Title

Assessing alternative imputation strategies for infrequently missing items on multi-item scales

Journal
Communications in Statistics Case Studies Data Analysis and Applications, 8(4)

ISSN
2373-7484

Authors
Rezvan, Panteha Hayati
Comulada, W Scott
Fernández, M Isabel
et al.

Publication Date
2022-10-02

DOI
10.1080/23737484.2022.2115430

Peer reviewed

HHS Public Access

Author manuscript

Commun Stat Case Stud Data Anal Appl. Author manuscript; available in PMC 2023 September 01.

Published in final edited form as:

Commun Stat Case Stud Data Anal Appl. 2022 ; 8(4): 682–713. doi:10.1080/23737484.2022.2115430.

Assessing Alternative Imputation Strategies for Infrequently Missing Items on Multi-item Scales

Panteha Hayati Rezvan[a,*], **W. Scott Comulada**[a,b], **M. Isabel Fernández**[c], **Thomas R. Belin**[a,d]

[a]Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, U.S.A.

[b]Department of Health Policy and Management, UCLA Fielding School of Public Health, Los Angeles, California, U.S.A.

[c]College of Osteopathic Medicine, Nova Southeastern University, Miami, Florida, U.S.A.

[d]Department of Biostatistics, UCLA Fielding School of Public Health, Los Angeles, California, U.S.A.

Abstract

Health-science researchers often measure psychological constructs using multi-item scales and encounter missing items on some participants. Multiple imputation (MI) has emerged as an alternative to *ad-hoc* methods (e.g., mean substitution) for handling incomplete data on multi-item scales, appealingly reflecting available information while accounting for uncertainty due to missing values in a unified inferential framework. However, MI can be implemented in a variety of ways. When the number of variables to impute gets large, some strategies yield unstable estimates of quantities of interest while others are not technically feasible to implement. These considerations raise pragmatic questions about the extent to which *ad-hoc* procedures would yield statistical properties that are competitive with theoretically motivated methods. Drawing on an HIV study where depression and anxiety symptoms are measured with multi-item scales, this empirical investigation contrasts *ad-hoc* methods for handling missing items with various MI implementations that differ as to whether imputation is at the item-level or scale-level and how auxiliary variables are incorporated. While the findings are consistent with previous reports favoring item-level imputation when feasible to implement, we found only subtle differences in statistical properties across procedures, suggesting that weaknesses of *ad-hoc* procedures may be muted when missing data percentages are modest.

[*]Corresponding Author: The University of California, Center for Community Health, 10920 Wilshire Blvd., Suite 350, Los Angeles, U.S.A. phayatirezvan@ucla.edu; Phone: (310) 794-6633.

Conflict of interest disclosures: The authors have declared that they have no competing or potential conflicts of interest in relation to the work described.

Ethical principles: The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data. The University of California Los Angeles (UCLA) Institutional Review Board approved the study (IRB #16-001674-AM-00006), and the trial was registered in www.Clinicaltrials.gov (#NCT03134833).

Declarations of interest: None

Keywords

Missing data; Multi-item scale; Multiple imputation

1. Introduction

Multi-item scales, by which we mean composite scores that are created by summing or averaging multiple self-report items that measure a common construct, are widely used in social and behavioral sciences to represent domains of interest that cannot reliably be measured by a single item. Missing data are frequently encountered on questionnaire items comprising multi-item scales, complicating statistical analysis and carrying the risk of bias and/or reduced precision if not handled properly.

Although the literature on handling missing data has grown rapidly, bolstered by the development of advanced imputation techniques for handling incomplete multivariate data, evaluation of the application of existing techniques to multi-item scales has received limited attention. In addition, researchers in the social and behavioral sciences often receive guidance to use *ad-hoc* techniques such as substitution of participant's mean score on other items for missing values, as opposed to implementing methods that explicitly account for systematic patterns in the data (e.g., higher or lower scores on certain items) as well as for variability in the data. The goal of this manuscript is to illustrate the application of different multiple imputation (MI) strategies for handling incomplete multi-item scales in the context of a study of HIV risk among youth where the analysis model of interest incorporates a moderate number of covariates and where multi-item scale scores exhibit infrequently missing data, a ubiquitous scenario in behavioral health-science research.

1.1. Common ad-hoc strategies for handling incomplete multi-item scales

Missing data on questionnaires can arise through item non-response or unit non-response. The former occurs when individuals refuse to respond to some of the items of a scale and the latter occurs when individuals fail to respond to all items within a scale. Some scoring manuals do not provide guidance on how to obtain scale scores if respondents have missing items (e.g., the 14-item Hospital Anxiety and Depression Scale (HADS; Zigmond and Snaith 1983, Bell et al. 2016) or the 7-item Generalized Anxiety Disorder scale (GAD-7; Spitzer et al. 2006)). When individuals have missing values on one or more items comprising a scale, it is a common practice to treat the scale scores as missing and use the data on individuals with a complete set of responses (termed complete-case analysis or listwise deletion). This approach restricts the statistical analysis to individuals with complete data, reducing the sample size and ordinarily resulting in loss of precision. Furthermore, complete-case analysis has the potential to induce bias in estimates of quantities of interest when individuals with incomplete observations differ systematically from those with complete data (Sterne et al. 2009, Greenland and Finkle 1995, Horton and Kleinman 2007, Little 1988).

One *ad-hoc* approach to obtain the scale score in the presence of missing items is averaging the observed items within a scale. This method, which is sometimes advised in user manuals and is widely used in practice, is also known as person-mean imputation (also referred to as "proration") since it is equivalent to replacing the missing items for each individual by the mean of an individual's observed items (Peyre, Leplège, and Coste 2011, Huisman 2000, Sijtsma and van der Ark 2003, Roth, Switzer, and Switzer 1999, Bernaards and Sijtsma 2000, Eekhout et al. 2014, Shrive et al. 2006, Bernaards and Sijtsma 1999, Schafer and Graham 2002, Hawthorne and Elliott 2005, Gmel 2001, Fayers, Curran, and Machin 1998).

Some scoring manuals (e.g., the 53-item Brief Symptom Inventory (BSI; Derogatis and Melisaratos 1983) or the 9-item Patient Health Questionnaire (PHQ-9; Kroenke et al. 2010, Kroenke, Spitzer, and Williams 2001)) outline rules treating the scale score as observed and equal to the mean of the available items if the number of observed items exceeds a specified threshold and that otherwise the scale score should be treated as missing. For instance, previous studies have replaced missing PHQ-9 item values with the mean value of the remaining items if the percentage of missing items was below 20% (Kocalevent, Hinz, and Brähler 2013, Kroenke et al. 2010) or 25% (Löwe et al. 2008) and have treated the scale score as missing if the percentage of missing items within the scale exceeded 20% or 25%.

A variant on such a threshold rule is known as the "half-rule", where missing items within a scale are replaced by the mean of the observed items if at least half of the items have been observed, with the scale score otherwise treated as missing. This approach has been applied to the 27-item Functional Assessment of Cancer Therapy (FACT-G; Fairclough and Cella 1996), the 23-item Pediatric Quality of Life Inventory (PedsQL; Varni, Burwinkle, and Seid 2006), and the 36-item Short Form Health Survey (SF-36; Ware et al. 1993).

Previous studies have advised caution with reference to mean-substitution strategies. One concern is that these strategies lack a theoretical justification either from a sampling or likelihood perspective (Schafer and Graham 2002). Another concern relates to the interpretation of constructs when rules for handling incomplete data depend on the rates and patterns of missing items. When scale scores are computed using different subsets of items on different individuals, the reliability and validity of scale-score measurements is called into question given that the scale score no longer unambiguously represents the sum or average of the items comprising the scale (Schafer and Graham 2002, Mazza, Enders, and Ruehlman 2015, Enders 2010, Lee et al. 2015, Downey and King 1998, Enders 2003). Mean-substitution strategies can lead to biased inference if the missing-data mechanism is not missing completely at random (MCAR; a scenario not expected to arise unless by design where missingness is independent of both observed and missing data) or if the items means and between-item correlation are not similar in magnitude (Enders 2003, van Ginkel, van der Ark, and Sijtsma 2007b, McDonald, Thurston, and Nelson 2000, Gmel 2001, Sijtsma and van der Ark 2003, Huisman 2000, Schafer and Graham 2002, Graham 2012, Enders 2010, Lee et al. 2015, Graham 2009).

1.2. Theoretically motivated methods for handling incomplete data

MI is an inference framework that uses standard statistical analyses that would have been conducted in the absence of missing data to "average over" uncertainty due to missing values. Introduced by Rubin (1987, 1978), the method has been elaborated and refined in myriad ways (Little and Rubin 2019, Rubin 1987, Schafer 1997, van Buuren 2018, Enders 2010, Carpenter and Kenward 2013, Schafer and Graham 2002, Su et al. 2011, Raghunathan, Berglund, and Solenberger 2018, Rubin 1996).

The method involves replacing missing values with multiple plausible values drawn independently from the posterior predictive distribution of the missing data conditional on observed data based on an appropriate statistical model (an approach that emerges naturally from a Bayesian perspective). The resulting multiple imputed datasets are then analyzed separately using statistical techniques applicable to the complete data, and the parameter estimates along with their estimated standard errors (SEs) are combined using rules that support an overall inference (Rubin 1987). Crucial to the method is an accounting for uncertainty in the missing data that combines average within-imputation variability (i.e., the squared SE of the estimate from each imputed dataset) and between-imputation variability (i.e., the sample variance of the point estimates across the datasets) (Little and Rubin 2019).

Along with MI, full information maximum likelihood (FIML) estimation (Enders 2010, Arbuckle 1996, Beale and Little 1975, Dempster, Laird, and Rubin 1977) has emerged as a successful framework for handling missing data. Like MI, the approach is often implemented under a missing at random (MAR) assumption, where the probability of data being missing is allowed to depend on the observed data but is not residually dependent on the underlying missing values (Rubin 1987, Little and Rubin 2019). Unlike MI, which is a two-stage procedure where the imputer and analyst might not be the same, FIML is ordinarily implemented in a unified manner, often using iterative numerical optimization methods (Arbuckle 1996). Because statistical findings between MI and FIML often parallel one another with multivariate normal data (e.g. Collins, Schafer, and Kam 2001) given sufficient sample sizes, we do not pursue FIML further here but anticipate that the findings reported here would be applicable to FIML approaches.

While several specialized procedures have been proposed in the literature for dealing with item-level missing data in questionnaires (van Ginkel, van der Ark, and Sijtsma 2007b, van Ginkel et al. 2010, Bernaards and Sijtsma 2000, van Ginkel et al. 2007, van Ginkel, van der Ark, and Sijtsma 2007a, Vermunt et al. 2008, Bernaards and Sijtsma 1999, van Ginkel 2010, Gmel 2001), it is appealing in applied research to have a relatively general, flexible, accessible method for producing imputations even if such an approach entails an added layer of approximation or modest loss of precision compared with methods tailored to a specific questionnaire (Hayati Rezvan, Lee, and Simpson 2015, Mackinnon 2010, Sterne et al. 2009). One approach is hot-deck imputation which is based on filling in missing values from matching subjects using an appropriate matching criterion and is often implemented using predictive mean matching (Little and Rubin 2019, Little 1988, Morris, White, and Royston 2014). Other approaches include MI via joint modeling and MI via fully conditional specification (FCS; Carpenter and Kenward 2013), also known as multivariate imputation by chained equations (MICE; van Buuren, Boshuizen, and Knook 1999, van Buuren et al. 2006, van Buuren 2016, 2018, 2015) or sequential regression multiple imputation or regression switching (Raghunathan et al. 2001). Here we focus on FCS, which specifies a sequence of overlapping regression models to impute missing values and allows each (typically univariate) regression model to be tailored to a particular variable type (e.g., binary, small count, semi-continuous) associated with the incomplete data. Such approaches have become readily accessible via widely available standard statistical software including Stata's mi impute module, the Stata user written command 'ice' (Royston and White 2011), the SAS Proc MI module, the SAS callable software application IVEware (Raghunathan, Berglund, and Solenberger 2018), the mi (Su et al. 2011) and mice (van Buuren 2018, 2021) packages in R, or the stand-alone imputation program Blimp (Keller and Enders 2019).

1.3. Previous literature on item-level and scale-level imputation

Missing data in multi-item scales can be handled at either scale-level or item-level. In the context of MI, the former treats the scale score as missing if at least one of the items comprising the scale has missing values and proceeds by deriving the scale score for cases with complete data on all the items and then imputing missing data at the scale-level for cases with partially observed items. The later begins with imputing missing data in the items of the scale prior to computing the scale score, and then deriving the scale score using the observed and imputed values of the items.

Previous studies have recommended using either item-level or scale-level MI over other missing-data handling strategies such as complete-case analysis, mean substitution, and hot-deck imputation (Huisman 2000, Bernaards et al. 2003, Burns et al. 2011, Shrive et al. 2006, Parent 2013). Belin et al. (1999) compared item-level and scale-level MI strategies empirically and found that the statistical significance of some predictors was sensitive to the choice of imputation strategy. Simulation findings have favored item-level imputation over scale-level imputation due to potential gains in precision (Gottschall, West, and Enders 2012, Eekhout et al. 2014, Simons et al. 2015). Gottschall, West, and Enders (2012) emphasized the potential for item-level MI to improve the precision of the estimates compared with scale-level MI even if the bias in downstream parameter estimates is not substantial. Eekhout et al. (2014) advised against using *ad-hoc* imputation strategies due to both bias in point estimates and underestimation of SEs with even modest amounts of missing data (e.g., a scenario where > 10% of individuals have missing data with > 25% missing items). They recommended item-level over scale-level MI regardless of missing item patterns or missing item percentages, since scale-level MI resulted in overestimation of SEs when the percentage of individuals with missing data was substantial (e.g., > 50%). Simons et al. (2015) found that item-level and scale-level MI provided similar results for large samples (> 500) with primarily unit non-response and for smaller samples (100 and 500) with a modest proportion of missing data (such as 5% or 10%) while also finding that item-level MI outperformed scale-level MI for large samples with substantial item nonresponse and for small samples with a larger proportion of missing data (e.g., 20% or 40%).

Mazza, Enders, and Ruehlman (2015) reported similar conclusions regarding the superior efficiency of handling missing data at the item level rather than the scale level using an FIML procedure incorporating a subset of the scale as additional variables in the imputation model. Nooraee et al. (2018) showed that missing data in longitudinal questionnaire outcome data can be best handled using a hybrid approach combining MI and FIML estimation (i.e., when the imputed scales are eliminated after MI if all items of that scales were originally missing) using predictive mean matching at the item-level. In contrast to prior research, Vera and Enders (2021) found that scale-level MI provided more precise estimates than item-level MI when all questionnaire items comprising a scale are missing in a longitudinal setting, with no improvement in precision observed using item-level MI when the number of items within a scale was large and the proportion of missing data was high.

1.4. Feasibility of item-level MI with larger numbers of items

In line with general guidance to avoid omitting important predictors (Rubin 1996), previous research has favored handling missing data on multi-item scale scores using item-level imputation when it is feasible to do so. As the number of items encompassed within multi-item scales increases, there is apt to be a corresponding appeal of using item-level imputation, but it might not be computationally feasible to implement established statistical methods when the number of variables grows (Nguyen, Carlin, and Lee 2021). When combined with recommendations in the literature (Graham 2012, Collins, Schafer, and Kam 2001) that it is advantageous to use an inclusive strategy incorporating all available variables in an analysis that are predictive of missingness and/or are correlated with incomplete variables, employing item-level imputation can lead to a breakdown of an imputation algorithm due to high correlations between variables (i.e., collinearity) or due to cells with zero counts in the cross-tabulations of categorical items (i.e., perfect prediction). Numerical issues might similarly arise when a large number of questionnaire items are included in the imputation model as auxiliary variables to impute missing scale scores, particularly in longitudinal studies when repeated measures of a variable require imputation, or when the number of parameters in the imputation model is larger than the sample size. Rombach et al. (2018) showed that both item-level and scale-level MI perform well for large sample sizes (≥ 500) and for small samples with $< 10\%$ of missing data, although the findings of their simulation and case study suggested that item-level MI is often infeasible and prone to convergence issues due to perfect prediction for small sample sizes with a substantial proportion of item nonresponse, particularly when the number of items increases.

1.5. Proposed solutions to incorporate item-level information when imputation model is infeasible

There has been a growing body of literature on solutions for incorporating item-level information when it becomes infeasible to fit certain types of imputation models due to there being a large number of variables. Typically, such approaches make use of dimension-reduction techniques (Enders 2010). Eekhout et al. (2015a) used a function or “parcel summary” of the observed items as auxiliary variables in a latent growth model with incomplete scale scores and showed that this approach improves the precision of the parameter estimates. The application of the method has been further illustrated using real data by Eekhout et al. (2015b).

Similarly, Howard, Rhemtulla, and Little (2015) applied principal components analysis (PCA) to reduce the number of auxiliary variables. They conducted a simulation evaluation based on a multivariate normal correlation model with an incomplete variable Y and a fully observed variable X, where the parameters of interest were marginal mean and variance of Y, as well as magnitude of correlation between X and Y. They used one principal component that contained most of the variation among all eight incomplete auxiliary variables in the missing data estimation process using FIML and found that the PCA strategy can perform as well as or even better than the inclusive strategy in terms of bias and efficiency. A recent study also favored PCA treatment of auxiliary variables over an inclusive MI strategy with an incomplete categorical variable Y, a fully observed normally distributed covariate X, along with eight continuous normally distributed auxiliary variables. They showed that the PCA approach provides less biased and more efficient results for mean and variance of Y as well as for the correlation between X and Y regardless of the number of categories of Y (Kim, Lee, and Little 2020).

Plumpton et al. (2016) proposed an adaption of MI that passively imputes scale scores after each iteration of an iterative-simulation estimation procedure. When the items of one scale are being imputed, scale scores of other scales are used as auxiliary variables for purposes of prediction instead of using all items of the other scales as predictors. Doing so incorporates item-level information in imputing missing scale scores in a manner that is feasible while simplifying imputation-model computations. Evaluations of the procedure document its feasibility and satisfactory statistical properties when a large number of variables are included. Evaluations of alternative methods by Mainzer et al. (2021) and Eekhout et al. (2018) similarly provided support for the use of scale scores, principal components or a parcel summary score as auxiliary variables in item-level MI when the inclusion of all individual items as auxiliary variables is not feasible.

1.6. The goal of the present investigation

Despite findings that MI is superior to complete-case analysis and person-mean imputation for handling incomplete multi-item scales in questionnaires, *ad-hoc* methods are still applied in many settings (Karahalios et al. 2012, Hayati Rezvan, Lee, and Simpson 2015, Eekhout et al. 2012, Bell et al. 2014, Mackinnon 2010, Powney et al. 2014, Wood, White, and Thompson 2004, Peugh and Enders 2004, Noble, Hollingworth, and Tilling 2012, Rousseau et al. 2012, Rombach et al. 2016, Schlomer, Bauman, and Card 2010). The present investigation aims to contrast alternative imputation methods in an empirical case study where scales of interest include depression symptoms measured by the PHQ-9 instrument (Kroenke, Spitzer, and Williams 2001) and generalized anxiety disorder symptoms measured by the GAD-7 instrument (Spitzer et al. 2006). While allowing for a general pattern of missing data, we are specifically interested in the impact of having a relatively modest amount of missing data on multi-item scales, which is a scenario frequently encountered in practice.

The empirical assessments investigated here use data on youth at-risk for HIV collected as part of an HIV prevention study implemented through the Adolescent Medicine Trials Network (ATN) consortium (Swendeman et al. 2019). As alternative methods for handling missing data, we implement (1) scale-level MI treating scale scores as missing if at least one of the items within the scale has missing values; (2) item-level MI including all items as auxiliary variables in the imputation model; (3) item-level MI including scale scores of other scales as auxiliary variables (i.e., “passive MI”); (4) item-level MI including principal components derived from items of other scales as auxiliary variables in the imputation model (i.e., "PCA MI"); (5) complete-case analysis; and (6) the "half-rule" method where the person-specific mean on other available items is used in place of missing items if at least half of the items on the scale are observed. The analysis of interest is to identify baseline covariates among demographic characteristics, risk behaviors, mental health summary scores, and indicators of protective acts that are predictive of internet seeking for social-service information (Comulada et al. 2021). We are specifically interested in the extent to which the above strategies would produce similar or discrepant final results, focusing on the extent to which statistical-significance conclusions regarding various predictors are affected by alternative methods for handling missing data. Of substantial interest from a pragmatic perspective is the extent to which there are any substantive differences in inferences from *ad-hoc* methods as compared to methods that have stronger theoretical motivation.

2. Background and overview on multiple imputation for multi-item scales with missing data

In this section, we first review general strategies for implementing MI and then introduce refinements of MI strategies applicable to studies involving multi-item scales.

2.1. Implementing MI via iterative algorithms

One general strategy for implementing MI is through fitting a multivariate model to incomplete data using a Markov chain Monte Carlo (MCMC) approach (Jackman 2000) such as data augmentation (Tanner and Wong 1987) or Gibbs sampling (Gelfand and Smith 1990; Casella and George 1992). Such "joint modeling" strategies translate associations in the observed portion of the data into plausible imputations that reflect those same patterns of association. Foundational methods for joint modeling that have been implemented in various statistical packages are described in Schafer (1997, 1999), Schafer and Olsen (1998), and Schafer and Yucel (2002).

Another general strategy that can be viewed as an approximation or analogy to Gibbs sampling is FCS. Although there might be incompatibilities in overlapping conditional distributions with FCS, the flexibility associated with using familiar regression models as steps within FCS and the absence of evidence that the validity of downstream inferences is substantially harmed by such incompatibilities have led to FCS being widely used in practice. While joint modeling can accommodate mixtures of incomplete continuous and categorical (i.e., binary, ordinal, and nominal) variables either through general location models (Schafer 1997) or underlying normal latent variables (Muthén and Muthén 1998-2017, Quartagno and Carpenter 2019, Quartagno and Carpenter 2020), FCS allows for a mix of variable types through the specified sequence of univariate regression models for each incomplete variable. In the analyses that follow, we implement the FCS approach to handle incomplete data in our case study.

2.2. MI strategies for handling incomplete data with more than one multi-item scale

To fix ideas for software algorithms to implement MI, consider a study where a variable O is the primary outcome of interest for a complete-data analysis. Suppose that among potential predictors of O, there are complete variables $\boldsymbol{X}$ (possibly multivariate) in addition to the incomplete predictors $Y_1, Y_2, \ldots, Y_p$ as well as two incomplete scale scores: a multi-item scale score of U, made up q items ($u_1, \ldots, u_q$); and a multi-item scale score of V, made up r items ($v_1, \ldots, v_r$). Suppose there are also s auxiliary variables ($A_1, \ldots, A_s$).

2.2.1. Scale-level MI—Imputation procedure using scale-level MI can be summarized as follows.

- Y_j ($j = 1, 2, \ldots, p$) are imputed conditioning on the observed and current imputed values of $O, Y_1, Y_2, \ldots, Y_{j-1}, Y_{j+1}, \ldots, Y_p, U, V, A_1, \ldots, A_s$, and $\boldsymbol{X}$.
- U is imputed conditioning on the observed and current imputed values of $O, Y_1, Y_2, \ldots, Y_p, V, A_1, \ldots, A_s$, and $\boldsymbol{X}$.
- V is imputed conditioning on the observed and current imputed values of $O, Y_1, Y_2, \ldots, Y_p, U, A_1, \ldots, A_s$, and $\boldsymbol{X}$.

2.2.2. Item-level MI—Imputation procedure using item-level MI can be described as follows.

- Y_j ($j = 1, 2, \ldots, p$) are imputed conditioning on the observed and current imputed values of $O, Y_1, Y_2, \ldots, Y_{j-1}, Y_{j+1}, \ldots, Y_p, U, V, A_1, \ldots, A_s$, and $\boldsymbol{X}$.
- u_i ($i = 1, 2, \ldots, q$) are imputed conditioning on the observed and current imputed values of $O, Y_1, Y_2, \ldots, Y_p, u_1, u_2, \ldots, u_{i-1}, u_{i+1} \ldots, u_q, v_1, v_2, \ldots, v_r, A_1, \ldots, A_s$, and $\boldsymbol{X}$.
- v_h ($h = 1, 2, \ldots, r$) are imputed conditioning on the observed and current imputed values of $O, Y_1, Y_2, \ldots, Y_p, u_1, u_2, \ldots, u_q, v_1, v_2, \ldots, v_{h-1}, v_{h+1} \ldots, v_r, A_1, \ldots, A_s$, and $\boldsymbol{X}$.

As noted earlier, employing detailed item-level information in imputation algorithms is not always feasible, with numerical issues arising in extreme scenarios involving a large number of questionnaire items and/or high correlations across the imputation variables. In this section, we pursue the strategies of Howard, Rhemtulla, and Little (2015) and Plumpton et al. (2016) to address associated estimation issues.

2.2.3. Passive MI - Item-level MI using scale scores of other scales as auxiliary variables—As an adaption of FCS, one can envision sampling from a sequence of conditional distributions predicting missing items within one scale using all other items of that scale as well as the scale score of other scales. Scale scores can be updated using passive imputation (van Buuren 2018) after each imputation iteration, incorporating imputed item values from the previous iteration along with updated scale scores as predictors to impute missing item values in the next imputation iteration. Using scale scores as auxiliary variables contains the size of the imputation model and can avoid statistical-computing convergence issues. It is important to note that the passively imputed scale scores must be used in the imputation model for imputation of items of other scales, otherwise convergence problems may arise due to multicollinearity between scale scores and the items comprising the same scale score. Applying this approach in the framework described above results in the following imputation procedures.

- Y_j ($j = 1, 2, \ldots, p$) are imputed conditioning on the observed and current imputed values of $O, Y_1, Y_2, \ldots, Y_{j-1}, Y_{j+1}, \ldots, Y_p, U, V, A_1, \ldots, A_s$, and $\boldsymbol{X}$.
- u_i ($i = 1, 2, \ldots, q$) are imputed conditioning on the observed and current imputed values of $O, Y_1, Y_2, \ldots, Y_p, u_1, u_2, \ldots, u_{i-1}, u_{i+1} \ldots, u_q, V, A_1, \ldots, A_s$, and $\boldsymbol{X}$.
- v_h ($h = 1, 2, \ldots, r$) are imputed conditioning on the observed and current imputed values of $O, Y_1, Y_2, \ldots, Y_p, U, v_1, v_2, \ldots, v_{h-1}, v_{h+1} \ldots, v_r, A_1, \ldots, A_s$, and $\boldsymbol{X}$.

2.2.4. PCA MI - Item-level imputation using principal components derived from items of other scales as auxiliary variables—As a general dimension-reduction strategy, PCA (Johnson and Wichern 2007, Everitt 1996) focuses on explaining the variance of a set of correlated variables through a number of independent linear combinations of the original variables (termed principal components). The choice of the number of principal components to include in an analysis can be made with the help of a scree plot displaying the proportion of the total variance explained by each principal component versus the number of principal components, based either on a gap in the proportion of variation explained, a change in the steepness of the plot, or a fixed threshold for the proportion of variance explained.

In the context of MI, PCA can be used prior to the item-level imputation process to reduce the size of the imputation model, replacing correlated items with a smaller set of uncorrelated principal components which can then be used as auxiliary variables. The need to fill in missing data on questionnaire items to implement PCA is a complication; strategies for addressing this concern include the use of multivariate normal imputation with a single imputation (Howard, Rhemtulla, and Little 2015), mean substitution, or taking a random draw from the observed marginal distribution of the same items (i.e., performing the initial step of an FCS algorithm).

Letting the principal components of the q-item scale U be represented by $(W_1, W_2, \ldots, W_k)$, letting the principal components of the r-item scale V be represented by $(Z_1, Z_2, \ldots, Z_l)$, and denoting the respective number of principal components retained as k and l, PCA-based MI can be described as follows.

- Y_j ($j = 1, 2, \ldots, p$) are imputed conditioning on the observed and current imputed values of $O, Y_1, Y_2, \ldots, Y_{j-1}, Y_{j+1}, \ldots, Y_p, U, V, A_1, \ldots, A_s$, and $\boldsymbol{X}$.
- u_i ($i = 1, 2, \ldots, q$) are imputed conditioning on the observed and current imputed values of $O, Y_1, Y_2, \ldots, Y_p, u_1, u_2, \ldots, u_{i-1}, u_{i+1} \ldots, u_q, Z_1, \ldots, Z_l, A_1, \ldots, A_s$, and $\boldsymbol{X}$.
- v_h ($h = 1, 2, \ldots, r$) are imputed conditioning on the observed and current imputed values of $O, Y_1, Y_2, \ldots, Y_p, W_1, \ldots, W_k, v_1, v_2, \ldots, v_{h-1}, v_{h+1} \ldots, v_r, A_1, \ldots, A_s$, and $\boldsymbol{X}$.

Note, after imputing missing data on the items, the scale scores are updated to be used as predictors in the imputation model of incomplete items in the next iteration.

3. Empirical illustration

Youth at-risk for HIV exposure participated in a multi-site study (ATN CARES 149) to evaluate interventions to prevent HIV infection. Outcomes of interest included HIV risk behavior (specifically condomless sex), engagement in HIV prevention activities (use of pre-exposure prophylaxis (PrEP) or post exposure prophylaxis (PEP)), as well as mental health measures, substance use, and housing insecurity. With follow-up data collection still ongoing, our empirical illustration is based on the baseline sample of 1487 youth, 14 – 24 years old, who were recruited at youth-serving agencies in high HIV prevalence neighborhoods in Los Angeles ($n = 839$) and New Orleans ($n = 647$). Additional details about study eligibility criteria and recruitment are provided in Swendeman et al. (2019).

3.1. Analysis model for HIV prevention study

The analysis model in this study was motivated by Comulada et al. (2021), where there was interest in identifying important baseline covariates (among demographic characteristics, mental health symptoms, risk behaviors, and indicators of protective acts) associated with seeking out sexual, general health, and social-service information via internet. The predictive models reported in Comulada et al. (2021) used machine learning variable selection methods including LASSO (Tibshirani 1996) and elastic net (Zou and Hastie 2005) while relying on complete-case analysis to handle missing data. In the current study, we accounted for missing-data uncertainty using a range of MI strategies, and we implemented complete-case analysis and the half-rule method for comparison.

The binary outcome of interest in this investigation was an indicator of seeking social-service information via the internet (SSI-internet), reflecting reports of using the internet to access case-work services, mental-health counselling, legal help (including information regarding updating records of one's name or gender identity), employment services, food assistance, transportation services, or other social services. A logistic regression model predicting SSI-internet considered thirty covariates (Table S1 in supplemental material), the majority of which were binary or categorical variables but also including three continuously-scaled variables: age at enrollment, GAD-7 scale score, and PHQ-9 scale score. The commonly used GAD-7 and PHQ-9 scores are based on multi-item instruments to evaluate anxiety and depression symptoms and are of central interest in the current study. The PHQ-9 is a 9-item questionnaire used to screen for depression symptoms during the past 2 weeks. An overall score is calculated by summing responses to each of nine Likert-scaled items that can be scored as 0 ("*Not at all*"), 1 ("*Several days*"), 2 ("*More than half the days*"), or 3 ("*Nearly every day*"), with higher scores indicating an increased frequency of occurrence of depression symptoms. The GAD-7 similarly utilizes 7 items to assess self-reported anxiety symptoms during the past 2 weeks, with an overall score calculated by summing all responses. In order to align the variability of regression coefficients for continuous variables with the variability of coefficients for binary covariates, we rescaled the continuous variables, dividing each by two times its standard deviation (Gelman 2008). Interest focused on inference for the coefficients of GAD-7 and PHQ-9 as well as on the marginal means of GAD-7 and PHQ-9 across different missing-data handling strategies.

3.2. Missing data in HIV prevention study

Descriptive summaries of the sample used in this case study are presented in Table S1. Eighty percent of the sample (n = 1195) had complete data on the outcome and all the variables included in the analysis model, meaning that the data on 292 individuals (20%) would be discarded if we perform a complete-case analysis. Overall, there was a general pattern of missing data without specific structure, with 9 distinct patterns seen across the analysis-model variables. Some variables in the analysis model were completely observed: assessment site, age at enrollment, sex assigned at birth, gender identity, race and ethnicity, health insurance coverage, psychiatric hospitalization, and involvement in substance abuse treatment programs. All other analysis-model variables had some missing values, with the percentage of missing data on the outcome variable around 2% and the percentages for other analysis variables ranging from 0.2% to approximately 8% (Table S1). For the GAD-7 and PHQ-9 measures in particular, 98% of individuals responded to all GAD-7 items, 97% responded to all PHQ-9 items, and the percentage of missing data was less than 2% for all specific items (Table S2).

3.3. Comparison of participants with and without complete data

In Table 1, we compared the baseline characteristics between participants who did and who did not provide data on analysis variables. There were meaningful differences between complete and incomplete cases on a number of characteristics, indicating that the participants who have complete observations on all the analysis variables would not be representative of all the participants in the study sample, and suggesting that a complete-case analysis would result in biased estimates.

3.4. Predictors of missingness in HIV prevention study

To evaluate departures from MCAR missingness, we examined the extent to which analysis variables predicted whether a case had complete measurements (Table S3). For this purpose, we considered a logistic regression model where the outcome variable was an indicator coded as 1 if at least one analysis variable was incomplete and coded 0 if all were complete. Investigating covariates one at a time, it was seen that assessment site, age at enrollment, sexual orientation, race and ethnicity, education, income, support services, health insurance coverage, GAD-7 scale score, history of PEP/PrEP use, consistent condom use, homelessness, number of sexual partners, hazardous drinking, marijuana use, and involvement in substance abuse treatment programs all were associated with missingness among analysis variables. Using the same strategy, we investigated predictors of missingness among any items of the GAD-7 and PHQ-9 scales. Incompleteness in GAD-7 items was seen to be associated with income, health insurance coverage, consistence condom use, homelessness, involvement in substance abuse treatment programs (Table S4). Incompleteness in PHQ-9 items was seen to be associated with a history of PEP/PrEP use, homelessness, sex exchange, opiates use, involvement in substance abuse treatment and programs (Table S5).

In order to identify auxiliary variables that are predictive of missingness, using the same strategy above, we examined the associations between four additional variables (i.e., emotional support, having healthcare provider, recent ER/Urgent care visit, and recent mental health outpatient care) and incompleteness among analysis variables, GAD-7 items, and PHQ-9 items. Incompleteness among GAD-7 and PHQ-9 items was seen to be associated with having healthcare provider. Correlations between items and scale scores ranged from 0.67 to 0.84 for GAD-7 and varied from 0.57 to 0.74 for PHQ-9, yielding estimates of 0.88 and 0.85, respectively for Cronbach's alpha (Table S6). Since including strong auxiliary variables in the imputation model can reduce bias and improve precision in comparison to a complete-case analysis, we examined correlations involving four potential auxiliary variables and analysis variables (Table S7), as well as among GAD-7 and PHQ-9 items (Table S8). While the correlations are not very strong, the findings suggest that including auxiliary variables in missing data models might be beneficial in predicting missing values (Collins, Schafer, and Kam 2001).

3.5. Setting up an imputation model

MI via FCS was implemented using the Stata 'ice' command (Royston and White 2011) with 100 cycles and applied to all the variables in the analysis model, as well as auxiliary variables. In one version of ice, scale-level imputations for GAD-7 and PHQ-9 were produced using predictive mean matching. Specifically, for each missing scale score, a pool of 10 candidate donors was formed from cases that had complete item data on the scale and that gave rise to a predicted scale score in the same decile as for the case with a missing scale score. Then, each missing scale score was replaced by the observed value of a randomly selected donor from the candidates in the pool.

In another version of ice, item-level imputations for GAD-7 and PHQ-9 ordinal items were produced using a sequence of ordinal logistic regression models. In addition, incomplete binary covariates were imputed using logistic regression, nominal categorical covariates (employment status and sexual orientation) were imputed using multinomial logistic regression, and ordinal categorical covariates (education level and number of sexual partners) were imputed using ordinal logistic regression models. For the item-level MI via PCA, missing item values were initially filled in using a simple hot-deck procedure, taking random draws from values of the same item observed on other study participants. After running the PCA step on the complete data, the number of principal components were chosen using scree plots, where it was noted that retaining two principal components explained 68% and 56% of the total variance in the original items of GAD-7 and PHQ-9, respectively. In the related study by Howard, Rhemtulla, and Little (2015), acceptable statistical properties were seen in downstream analyses when the proportion of variance explained by principal components used as auxiliary variables in an imputation procedure was at least 40%.

All four MI strategies used a set of thirty covariates capturing baseline characteristics including demographic, mental health, risk behaviors, and protective acts (Tables S9 - S14). The analysis-model outcome variable, SSI-internet, was also included in the imputation model to make the imputation model congenial with the analysis model (Moons et al. 2006, Meng 1994) and avoid producing biased estimates of regression coefficients. In addition to the analysis variables, four auxiliary variables, each with less than 1% missing data, were included in the imputation models to improve precision and to make the assumption of an MAR mechanism more plausible (Graham 2012, Collins, Schafer, and Kam 2001). In line with the recommendation by White, Royston, and Wood (2011) that the number of imputations should be greater than the percentage of missing data in the analysis variables, we used 25 imputations for all MI strategies.[1] Finally, to check the imputation models and assess whether the imputed data are reasonable (Nguyen, Carlin, and Lee 2017), we used graphical displays and compared the distributions of imputed values of GAD-7 and PHQ-9 scale scores obtained from the four MI strategies with the density function for the observed values of GAD-7 and PHQ-9 scale scores in complete-case analyses. All the analysis and imputation procedures were conducted in Stata SE version 16 (StataCorp. 2019).

4. Results

The estimated marginal means for the GAD-7 and PHQ-9 scale scores were similar across different missing-data handling methods (Figure S1 in supplementary materials). For prediction of SS-internet based on complete-case analysis, the half-rule method, and each of four MI strategies (scale-level MI, item-level MI, passive MI, and PCA MI) estimated odds ratios (ORs) and associated 95% confidence intervals (CIs) are presented in Figure 1 for demographic covariates, in Figure 2 for mental-health covariates and indicators of engagement in HIV prevention activity, and in Figure 3 for risk behaviors. For some covariates (other sexual orientation, having completed higher education, GAD-7 scale score, involvement in HIV prevention/intervention programs, consistent condom use, marijuana use), the CIs obtained from complete-case analysis and the half-rule were substantially wider than those obtained using the MI strategies. The SEs were nearly identical for some covariates, while for others, the SEs obtained from the MI strategies were smaller than those obtained from either complete-case analysis or the half-rule. The exceptions were for being Black/African American, PHQ-9 scale score, having 1-2 sexual partners, and hazardous drinking, where performing MI strategies led to larger SEs than a complete-case analysis (Figures S2 - S4).

Estimated ORs were essentially indistinguishable across the four MI strategies; 95% CIs were similar across MI methods for most covariates although were slightly wider for scale-level MI for some covariates. Most findings of statistical significance were also similar across methods, with the odds of SSI-internet seen to be lower among Black/African American participants and participants assigned female at birth, and the odds seen to be higher among bisexual youth, those with some higher education, and those who had received support services (Figure 1). However, the choice among the incomplete-data strategies impacted some conclusions. Specifically, being transgender/gender diverse and having health insurance coverage (Figure 1) were associated with higher odds of SSI-internet only in the MI approaches, while having higher score of GAD-7 scale (Figure 2), hazardous drinking, and marijuana use (Figure 3) were seen as significant predictors of the outcome using complete-case analysis and the half-rule. In addition, some predictors

[1]For comparative purposes, we applied the two-stage algorithm developed by von Hippel (2018) which indicates the required number of imputations ensuring replicable SEs estimates if missing data were imputed again. The algorithm suggested 8 imputed datasets were required to estimate SEs of the covariates with the desired precision. The Monte Carlo SEs for the 30 estimated regression coefficients, which indicate variability of the estimates across repeated MI procedure, showed minor variation (Footnotes of Tables S9 - S14).

showed borderline significant associations using some methods but not others (e.g., being gay/lesbian in Figure 1, PHQ-9 scale score, and involvement in HIV intervention/prevention programs in Figure 2). Density plots of the observed values (solid black line) and each of the imputed datasets (25 dashed grey lines) are shown for GAD-7 (Figure S5) and PHQ-9 (Figure S6) scale scores. A salient feature of the plots is that the multiple dotted lines reflecting the distributions emerging from predictive distributions for imputed values are more similar to one another than to the solid lines reflecting empirical distributions of variables. An implication of the predictive distributions of the missing values given observed values differing from the empirical distribution of the observed values is that the data are not MCAR. The differences between the solid line and dashed lines reflect differences in case mix between complete and incomplete cases. While all MI strategies reproduce skewness in GAD-7 and PHQ-9 as seen in individuals for whom the scale scores were observed, the item-level MI strategies exhibited more variation across imputed values and yielded distributions more similar to the observed value distributions than scale-level MI.

5. Discussion

In this paper, we investigated the extent to which *ad-hoc* techniques such as complete-case analysis and the “half-rule” approach to person-mean imputation, produced substantively different inferences compared to theoretically motivated MI strategies. We also investigated the extent to which inferences would differ across alternative MI strategies, specifically considering imputation at the item level or the scale level as well as alternative hybrid strategies for incorporating auxiliary variables. Our empirical investigation underscored how the analysis of multi-item scales scores can be complicated by even a modest number of missing item responses.

While the interpretation of the findings was often not impacted by the approach taken to address missing data, our regression analysis findings were somewhat sensitive to the choice of imputation strategy despite the small percentage of missing items on the two multi-item scale instruments of interest. For instance, the results obtained from *ad-hoc* techniques showed evidence of association between SSI-internet and GAD-7 scale score, though, no such association was observed when using the MI strategies.

As noted earlier, FIML could be considered as an alternate strategy for handling incomplete data, accommodating a range of missing data patterns and incorporating auxiliary information in models. In the sense defined by (Collins, Schafer, and Kam 2001), FIML results would mirror MI results within a multivariate normal modeling framework under the same model specification and a sufficiently large sample size. In the context of structural equation modeling (SEM), where FIML is routinely employed for addressing missing data, we would note that limitations of FIML include the challenge of developing a detailed structural equation model for item-level data with dozens of variables and the potential impact of model misspecification in generating imputations. A recent comparison of FIML and MI in the context of SEM by Lee and Shi (2021) revealed that although both procedures tended to yield equivalent results with correctly specified models, under realistic scenarios with misspecified models, FIML-based parameter estimates became more discrepant from underlying estimates (obtained from complete data analysis via the standard maximum likelihood method) with greater percentages of missingness and level of model misfit, while MI-based parameter estimates were more robust to the amount of missing data and degree of model misfit. In line with Enders and Bandalos (2001) and Enders and Mansolf (2018), we agree that further comparison of FIML and MI in SEM settings is worthy of additional research.

Previous studies using cross-sectional data have favored producing imputations at the item-level in the imputation model over strategies that collapse variables first and then attempt to handle missing data directly at the scale-level (Eekhout et al. 2014, Simons et al. 2015, Gottschall, West, and Enders 2012). While our findings aligned with conclusions from previous studies in cross-sectional settings that were favorable to item-level imputation when it is feasible to implement, we found only subtle differences in statistical properties across procedures in this empirical evaluation, suggesting that weaknesses of *ad-hoc* procedures are apt to be muted in settings where the percentage of missing data is modest. In the present study, the advantages of item-level MI were slight for some covariates and difficult to discern for other covariates. It stands to reason that item-level MI would be more impactful with increasing amounts of item non-response, but in our case study, the percentages of item-level missing data were generally modest.

In the present investigation, we recognized a distinction between performing imputation at the scale level and using collapsed versions of scales as auxiliary variables in imputation procedures. Most of our findings suggested that using scale scores as auxiliary variables in imputation models or using principal components derived from items of other scales as auxiliary variables (i.e., hybrid strategies) performed comparably to including individual items as auxiliary variables. Analyses of the PHQ-9 scale score gave rise to an exception, with passive and PCA imputation yielding associations with SSI-internet that were just barely statistically significant. Although including all available items in imputation models is considered ideal, imputation at the item level is prone to numerical issues and is sometimes not viable, particularly in settings where large numbers of questionnaire items would induce explosive numbers of parameters in models allowing general patterns of association. In such scenarios, where fitting a fully general model may be infeasible, hybrid strategies such as passive and PCA imputation emerge as practical approaches, allowing for imputation of individual items in a particular scale using either scale scores from other scales or principal components derived from items of other scales as predictors in associated imputation procedures.

In this paper, we have focused on what McNeish and Wolf (2020) call “sum scoring”, where composite variables are obtained by adding or averaging responses to multiple questionnaire items. Sum scoring has the appeal of arithmetic simplicity, but while noting that rough approximations might suffice in some contexts, McNeish and Wolf (2020) point out that when viewed within the broader arena of latent-variable modeling, the assumptions underlying sum scores correspond to model constraints that might be unnecessarily restrictive. The flexibility that accompanies latent-variable modelling might contribute to the favorable performance of the hybrid methods in our analysis; meanwhile, additional investigation is warranted to gain further insight into the psychometric properties of methods that rely on varying degrees of approximation in accounting for variation in observed data values.

Our study has a number of limitations. While we have provided a detailed illustration of the application of different MI strategies for handling incomplete multi-item scales, our findings are built on a single empirical case study. In our case study application, the amount of missing item data was modest, with 20% of individuals having missing data on some variables and with most individual items having no more than 2% missing data. In general, we would expect the impact of imputation procedures to be modest under such a scenario and to be greater when a greater proportion of cases are affected by missing data.

In the imputation procedures we implemented, we focused on additive and linear effects of predictors and did not further investigate the impact of including non-linear effects such as interactions and polynomial terms in the analysis model. Of note, FCS imputation can introduce bias in subsequent analyses when there are incompatibilities between an imputation model that omits interactive or non-linear effects and an analysis model that appropriately includes interactive or non-linear effects. Alternative strategies to accommodate missing data in interactions or polynomial effects include model-based imputation approaches (Ibrahim, Chen, and Lipsitz 2002, Ludtke, Robitzsch, and West 2020, Enders, Du, and Keller 2020, Erler et al. 2016, Kim, Belin, and Sugar 2018, Kim, Sugar, and Belin 2015) and substantive model-compatible imputation – an extension of the FCS imputation approach (Bartlett et al. 2015).

In implementing MI for GAD-7 and PHQ-9, which are comprised of ordinal items, we used ordinal logistic regression within an FCS algorithm when imputing incomplete values at the item level, and we used predictive mean matching when imputing missing scale scores. Predictive mean matching provides flexibility to reflect skewed distributions of incomplete variables, avoiding unrealistic normality assumptions for distributions for scale scores. In our application, we did not encounter the numerical issues that can arise when zero cell counts give rise to perfect prediction when fitting ordinal logistic models to item-level data; alternatives that could be considered when such concerns arise include the use of predictive mean matching for item-level imputation or imputation of scale scores through linear regression. Although different imputation procedures can give rise to similar inferences, it is also possible for such alternative implementations of MI to yield meaningfully different results.

With PCA imputation, there remains ambiguity regarding how many principal components to use as auxiliary variables in an imputation model. In the application studied here, we chose to use two principal components based on examining the scree plots, which yielded percentages of explained variability exceeding a threshold (40%) that had been identified in an earlier investigation as being associated with satisfactory statistical properties (Howard, Rhemtulla, and Little 2015). Future research could provide guidance on the implications of such decisions when implementing PCA MI. Furthermore, it would be of interest to compare PCA-based methods with machine-learning variable selection algorithms (e.g., Hastie, Tibshirani, and Friedman 2013) to assess whether certain dimension-reduction techniques have advantages when selecting auxiliary variables to be included in imputation models.

Although not typically recommended when there is evidence that the missing data mechanism departs from MCAR, complete-case analysis is still commonly used in the presence of missing data. In our evaluations, we included both complete case analysis and half-rule, another *ad-hoc* imputation method that has been recommended in user manuals for multi-item scales. In line with previous studies, we found that these *ad-hoc* methods tended to yield less-than-nominal interval-estimate coverage; however, the magnitude of the undercoverage was typically modest.

A ubiquitous concern with missing data is the prospect that patterns seen among observed data values might not carry over to unobserved data values. We kept the focus of this investigation on approaches could be expected to accommodate MAR mechanisms. The MAR assumption is often considered a reasonable starting point for studies with a substantial amount of relevant covariate information, although it remains of scientific interest to consider the robustness of inferences when missingness could be missing not at random (MNAR), where the probability of values being missing is allowed to depend on the unobserved values. Consideration of MNAR mechanisms was beyond the scope of the current paper, but it remains of interest to pursue sensitivity analyses through the use of selection modeling (Carpenter and Smuk 2021, Hayati Rezvan et al. 2015, Carpenter, Kenward, and White 2007, Beesley and Taylor 2021) or pattern mixture modeling (Tompsett et al. 2018, Hayati Rezvan, Lee, and Simpson 2018, Ratitch, O'Kelly, and Tosiello 2013, Tompsett et al. 2020).

6. Conclusions

Behavioral health-science researchers frequently use multi-item scale scores to address substantive research questions, and they are often faced with missing data problems. This research offers insight into the relative merits of scale-level, item-level, and hybrid imputation strategies, and contributes to the literature using a new dataset to illustrate applications of these imputation strategies for handling incomplete questionnaire items when inference on the scale scores is of interest. Since many user manuals of multi-item questionnaires were developed prior to wide accessibility of imputation techniques for handling incomplete multivariate data, it is important to consider whether strategies for handling missing data on multi-item scales can be improved. Our findings do not suggest that complete-case analysis and the half-rule have dramatically misleading implications when used in settings with modest amounts of missing data. While those findings are reassuring, we still caution against the use of *ad-hoc* strategies for handling missing items, especially when the rate of missing data on the items is larger than seen in the application studied here. Given that scale-level MI and item-level MI strategies yielded similar results and given that these results sometimes departed from the findings of *ad-hoc* strategies, an overarching implication of our findings is that is better to address missing data by pursuing one of the varieties of multiple imputation strategies than ignores its’ presence and perform a complete-case analysis. Meanwhile, recognizing the potential for auxiliary variables to mitigate bias and offer precision gains, hybrid strategies that incorporate information in the imputation model as auxiliary variables, whether in the form of scale scores or through principal components derived from available items, seem to be promising alternatives when including all individual items in an imputation model is infeasible.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

We would like to thank the study participants for their time commitment in participating in the Adolescent Trials Network (ATN) CARES study and acknowledge the ATN CARES Team members: Sue Ellen Abdalian, Elizabeth Mayfield Arnold, Robert Bolan, Yvonne Bryson, W. Scott Comulada, Ruth Cortado, M. Isabel Fernandez, Risa Flynn, Panteha Hayati Rezvan, Tara Kerin, Jeffrey Klausner, Marguerita Lightfoot, Norweeta Milburn, Karin Nielsen, Manuel Ocasio, Wilson Ramos, Cathy Reback, Mary Jane Rotheram-Borus, Dallas Swendeman, Wenze Tang, and Robert E. Weiss. We also thank the reviewers for improving the quality of this manuscript including a reviewer who added to our discussion linking FIML to MI for structural equation modeling.

Funding:

The following funding agencies supported the investigators to work on the topic of adolescent HIV prevention and treatment strategies: the Adolescent Medicine Trials Network (ATN) for HIV/AIDS Interventions [U19HD089886] of the Eunice Kennedy National Institute of Child Health and Human Development (NICHD) with support of the National Institute of Mental Health (NIMH), National Institute of Drug Abuse (NIDA), and National Institute on Minority Health and Health Disparities (NIMHD); National Institute of Mental Health (NIMH) [T32MH109205], the UCLA Center for HIV Identification, Prevention, and Treatment Services (CHIPTS) grant [P30MH58107], and the UCLA Clinical Translational Science Institute (CTSI) National Center for Advancing Translational Sciences of the NIH ((NIH/NCATS) grant [UL1TR001881].

Role of the Funders/Sponsors:

None of the funders or sponsors of this research had any role in the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Abbreviations

ATN	Adolescent Medicine Trials Network
CI	Confidence interval
FCS	Fully conditional specification
FIML	full information maximum likelihood
MAR	Missing at random
MCAR	Missing completely at random
MCMC	Markov chain Monte Carlo
MI	Multiple imputation
MICE	Multiple imputation by chained equations
MNAR	Missing not at random
OR	Odds ratio
PEP	Post-exposure prophylaxis
PCA	Principal components analysis

PrEP	Post-exposure prophylaxis
SE	Standard error
SEM	Structural equation modeling

References

Arbuckle JL . 1996. "Full information estimation in the presence of incomplete data." In Advanced Structural Equation Modeling, edited by Marcoulides George A. and Schumacker Randall E., 243–277. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Bartlett Jonathan W., Seaman Shaun R., White Ian R., and Carpenter James R.. 2015. "Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model." Statistical Methods in Medical Research 24 (4):462–487. [PubMed: 24525487]

Beale EML, and Little RJA. 1975. "Missing values in multivariate analysis." Journal of the Royal Statistical Society Series B (Statistical Methodology) 37 (1):129–145. doi: 10.1111/j.2517-6161.1975.tb01037.x.

Beesley Lauren J., and Taylor Jeremy M. G.. 2021. "Accounting for not-at-random missingness through imputation stacking." Statistics in Medicine. doi: 10.1002/sim.9174.

Belin TR, Datt M, Desmond K, and Ganz PA. 1999. "Comparing imputation of entire subscales versus individual items in a study of quality of life following breast cancer." Survey Research Methods Section.

Bell Melanie L., Fairclough Diane L., Fiero Mallorie H., and Butow Phyllis N.. 2016. "Handling missing items in the Hospital Anxiety and Depression Scale (HADS): a simulation study." BMC Research Notes 9 (1):479–479. doi: 10.1186/s13104-016-2284-z. [PubMed: 27770833]

Bell Melanie L., Fiero Mallorie, Horton Nicholas J., and Hsu Chiu-Hsieh. 2014. "Handling missing data in RCTs; a review of the top medical journals." BMC Medical Research Methodology 14 (1):118–118. doi: 10.1186/1471-2288-14-118. [PubMed: 25407057]

Bernaards CA, Farmer MM, Karen Qi, Dulai GS, Ganz PA, and Kahn KL. 2003. "Comparison of Two Multiple Imputation Procedures in a Cancer Screening Survey." Journal of Data Science.

Bernaards CA, and Sijtsma K. 1999. "Factor analysis of multidimensional polytomous item response data suffering from ignorable item nonresponse." Multivariate Behavioral Research 34 (3):277–313. doi: 10.1207/S15327906MBR3403_1.

Bernaards CA, and 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. doi: 10.1207/S15327906MBR3503_03. [PubMed: 26745335]

Burns Richard A., Butterworth Peter, Kiely Kim M., Bielak Allison A. M., Luszcz Mary A., Mitchell Paul, Christensen Helen, Von Sanden Chwee, and Anstey Kaarin J.. 2011. "Multiple imputation was an efficient method for harmonizing the Mini-Mental State Examination with missing item-level data." Journal of Clinical Epidemiology 64 (7):787–793. doi: 10.1016/j.jclinepi.2010.10.011. [PubMed: 21292440]

Carpenter James R., and Kenward Michael G.. 2013. Multiple imputation and its application. London, UK: John Wiley & Sons.

Carpenter James R., Kenward Michael G., and White Ian R.. 2007. "Sensitivity analysis after multiple imputation under missing at random: A weighting approach." Statistical Methods in Medical Research 16 (3):259–275. doi: 10.1177/0962280206075303. [PubMed: 17621471]

Carpenter James R., and Smuk Melanie. 2021. "Missing data: A statistical framework for practice." Biometrical Journal 63 (5):915–947. doi: 10.1002/bimj.202000196. [PubMed: 33624862]

Collins LM, Schafer JL, and Kam CM. 2001. "A comparison of inclusive and restrictive strategies in modern missing data procedures." Psychological Methods 6 (4):330–351. doi: 10.1037//1082-989X.6.4.330. [PubMed: 11778676]

Comulada WS, Goldbeck Cameron, Almirol Ellen, Gunn Heather J., Ocasio Manuel A., Fernández M. Isabel, Arnold Elizabeth Mayfield, Romero-Espinoza Adriana, Urauchi Stacey, Ramos Wilson, Rotheram-Borus Mary Jane, Klausner Jeffrey D., Swendeman Dallas, and Cares Team

Adolescent Medicine Trials Network. 2021. "Using Machine Learning to Predict Young People's Internet Health and Social Service Information Seeking." Prevention Science. doi: 10.1007/s11121-021-01255-2.

Dempster AP, Laird NM, and Rubin DB. 1977. "Maximum likelihood from incomplete data via the EM algorithm." Journal of the Royal Statistical Society Series B (Statistical Methodology) 39 (1):1–22. doi: 10.1111/j.2517-6161.1977.tb01600.x.

Derogatis Leonard R., and Melisaratos Nick. 1983. "The Brief Symptom Inventory: an introductory report." Psychological Medicine 13 (3):595–605. doi: 10.1017/S0033291700048017. [PubMed: 6622612]

Downey RG, and King C. 1998. "Missing data in Likert ratings: A comparison of replacement methods." The Journal of General Psychology 125 (2):175. [PubMed: 9935342]

Eekhout Iris, de Boer Michiel R., Twisk Jos W. R., de Vet Henrica C. W., and Heymans Martijn W.. 2012. "Missing Data A Systematic Review of How They Are Reported and Handled." Epidemiology (Cambridge, Mass.) 23 (5):729–732. doi: 10.1097/EDE.0b013e3182576cdb. [PubMed: 22584299]

Eekhout Iris, de Vet Henrica C. W., de Boer Michiel R., Twisk Jos W. R., and Heymans Martijn W.. 2018. "Passive imputation and parcel summaries are both valid to handle missing items in studies with many multi-item scales." Statistical Methods in Medical Research 27 (4):1128–1140. doi: 10.1177/0962280216654511. [PubMed: 27334917]

Eekhout Iris, de Vet Henrica C. W., Twisk Jos W. R., Brand Jaap P. L., de Boer Michiel R., and Heymans Martijn 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. doi: 10.1016/j.jclinepi.2013.09.009. [PubMed: 24291505]

Eekhout Iris, Enders Craig K., Twisk Jos W. R., de Boer Michiel R., de Vet Henrica C. W., and Heymans Martijn W.. 2015a. "Analyzing Incomplete Item Scores in Longitudinal Data by Including Item Score Information as Auxiliary Variables." Structural Equation Modeling 22 (4):588–602. doi: 10.1080/10705511.2014.937670.

Eekhout Iris, Enders Craig K., Twisk Jos W. R., de Boer Michiel R., de Vet Henrica C. W., and Heymans Martijn W.. 2015b. "Including auxiliary item information in longitudinal data analyses improved handling missing questionnaire outcome data." Journal of Clinical Epidemiology 68 (6):637–645. doi: 10.1016/j.jclinepi.2015.01.012. [PubMed: 25724894]

Enders CK, and Mansolf M. 2018. "Assessing the fit of structural equation models with multiply imputed data." Psychol Methods 23 (1):76–93. doi: 10.1037/met0000102. [PubMed: 27893216]

Enders Craig K. 2003. "Using the Expectation Maximization Algorithm to Estimate Coefficient Alpha for Scales With Item-Level Missing Data." Psychological Methods 8 (3):322–337. doi: 10.1037/1082-989X.8.3.322. [PubMed: 14596494]

Enders Craig K. 2010. Applied missing data analysis, Methodology in the social sciences series. New York, NY Guilford Press.

Enders Craig K., and Bandalos Deborah L.. 2001. "The Relative Performance of Full Information Maximum Likelihood Estimation for Missing Data in Structural Equation Models." Structural Equation Modeling: A Multidisciplinary Journal 8 (3):430–457. doi: 10.1207/S15328007SEM0803_5.

Enders Craig K., Du H, and Keller BT. 2020. "A model-based imputation procedure for multilevel regression models with random coefficients, interaction effects, and nonlinear terms." Psychological Methods 25 (1):88–112. doi: 10.1037/met0000228. [PubMed: 31259566]

Erler Nicole S., Rizopoulos Dimitris, van Rosmalen Joost, Jaddoe Vincent W. V., Franco Oscar H., and Lesaffre Emmanuel M. E. H.. 2016. "Dealing with missing covariates in epidemiologic studies: A comparison between multiple imputation and a full Bayesian approach." Statistics in Medicine 35 (17):2955–2974. doi: 10.1002/sim.6944. [PubMed: 27042954]

Everitt BS 1996. Making sense of statistics in psychology: a second level course. New York: Oxford University Press Inc.

Fairclough DL , and Cella DF. 1996. "Functional Assessment of Cancer Therapy (FACT-G): non-response to individual questions." Quality of Life Research 5:321–329. [PubMed: 8763800]

Fayers PM, Curran D, and Machin D. 1998. "Incomplete quality of life data in randomized trials: missing items." Statistics in Medicine 17 (5-7):679. [PubMed: 9549816]

Gelman Andrew. 2008. "Scaling regression inputs by dividing by two standard deviations." Statistics in Medicine 27 (15):2865–2873. doi: 10.1002/sim.3107. [PubMed: 17960576]

Gmel G 2001. "Imputation of missing values in the case of a multiple item instrument measuring alcohol consumption." Statistics in Medicine 20 (15):2369–2381. doi: 10.1002/sim.837. [PubMed: 11468769]

Gottschall Amanda C., West Stephen G., and Enders Craig K.. 2012. "A Comparison of Item-Level and Scale-Level Multiple Imputation for Questionnaire Batteries." Multivariate Behavioral Research 47 (1):1–25. doi: 10.1080/00273171.2012.640589.

Graham JW 2009. "Missing Data Analysis : Making It Work in the Real World." Annual Review of Psychology 60 (1):549–576. doi: 10.1146/annurev.psych.58.110405.085530.

Graham JW 2012. Missing data: Analysis and design. 1st ed. New York: Springer.

Greenland Sander, and Finkle William D.. 1995. "A critical look at methods for handling missing covariates in epidemiologic regression analyses." American Journal of Epidemiology 142 (12):1255–1264. doi: 10.1093/oxfordjournals.aje.a117592. [PubMed: 7503045]

Hastie Trevor, Tibshirani Robert, and Friedman Jerome. 2013. The elements of statistical learning: Data mining, inference, and prediction. New York, NY: Springer.

Hawthorne Graeme, and Elliott Peter. 2005. "Imputing cross-sectional missing data: comparison of common techniques." Australasian Psychiatry : Bulletin of the Royal Australian and New Zealand College of Psychiatrists 39 (7):583–590. doi: 10.1080/j.1440-1614.2005.01630.x.

Hayati Rezvan P , Lee KJ, and Simpson JA. 2015. "The rise of multiple imputation: A review of the reporting and implementation of the method in medical research." BMC Medical Research Methodology 15 (1):30. doi: 10.1186/s12874-015-0022-1. [PubMed: 25880850]

Hayati Rezvan P , Lee Katherine J., and Simpson JA. 2018. "Sensitivity analysis within multiple imputation framework using delta-adjustment: Application to longitudinal study of Australian children." Longitudinal and Life Course Studies 9 (3):259–278. doi: 10.14301/llcs.v9i3.503.

Hayati Rezvan P , White Ian R., Lee Katherine J., Carlin John B., and Simpson Julie A.. 2015. "Evaluation of a weighting approach for performing sensitivity analysis after multiple imputation." BMC Medical Research Methodology 15 (1). doi: 10.1186/s12874-015-0074-2.

Horton Nicholas J., and Kleinman Ken P.. 2007. "Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models." The American Statistician 61 (1):79–90. doi: 10.1198/000313007X172556. [PubMed: 17401454]

Howard WJ, Rhemtulla M, and Little TD. 2015. "Using Principal Components as Auxiliary Variables in Missing Data Estimation." Multivariate Behavioral Research 50 (3):285–299. doi: 10.1080/00273171.2014.999267. [PubMed: 26610030]

Huisman Mark. 2000. "Imputation of missing item responses: Some simple techniques." Quality & Quantity 34 (4):331–351. doi: 10.1023/A:1004782230065.

Ibrahim Joseph G., Chen Ming-Hui, and Lipsitz Stuart R.. 2002. "Bayesian methods for generalized linear models with covariates missing at random." Canadian Journal of Statistics 30 (1):55–78. doi: 10.2307/3315865.

Johnson RA , and Wichern DW. 2007. 6th ed, Applied multivariate statistical analysis. Englewood Cliffs, N.J: Prentice-Hall.

Karahalios Amalia, Baglietto Laura, Carlin John B., English Dallas R., and Simpson Julie A.. 2012. "A review of the reporting and handling of missing data in cohort studies with repeated assessment of exposure measures." BMC Medical Research Methodology 12 (1):96–96. doi: 10.1186/1471-2288-12-96. [PubMed: 22784200]

Keller Brian T., and Enders Craig K.. 2019. Blimp User’s Manual (Version 2.1). Los Angeles, CA.

Kim Soeun, Belin Thomas R., and Sugar Catherine A.. 2018. "Multiple imputation with non-additively related variables: Joint-modeling and approximations." Statistical Methods in Medical Research 27 (6):1683–1694. doi: 10.1177/0962280216667763. [PubMed: 27647811]

Kim Soeun, Sugar Catherine A., and Belin Thomas R.. 2015. "Evaluating model-based imputation methods for missing covariates in regression models with interactions." Statistics in Medicine 34 (11):1876–1888. doi: 10.1002/sim.6435. [PubMed: 25630757]

Kim Y, Lee J, and Little TD. 2020. "Multiple Imputation with Principal Components for Non-Normal Categorical Data." Multivariate Behavioral Research:1. doi: 10.1080/00273171.2020.1869516.

Kocalevent Rüya-Daniela Ph D. M. P. H., Hinz Andreas, and Brähler Elmar. 2013. "Standardization of the depression screener Patient Health Questionnaire (PHQ-9) in the general population." General Hospital Psychiatry 35 (5):551–555. doi: 10.1016/j.genhosppsych.2013.04.006. [PubMed: 23664569]

Kroenke Kurt , Spitzer Robert L., Williams Janet B. W., and Löwe Bernd. 2010. "The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review." General Hospital Psychiatry 32 (4):345–359. doi: 10.1016/j.genhosppsych.2010.03.006. [PubMed: 20633738]

Kroenke Kurt, Spitzer Robert L., and Williams Janet B. W.. 2001. "The PHQ-9: Validity of a brief depression severity measure." Journal of General Internal Medicine 16 (9):606–613. doi: 10.1046/j.1525-1497.2001.016009606.x. [PubMed: 11556941]

Lee Matthew R., Bartholow Bruce D., McCarthy Denis M., Pedersen Sarah L., and Sher Kenneth J.. 2015. "Two Alternative Approaches to Conventional Person-Mean Imputation Scoring of the Self-Rating of the Effects of Alcohol Scale (SRE)." Psychology of Addictive Behaviors 29 (1):231–236. doi: 10.1037/adb0000015. [PubMed: 25134022]

Lee T, and Shi D. 2021. "A comparison of full information maximum likelihood and multiple imputation in structural equation modeling with missing data." Psychol Methods 26 (4):466–485. doi: 10.1037/met0000381. [PubMed: 33507765]

Little Roderick J. A. 1988. "Missing-Data Adjustments in Large Surveys." Journal of Business & Economic Statistics 6 (3):287–296. doi: 10.1080/07350015.1988.10509663.

Little Roderick J. A., and Rubin Donald B.. 2019. Statistical analysis with missing data. 3rd ed. US: Wiley-Interscience.

Löwe Bernd, Spitzer Robert L., Williams Janet B. W., Mussell Monika, Schellberg Dieter, and Kroenke Kurt. 2008. "Depression, anxiety and somatization in primary care: syndrome overlap and functional impairment." General Hospital Psychiatry 30 (3):191–199. doi: 10.1016/j.genhosppsych.2008.01.001. [PubMed: 18433651]

Ludtke O, Robitzsch A, and West SG. 2020. "Regression models involving nonlinear effects with missing data: A sequential modeling approach using Bayesian estimation." Psychological Methods 25 (2):157–181. doi: 10.1037/met0000233. [PubMed: 31478719]

Mackinnon A 2010. "The use and reporting of multiple imputation in medical research - a review." Journal of Internal Medicine 268 (6):586–593. doi: 10.1111/j.1365-2796.2010.02274.x. [PubMed: 20831627]

Mainzer Rheanna, Apajee Jemishabye, Nguyen Cattram D., Carlin John B., and Lee Katherine J.. 2021. "A comparison of multiple imputation strategies for handling missing data in multi-item scales: Guidance for longitudinal studies." Statistics in Medicine. doi: 10.1002/sim.9088.

Mazza Gina L., Enders Craig K., and Ruehlman Linda S.. 2015. "Addressing Item-Level Missing Data: A Comparison of Proration and Full Information Maximum Likelihood Estimation." Multivariate Behavioral Research 50 (5):504–519. doi: 10.1080/00273171.2015.1068157. [PubMed: 26610249]

McDonald Robert A., Thurston Paul W., and Nelson Mark R.. 2000. "A Monte Carlo Study of Missing Item Methods." Organizational Research Methods 3 (1):71–92. doi: 10.1177/109442810031003.

McNeish Daniel, and Wolf Melissa Gordon. 2020. "Thinking twice about sum scores." Behavior research methods 52 (6):2287–2305. doi: 10.3758/s13428-020-01398-0. [PubMed: 32323277]

Meng Xiao-Li. 1994. "Multiple-imputation inferences with uncongenial sources of input." Statistical Science 9 (4):538–558. doi: 10.1214/ss/1177010269.

Moons Karel G. M., Donders Rogier A. R. T., Stijnen Theo, and Harrell Frank E.. 2006. "Using the outcome for imputation of missing predictor values was preferred." Journal of Clinical Epidemiology 59 (10):1092–1101. doi: 10.1016/j.jclinepi.2006.01.009. [PubMed: 16980150]

Morris Tim P., White Ian R., and Royston Patrick. 2014. "Tuning multiple imputation by predictive mean matching and local residual draws." BMC Medical Research Methodology 14 (1):75–75. doi: 10.1186/1471-2288-14-75. [PubMed: 24903709]

Mplus user’s guide. Muthén & Muthén., Los Angeles, CA.

Nguyen Cattram D., Carlin John B., and Lee Katherine J.. 2017. "Model checking in multiple imputation: an overview and case study." Emerging Themes in Epidemiology 14 (1):8–8. doi: 10.1186/s12982-017-0062-6. [PubMed: 28852415]

Nguyen Cattram D., Carlin John B., and Lee Katherine J.. 2021. "Practical strategies for handling breakdown of multiple imputation procedures." Emerging Themes in Epidemiology 18 (1):5–5. doi: 10.1186/s12982-021-00095-3. [PubMed: 33794933]

Noble Sian Marie, Hollingworth William, and Tilling Kate. 2012. "Missing data in trial-based cost-effectiveness analysis: the current state of play." Health Economics 21 (2):187–200. doi: 10.1002/hec.1693. [PubMed: 22223561]

Nooraee Nazanin, Molenberghs Geert, Ormel Johan, and van den Heuvel Edwin R.. 2018. "Strategies for handling missing data in longitudinal studies with questionnaires." Journal of Statistical Computation and Simulation 88 (17):3415–3436. doi: 10.1080/00949655.2018.1520854.

Parent Mike C. 2013. "Handling Item-Level Missing Data: Simpler Is Just as Good." The Counseling Psychologist 41 (4):568–600. doi: 10.1177/0011000012445176.

Peugh James L., and Enders Craig K.. 2004. "Missing Data in Educational Research: A Review of Reporting Practices and Suggestions for Improvement." Review of Educational Research 74 (4):525–556. doi: 10.3102/00346543074004525.

Peyre Hugo, Leplège Alain, and Coste Joël. 2011. "Missing data methods for dealing with missing items in quality of life questionnaires. A comparison by simulation of personal mean score, full information maximum likelihood, multiple imputation, and hot deck techniques applied to the SF-36 in the French 2003 decennial health survey." Quality of life research 20 (2):287–300. doi: 10.1007/s11136-010-9740-3. [PubMed: 20882358]

Plumpton Catrin O., Morris Tim, Hughes Dyfrig A., and White Ian R.. 2016. "Multiple imputation of multiple multi-item scales when a full imputation model is infeasible." BMC Research Notes 9 (44):45–45. doi: 10.1186/s13104-016-1853-5. [PubMed: 26809812]

Powney Matthew, Williamson Paula, Kirkham Jamie, and Kolamunnage-Dona Ruwanthi. 2014. "A review of the handling of missing longitudinal outcome data in clinical trials." Trials 15 (1):237–237. doi: 10.1186/1745-6215-15-237. [PubMed: 24947664]

Package ‘jomo' (Package).

Quartagno Matteo, and Carpenter James R.. 2019. "Multiple imputation for discrete data: Evaluation of the joint latent normal model." Biometrical Journal 61 (4):1003–1019. doi: 10.1002/bimj.201800222. [PubMed: 30868652]

Raghunathan TE, Berglund Patricia A., and Solenberger Peter W.. 2018. Multiple imputation in practice: With examples using IVEware. 2 ed. Vol. 1: CRC Press.

Raghunathan TE, Lepkowski JM, Hoewyk JV, and Solenberger P. 2001. "A multivariate technique for multiply imputing missing values using asequence of regression models." Survey Methodology 27:85–95.

Ratitch Bohdana, O'Kelly Michael, and Tosiello Robert. 2013. "Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models." Pharmaceutical Statistics 12 (6):337. [PubMed: 23292975]

Rombach Ines, Gray Alastair M., Jenkinson Crispin, Murray David W., and Rivero-Arias Oliver. 2018. "Multiple imputation for patient reported outcome measures in randomised controlled trials: advantages and disadnantages of imputing at the item, subscale or composite score level." BMC Medical Research Methodology 18 (1):87. [PubMed: 30153796]

Rombach Ines, Rivero-Arias Oliver, Gray Alastair M., Jenkinson Crispin, and Burke Órlaith. 2016. "The current practice of handling and reporting missing outcome data in eight widely used PROMs in RCT publications: a review of the current literature." Quality of Life Research 25 (7):1613–1623. doi: 10.1007/s11136-015-1206-1. [PubMed: 26821918]

Roth Philip L., Switzer Fred S., and Switzer Deborah M.. 1999. "Missing Data in Multiple Item Scales: A Monte Carlo Analysis of Missing Data Techniques." Organizational Research Methods 2 (3):211–232. doi: 10.1177/109442819923001.

Rousseau Michel, Rousseau Michel, Simon Marielle, Simon Marielle, Bertrand Richard, Bertrand Richard, Hachey Krystal, and Hachey Krystal. 2012. "Reporting missing data: a study of selected articles published from 2003–2007." Quality & Quantity 46 (5):1393–1406. doi: 10.1007/s11135-011-9452-y.

Royston Patrick, and White Ian R.. 2011. "Multiple Imputation by Chained Equations (MICE): Implementation in Stata." Journal of Statistical Software 45 (4):1–20.

Rubin Donald B. 1978. "Multiple imputations in sample surveys - A phenomenological Bayesian approach to nonresponse." Survey Research Methods Section, American Statistical Association (ASA)

Rubin Donald B. 1987. Multiple imputation for nonresponse in surveys. 1st ed, Wiley series in probability and mathematical statistics: Applied probability and statistics. New York: Wiley.

Rubin Donald B. 1996. "Multiple imputation after 18+ years." Journal of the American Statistical Association 91:473–489.

Schafer Joseph L. 1997. Analysis of incomplete multivariate data: Chapman and Hall.

Schafer Joseph L. 1999. "Multiple imputation: A primer." Statistical Methods in Medical Research 8 (1):3–15. doi: 10.1191/096228099671525676. [PubMed: 10347857]

Schafer Joseph L., and Graham John W.. 2002. "Missing data: Our view of the state of the art." Psychological Methods 7 (2):147–177. doi: 10.1037/1082-989X.7.2.147. [PubMed: 12090408]

Schafer Joseph L., and Olsen MK. 1998. "Multiple imputation for multivariate missing-data problems: A data analyst's perspective." Multivariate Behavioral Research 33 (4):545–571. doi: 10.1207/s15327906mbr3304_5. [PubMed: 26753828]

Schafer Joseph L., and Yucel Recai M.. 2002. "Computational Strategies for Multivariate Linear Mixed-Effects Models With Missing Values." Journal of Computational and Graphical statistics 11 (2):437–457. doi: 10.1198/106186002760180608.

Schlomer Gabriel L., Bauman Sheri, and Card Noel A.. 2010. "Best Practices for Missing Data Management in Counseling Psychology." Journal of Counseling Psychology 57 (1):1–10. doi: 10.1037/a0018082. [PubMed: 21133556]

Shrive Fiona M., Stuart Heather, Quan Hude, and Ghali William A.. 2006. "Dealing with missing data in a multi-question depression scale: a comparison of imputation methods." BMC Medical Research Methodology 6 (1):57–57. doi: 10.1186/1471-2288-6-57. [PubMed: 17166270]

Sijtsma K, and van der Ark LA. 2003. "Investigation and treatment of missing item scores in test and questionnaire data." Multivariate Behavioral Research 38 (4):505–528. doi: 10.1207/s15327906mbr3804_4. [PubMed: 26777444]

Simons Claire L., Rivero-Arias Oliver, Yu Ly-Mee, and Simon Judit. 2015. "Multiple imputation to deal with missing EQ-5D-3L data: Should we impute individual domains or the actual index?" Quality of Life Research 24 (4):805–815. doi: 10.1007/s11136-014-0837-y. [PubMed: 25471286]

Spitzer Robert L., Kroenke Kurt, Williams Janet B. W., and Löwe Bernd. 2006. "A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7." Archives of Internal Medicine 166 (10):1092–1097. doi: 10.1001/archinte.166.10.1092. [PubMed: 16717171]

Stata Statistical Software: Release 16, College Station, TX: Stata Corp LLC.

Sterne Jonathan A. C., White Ian R., Carlin John B., Spratt Michael, Royston Patrick, Kenward Michael G., Wood Angela M., and Carpenter James R.. 2009. "Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls." BMJ 338:b2393. doi: 10.1136/bmj.b2393. [PubMed: 19564179]

Su Yu-Sung, Gelman Andrew, Hill Jennifer, and Yajima Masanao. 2011. "Multiple Imputation with Diagnostics (mi) in R: Opening Windows into the Black Box." Journal of Statistical Software 45 (2):1–31.

Swendeman Dallas, Elizabeth Mayfield Arnold Danielle Harris, Fournier Jasmine, Comulada W. Scott, Reback Cathy, Koussa Maryann, Ocasio Manuel, Lee Sung-Jae, Kozina Leslie, Fernández Maria Isabel, Rotheram Mary Jane, and Cares Team Adolescent Medicine Trials Network. 2019. "Text-messaging, online peer support group, and coaching strategies to optimize the HIV prevention continuum for youth: Protocol for a randomized controlled trial." JMIR Research Protocols 8 (8):e11165. [PubMed: 31400109]

Tibshirani Robert. 1996. "Regression shrinkage and selection via the lasso." Journal of the Royal Statistical Society Series B (Statistical Methodology) 58 (1):267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x.

Tompsett Daniel, Leacy F, Moreno-Betancur M, Heron J, and White IR. 2018. "On the use of the not-at-random fully conditional specification (NARFCS) procedure in practice." Statistics in Medicine 37 (15):2338–2353. doi: 10.1002/sim.7643. [PubMed: 29611205]

Tompsett Daniel, Sutton Stephen, Seaman Shaun R., and White Ian R.. 2020. "A general method for elicitation, imputation, and sensitivity analysis for incomplete repeated binary data." Statistics in Medicine 39 (22):2921–2935. doi: 10.1002/sim.8584. [PubMed: 32677726]

van Buuren S 2015. "Chapter 13: Fully Conditional Specification." In Handbook of Missing Data Methodology, 267–294. Chapman & Hall/CRC.

van Buuren S 2016. "Multiple imputation of discrete and continuous data by fully conditional specification." Statistical Methods in Medical Research 16 (3):219–242. doi: 10.1177/0962280206074463.

van Buuren S 2018. Flexible imputation of missing data. 2nd ed: CRC Press.

Multivariate Imputation by Chained Equations 3.13.0 (Package).

van Buuren S, Boshuizen HC, and Knook DL. 1999. "Multiple imputation of missing blood pressure covariates in survival analysis." Statistics in Medicine 18 (6):681–694. [PubMed: 10204197]

van Buuren S, Brand JPL, Groothuis-Oudshoorn C, and Rubin DB. 2006. "Fully conditional specification in multivariate imputation." Journal of Statistical Computation & Simulation 76 (12):1049–1064. doi: 10.1080/10629360600810434.

van Ginkel JR 2010. "Investigation of multiple imputation in low-quality questionnaire data." Multivariate Behavioral Research 45 (3):574–598. doi: 10.1080/00273171.2010.483373. [PubMed: 26760493]

van Ginkel JR, Sijtsma K, van der Ark LA, and Vermunt JK. 2010. "Incidence of Missing Item Scores in Personality Measurement, and Simple Item-Score Imputation." Methodology 6 (1):17–30. doi: 10.1027/1614-2241/a000003.

van Ginkel JR, van der Ark LA, and Sijtsma K. 2007a. "Multiple imputation of item scores in test and questionnaire data, and influence on psychometric results." Multivariate Behavioral Research 42 (2):387–414. doi: 10.1080/00273170701360803. [PubMed: 26765492]

van Ginkel JR, van der Ark LA, and Sijtsma K. 2007b. "Multiple imputation of item scores when test data are factorially complex." British Journal of Mathematical & Statistical Psychology 60 (2):315–337. doi: 10.1348/000711006X117574. [PubMed: 17971272]

van Ginkel JR, van der Ark LA, Sijtsma K, and Vermunt JK. 2007. "Two-way imputation: A Bayesian method for estimating missing scores in tests and questionnaires, and an accurate approximation." Computational Statistics & Data Analysis 51 (8):4013–4027. doi: 10.1016/j.csda.2006.12.022.

Varni JW, Burwinkle TM, and Seid M. 2006. "The PedsQL (TM) 4.0 as a school population health measure: Feasibility, reliability, and validity." Quality of Life Research 15 (2):203–215. doi: 10.1007/s11136-005-1388-z. [PubMed: 16468077]

Vera Juan Diego, and Enders Craig K.. 2021. "Is item imputation always better? An investigation of wave-missing data in growth models." Structural Equation Modeling. doi: 10.1080/10705511.2020.1850289.

Vermunt Jeroen K., van Ginkel Joost R, van der Ark L. Andries, and Sijtsma Klaas. 2008. "Multiple Imputation of Incomplete Categorical Data Using Latent Class Analysis." Sociological Methodology 38 (1):369–397. doi: 10.1111/j.1467-9531.2008.00202.x.

von Hippel Paul T. 2018. "How many imputations do you need? A two-stage calculation using a quadratic rule." Sociological Methods & Research 49 (3):004912411774730–718. doi: 10.1177/0049124117747303.

Ware JE Jr. , Snow KK, Kosinski M, and Gandek B. 1993. SF-36 health survey manual and interpretation guide. Boston: New England Medical Centre.

White IR, Royston P, and Wood AM. 2011. "Multiple imputation using chained equations: Issues and guidance for practice." Statistics in Medicine 30 (4):377–399. doi: 10.1002/sim.4067. [PubMed: 21225900]

Wood Angela M., White Ian R., and Thompson Simon G.. 2004. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. Thousand Oaks, CA: Sage Publications.

Zigmond AS, and Snaith RP. 1983. "The hospital anxiety and depression scale." Acta Psychiatrica Scandinavica 67 (6):361–370. [PubMed: 6880820]

Zou Hui, and Hastie Trevor. 2005. "Regularization and variable selection via the elastic net." Journal of the Royal Statistical Society Series B (Statistical Methodology) 67 (2):301–320. doi: 10.1111/j.1467-9868.2005.00503.x.

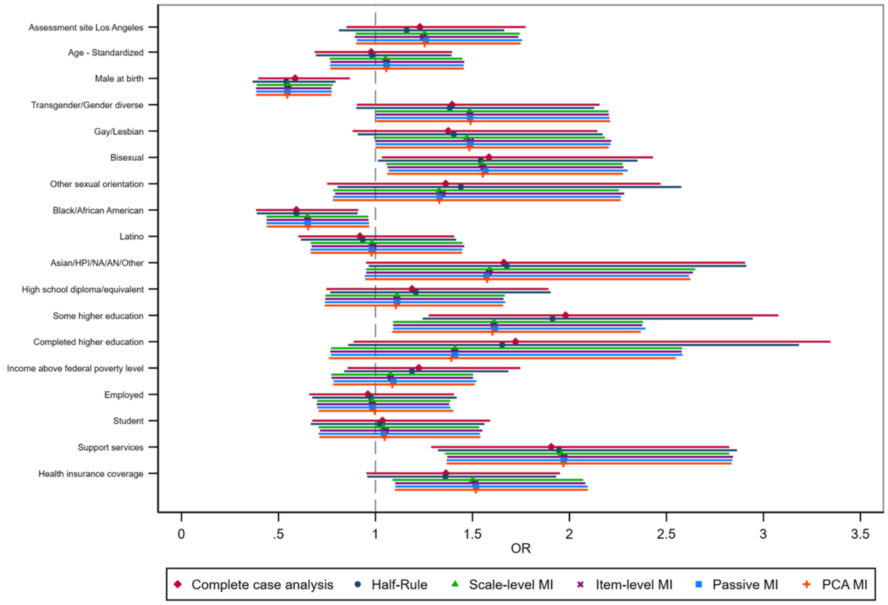

Figure 1.

Estimated ORs and 95% CIs for regression coefficients of demographic predictors of internet use for social services across a complete case analysis, *ad-hoc* half-rule, scale-level MI, item-level MI, item-level MI via passive imputation, and item-level MI via PCA imputation.

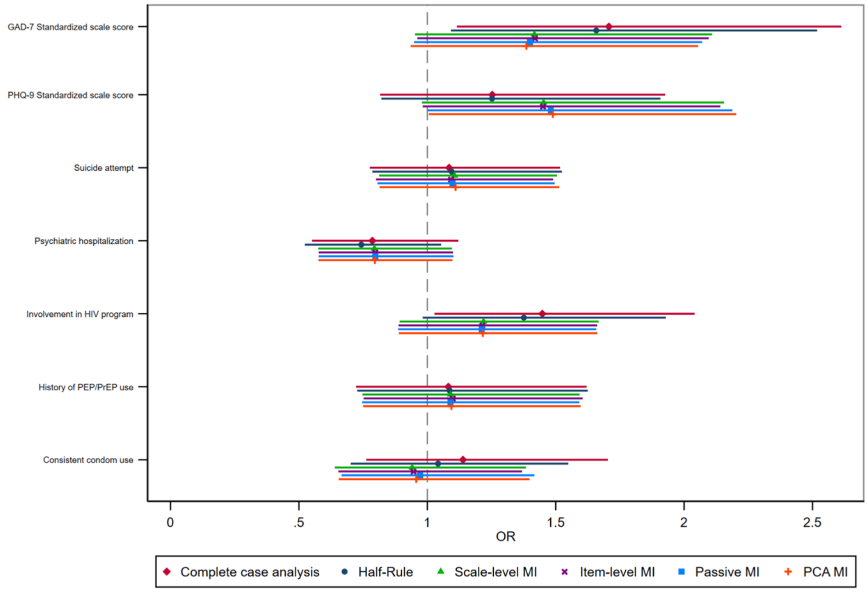

Figure 2.
Estimated ORs and 95% CIs for regression coefficients of mental health and engagement in HIV prevention predictors of internet use for social services across a complete case analysis, *ad-hoc* half-rule, scale-level MI, item-level MI, item-level MI via passive imputation, and item-level MI via PCA imputation.

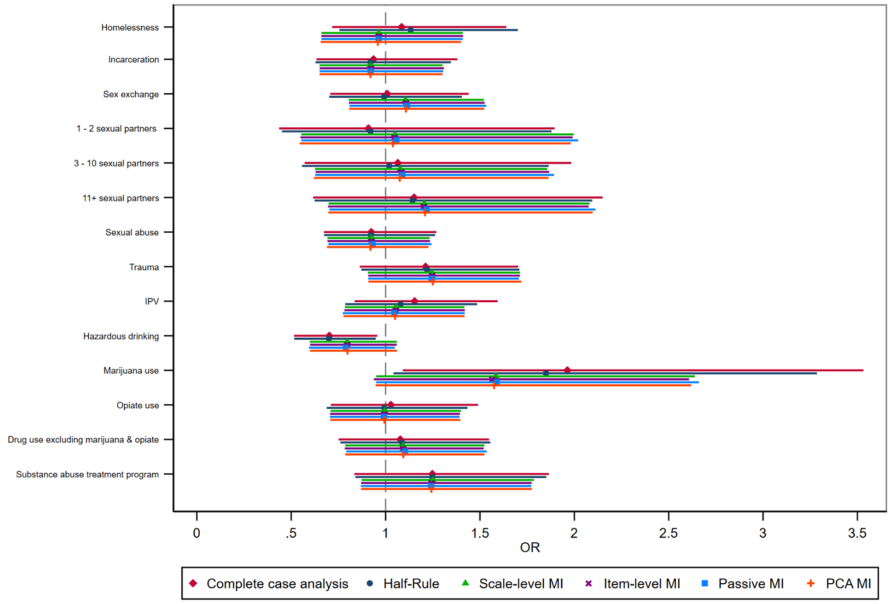

Figure 3.

Estimated ORs and 95% CIs for regression coefficients of HIV risk predictors of internet use for social services across a complete case analysis, *ad-hoc* half-rule, scale-level MI, item-level MI, item-level MI via passive imputation, and item-level MI via PCA imputation.

Table 1

Comparison of participants with and without complete data in the analysis variables

Variables	Complete cases		Incomplete cases	
	n	**%**	***n***	**%**
Outcome variable				
Internet use for social services[1]	307	25.7	58	22.3
Predictors				
Demographic				
Age**	21.0	2.1	20.6	2.2
Sex assigned at birth**				
Female	215	18.0	62	21.2
Male	980	82.0	230	78.8
Gender identity				
Cisgender	1035	86.6	255	87.3
Transgender/Gender diverse	160	13.4	37	12.7
Sexual orientation***				
Heterosexual	280	23.4	121	42.8
Gay or lesbian	510	42.7	92	32.5
Bisexual	301	25.2	41	14.5
Other sexual orientation	104	8.7	29	10.3
Race & Ethnicity***				
Black/African American	523	43.8	176	60.3
Latino	358	30.0	57	19.5
White	217	18.2	34	11.6
Asian/HPI/NA/AN/Other	97	8.1	25	8.6
Assessment site***				
Los Angeles	702	58.7	136	46.6
New Orleans	493	41.3	156	53.4
Education***				
Below high school (HS)	255	21.3	94	34.6

Variables	Complete cases		Incomplete cases	
	n	**%**	***n***	**%**
HS diploma/equivalent	299	25.0	80	29.5
Some higher education (HE)	521	43.6	84	30.9
Completed HE	120	10.0	14	5.2
Income above the federal poverty level	361	30.2	63	22.3
Employment				
Employed	531	44.4	118	45.2
Student	329	27.5	63	24.1
Unemployed	335	28.0	80	30.7
Support services **	582	48.7	163	56.8
Health insurance coverage ***	906	75.8	183	62.7
<u>Mental Health</u>				
GAD-7 scale score **	6.7	5.4	5.7	5.9
PHQ-9 scale score	7.2	5.7	6.5	6.5
Suicide attempt	406	34.0	78	30.6
Psychiatric hospitalization	348	29.1	93	31.9
<u>Engagement in HIV Prevention</u>				
Involvement in HIV prevention program *	249	20.8	61	21.2
History of PEP/PrEP use ***	188	15.7	22	8.0
Consistent condom use with all partners ***	239	20.0	73	40.3
<u>Risk Behaviors and Protective Acts</u>				
Homelessness ***	543	45.4	178	66.7
Incarceration	285	23.9	81	28.3
Sex exchange	289	24.2	74	26.1
Sexual partners **				
None	88	7.4	33	11.5
1 – 2	124	10.4	32	11.2
3 – 10	480	40.2	127	44.4
11 or more	503	42.1	94	32.9
Sexual abuse	551	46.1	120	42.1

Variables	Complete cases		Incomplete cases	
	n	%	*n*	%
Trauma	785	65.7	200	69.4
Intimate partner violence (IPV)	437	36.6	94	38.4
Hazardous drinking **	487	40.8	85	30.7
Marijuana use **	1061	88.8	237	82.9
Opiates use[3]	275	23.0	77	27.4
Drug use excluding marijuana & opiates ***	727	60.8	143	49.5
Involvement in substance abuse treatment program **	225	18.8	73	25.0
Auxiliary variables				
Emotional support	485	40.6	105	36.3
Currently have a health care provider	829	69.5	189	65.6
Recent ER/Urgent care	374	31.3	79	27.3
Recent mental health outpatient care	338	28.3	78	26.9

* *p-value* < 0.1

** *p-value* < .05

*** *p-value* < .001.