistical Software*, *45*(4), 1–20.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*, *91*, 473–489. doi:10.1080/01621459.1996.10476908

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147–177. doi:10.1037/1082-989X.7.2.147

Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., … Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *BMJ: British Medical Journal*, *338*, b2393–b2393. doi:10.1136/bmj.b2393

Van Beijsterveldt, C., Van Boxtel, M., Bosma, H., Houx, P., Buntinx, F., & Jolles, J. (2002). Predictors of attrition in a longitudinal cognitive aging study: The Maastricht aging study (MAAS). *Journal of Clinical Epidemiology*, *55*, 216–223. doi:10.1016/S0895-4356(01)00473-5

Van Buuren, S. (2012). *Flexible imputation of missing data*. Boca Rotan, FL: CRC press.

Van Buuren, S., Boshuizen, H. C., & Knook, D. L. (1999). Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine*, *18*, 681–694. doi:10.1002/(SICI)1097-0258(19990330)18:6<681:AID-SIM71>3.0.CO;2-R

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). MICE: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67.

Vink, G., Frank, L. E., Pannekoek, J., & Van Buuren, S. (2014). Predictive mean matching imputation of semicontinuous variables. *Statistica Neerlandica*, *68*, 61–90. doi:10.1111/stan.12023

Walker, N., Howe, C., Bullen, C., Grigg, M., Glover, M., McRobbie, H., … Whittaker, R. (2012). The combined effect of very low nicotine content cigarettes, used as an adjunct to usual quitline care (nicotine replacement therapy and behavioural support), on smoking cessation: A randomized controlled trial. *Addiction*, *107*, 1857–1867.

Windle, M. (1990). A longitudinal study of antisocial behaviors in early adolescence as predictors of late adolescent substance use: Gender and ethnic group differences. *Journal of Abnormal Psychology*, *99*, 86–91. doi:10.1037/0021-843X.99.1.86

Woicik, P. A., Stewart, S. H., Pihl, R. O., & Conrod, P. J. (2009). The substance use risk profile scale: A scale measuring traits linked to reinforcement-specific substance use profiles. *Addictive Behaviors*, *34*, 1042–1055. doi:10.1016/j.addbeh.2009.07.001

Wolke, D., Waylen, A., Samara, M., Steer, C., Goodman, R., Ford, T., & Lamberts, K. (2009). Selective drop-out in longitudinal studies and non-biased prediction of behaviour disorders. *The British Journal of Psychiatry*, *195*, 249–256. doi:10.1192/bjp.bp.108.053751

Wong, M. M., Nigg, J. T., Zucker, R. A., Puttler, L. I., Fitzgerald, H. E., Jester, J. M., … Adams, K. (2006). Behavioral control and resiliency in the onset of alcohol and illicit drug use: A prospective study from preschool to adolescence. *Child Development*, *77*, 1016–1033. doi:10.1111/j.1467-8624.2006.00916.x

Wothke, W. (2000). Longitudinal and multigroup modeling with missing data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multiple group data: Practical issues, applied approaches and specific examples* (pp. 219–240). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.