INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research

International Journal of Quantitative Research in Education

ISSN online: 2049-5994 - ISSN print: 2049-5986
https://www.inderscience.com/ijqre

On the use of inclusive strategy when some participants fail to provide data on all studied variables

Yan Xia, Yachen Luo, Mingya Huang, Yanyun Yang

DOI: 10.1504/IJQRE.2024.10055004

Article History:

Received:	17 March 2021
Last revised:	07 June 2022
Accepted:	12 July 2022
Published online:	28 March 2023

On the use of inclusive strategy when some participants fail to provide data on all studied variables

Yan Xia*

Department of Educational Psychology,
University of Illinois at Urbana-Champaign, USA
Email: yx18@illinois.edu
*Corresponding author

Yachen Luo

Department of Educational Psychology and Learning Systems,
Florida State University, USA
Email: yluo3@fsu.edu

Mingya Huang

Department of Educational Psychology,
University of Wisconsin at Madison, USA
Email: mhuang233@wisc.edu

Yanyun Yang

Department of Educational Psychology and Learning Systems,
Florida State University, USA
Email: yyang3@fsu.edu

Abstract: Behavioural research scientists have become increasingly aware of the importance of missing data methods. Including auxiliary variables in data analysis can increase the plausibility of meeting the missing at random assumption, leading to increased parameter estimation accuracy and a more trustworthy goodness-of-fit evaluation. This study addresses a missing data pattern typically mishandled by using listwise deletion. The missing data pattern echoes a common research scenario in which some participants fail to respond to all the studied variables but provide information on auxiliary variables. Researchers commonly delete these participants from further data analyses in practice. Using confirmatory factor analysis models, this study shows that including effective auxiliary variables to analyse data with this missing data pattern can substantially improve the estimation accuracy, particularly when auxiliary variables correlate with latent factors.

Keywords: structural equation modelling; SEM; missing at random; full information maximum likelihood; FIML; auxiliary variables.

Reference to this paper should be made as follows: Xia, Y., Luo, Y., Huang, M. and Yang, Y. (2022) 'On the use of inclusive strategy when some participants fail to provide data on all studied variables', *Int. J. Quantitative Research in Education*, Vol. 5, No. 4, pp.356–378.

Biographical notes: Yan Xia received his PhD in Measurement and Statistics from Florida State University in 2016. He is currently an Assistant Professor in the Department of Educational Psychology at the University of Illinois at Urbana-Champaign. His research interests include structural equation modelling, missing data analysis, and categorical data analysis.

Yachen Luo is a PhD candidate at Florida State University. She also dual-enrolled in the Applied Statistic in the Department of Statistics at FSU and obtained her Master of Science degree in 2021. Her research focuses on structural equation modelling, categorical data analysis, and large-scale assessments.

Mingya Huang is currently a PhD student in Quantitative Methods, with an MS in Statistics and a minor in Computer Science at the University of Wisconsin-Madison. Her research focuses on Bayesian statistics, machine learning, and causal inference.

Yanyun Yang is a Professor of Measurement and Statistics in the Department of Educational Psychology and Learning Systems at Florida State University. Her research focuses on structural equation modelling, reliability estimation methods, and applications of advanced statistical procedures to subject areas such as mental health and educational psychology.

1 Introduction

Due to the vast development of software implementation of full information maximum likelihood (FIML) and multiple imputation (MI), behavioural research scientists have become increasingly aware of these two state-of-the-art methods for analysing missing data (Schafer and Graham, 2002). Compared with traditional methods such as listwise deletion, pairwise deletion, and single imputation, FIML and MI can provide unbiased parameter estimates, accurate standard errors, and more trustworthy model-data fit evaluation under certain conditions (Schafer, 1997). FIML is primarily implemented in structural equation modelling (SEM) software programs, such as M*plus* (Muthén and Muthén, 1998–2017) and the *lavaan* package in R (Rosseel, 2012). Because many traditional statistical procedures such as regression analysis, ANOVA, and the independent t-test can be viewed as special cases of SEM (e.g., Bollen, 1989), researchers can also use FIML for these traditional statistical analyses. MI, when conducted appropriately, produces similar results to FIML under many conditions (e.g., Collins et al., 2001; Schafer, 2003) and, therefore, FIML is the focus of this study.

Little and Rubin (2002) defined missing at random (MAR) and missing completely at random (MCAR). MAR states that the probability of missingness is independent of the missing values, while MCAR states that whether a response is missing is entirely independent of both the missing and observed values. MAR is a less restrictive assumption than MCAR and is a necessary condition for producing unbiased parameter estimates for FIML (Rubin, 1976). However, when the MAR assumption is violated, that

is, when missingness is due to missing values (MNAR), FIML produces biased results (Little and Rubin, 2002). Unfortunately, the MAR assumption is not testable despite its importance in securing the accuracy of FIML.

To improve the practice of FIML, increasing the plausibility of meeting the MAR assumption is critical. Methodologies have proposed experimental design methods and statistical methods to increase the likelihood of meeting the MAR assumption. Research scientists can employ planned missing data designs before data are collected (e.g., Graham et al., 2006). Such designs randomly assign groups of participants to skip a subset of variables and thus reduce participants' burden. Because of the random selection in such designs, the missing data mechanism is MCAR. After data collection is completed, research scientists can employ inclusive analytical strategies to increase the chance of meeting the MAR assumption when missing data occur, thus reducing bias and increasing efficiency (Collins et al., 2001; Little and Rubin, 2002; Raykov and Marcoulides, 2014; Schafer and Graham, 2002). In the context of MI, Collins et al. (2001) also suggested the inclusion of extra variables that have high correlations with the studied variables that contain missing values because such variables can potentially increase the estimation accuracy. Collins et al. (2001) referred to the extra variables as auxiliary variables (AVs; see also Graham, 2003).

For FIML, Graham (2003) proposed the extra-dependent variable model and saturated model to include AVs. Specifically, the extra-dependent model (Graham, 2003; Graham et al., 1997), available in *lavaan*, requires that all existing predictors predict AVs in a regression model, residuals of AVs are correlated with each other and that the residual of every AV is correlated with the residual of every latent endogenous variable. The saturated model, implemented in *lavaan* and M*plus*, requires that every AV correlate with every exogenous manifest variable in the model of interest and the residual of every endogenous manifest variable. Additionally, all AVs are correlated with each other. Adding AVs contributes to zero addition to the model degrees of freedom with the saturated model, producing similar results to the extra-dependent model. Through a simulation study, Graham (2003) showed that including AVs does not compromise the model of substantive interest. Including the AVs that account for the non-responses meets the MAR assumption and thus improves the estimation accuracy. Additionally, including the AVs which do not account for non-responses does not degrade the parameter estimation accuracy, regardless of the distribution of the AVs.

The advantages in parameter estimation accuracy introduced by the inclusion of AVs in Graham (2003) have been echoed by other studies (e.g., Enders, 2008; Raykov and Marcoulides, 2014; Savalei and Bentler, 2009; Yuan et al., 2015). Effective AVs that can increase the estimation accuracy are either the direct causes of missingness, correlates of missingness, or correlates of the studied variables with missing values (Collins et al., 2001).

2 A missing data pattern commonly being mishandled

Despite the general recognition of FIML as an appropriate method to address non-responses, the following missing data pattern, shown in Figure 1, has not been handled appropriately in practice, based on our observation. Suppose a research scientist developed a new assessment with ten items (i.e., X_1–X_{10} shown in Figure 1) that aims to measure the perception of online violence. For some pilot data evaluating the

psychometric properties, the ten items were included as the last few items in a lengthy online survey relevant to the experience of internet use. Therefore, the 50 items preceding X_1–X_{10} could be considered AVs (AV_1–AV_{50}) when evaluating the measurement model for X_1–X_{10}. Suppose 1,000 participants answered the online survey, but only 600 completed the last ten items, while the other 400 quit the survey before they reached the last ten items. When evaluating the psychometric properties of the last ten items (e.g., through factor analysis), it is a common practice to only use the data from the 600 participants because the other 400 individuals provided no responses to any of the ten items involved in the factor analysis.

Figure 1 An illustration of the missing data scenario where a proportion of participants fail to answer any studied variable but provide responses to auxiliary variables

ID	AV_1	AV_2	…	AV_{50}	X_1	X_2	…	X_{10}
1	Obs.	Obs.	Obs.	Obs.	Obs.	Obs.	Obs.	Obs.
2	Obs.	Obs.	Obs.	Obs.	Obs.	Obs.	Obs.	Obs.
…	Obs.	Obs.	Obs.	Obs.	Obs.	Obs.	Obs.	Obs.
600	Obs.	Obs.	Obs.	Obs.	Obs.	Obs.	Obs.	Obs.
601	Obs.	Obs.	Obs.	Obs.	**Mis.**	**Mis.**	**Mis.**	**Mis.**
…	Obs.	Obs.	Obs.	Obs.	**Mis.**	**Mis.**	**Mis.**	**Mis.**
1,000	Obs.	Obs.	Obs.	Obs.	**Mis.**	**Mis.**	**Mis.**	**Mis.**
					Studied variables			

Notes: 'Obs.' denotes observed values. 'Mis.' denotes missing values.

Intuitively, excluding the 400 participants from the factor analysis seems justifiable because they answered none of the ten items for assessing the perception of online violence. Additionally, the remaining sample size of 600 appeared to be large enough for factor analysis with ten items. However, excluding the 400 participants is essentially the application of listwise deletion, which will result in biased estimates if the missingness in X_1–X_{10} is relevant to AV_1–AV_{50}. For example, if participants who did not enjoy spending time online tended to quit the survey early, the 600 participants then formed a biased sample because the sample contained the participants associated with lower levels of internet involvement, which could be correlated with perceived online violence.

The above missing data patterns may also be observed in longitudinal datasets. For example, the National Education Longitudinal Study of 1988 datasets (Curtin et al., 2002) contained students who dropped out of the school, transferred to other schools, or other reasons for quitting the study in certain waves. Excluding these students from the analyses results in a biased sample (i.e., a sample of students with relatively higher

academic performance than students who dropped out). On the other hand, their data collected in the previous waves could serve as AVs, potentially increasing the accuracy of statistical results.

While the above scenario fits the realm of missing data analysis, excluding the participants who did not respond to any of the studied variables appears to be a more popular choice, according to our experience. While existing missing data literature (e.g., Collins et al., 2001; Graham, 2003) recommends including AVs in the analysis, research scientists may be reluctant to do so for the missing data pattern shown in Figure 1. Potential reasons for the reluctance can be several. First, why would one run a factor analysis with participants who provide no data on any of the studied items in the factor analysis? Including these participants is rather counter-intuitive. Second, it is not unusual that the proportion of participants who skip all the studied variables is large (e.g., greater than 50%). It seems inappropriate to include this portion of participants because the effective sample size appears to be artificially increased. Third, the remaining sample size (e.g., 600) appears acceptable for the planned analysis (e.g., the factor analysis with ten items). Additionally, no studies have investigated the benefits of including the participants who answer none of the studied items in statistical analyses. It is not clear whether including the participants that fail to respond to any studied item can substantially increase the parameter estimation.

The above reasons motivate practitioners to use listwise deletion instead of using an inclusive strategy in conjunction with FIML to handle missing data of this pattern. However, most simulation studies for missing data analyses simulate data that contain participants who skipped some but not all of the studied items (e.g., Collins et al., 2001; Enders, 2008; Graham, 2003; Savalei & Bentler, 2009). We are not aware of any methodological study that directly addressed the missing data pattern of this type, making it difficult to provide suggestions to research scientists in applied fields to handle this type of missing data pattern appropriately.

The purpose of the present simulation study is to evaluate parameter estimation accuracy when the inclusive strategy is applied to scenarios in which a portion of participants fails to answer any studied item but provides data for AVs. We expected that including the AVs that correlate with the studied variables can increase the parameter estimation accuracy compared with listwise deletion.

3 Method

3.1 The data-generation model

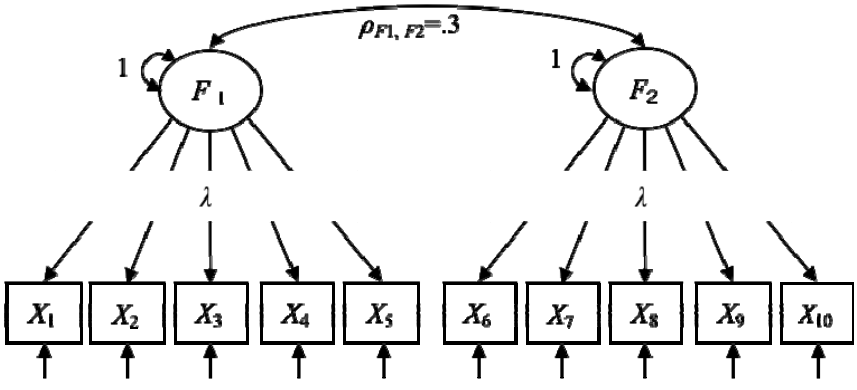
The data-generation model was a confirmatory factor analysis (CFA) model. A CFA model was chosen because it is one of the popular statistical techniques used in many disciplines. The model is shown in Figure 2, which consists of two factors measured by 10 items. The first factor (F_1) was measured by X_1–X_5 and the second factor (F_2) by X_6–X_{10}. The model was expressed as

$$x_i = \tau_i + \lambda_{ik} F_k + e_i ,$$

where x_i is the score for the i^{th} item (i = 1, …, 10), τ_i is the intercept for the i^{th} item, λ_{ik} is the loading of the i^{th} item on the k^{th} latent factor (k = 1, 2), F_k is the score for the k^{th} latent factor, and e_i is the residual score for the i^{th} item. The mean and variance of each factor

were 0 and 1, respectively. The correlation between F_1 and F_2 was fixed at 0.30, which corresponded to distinguishable factors that were correlated in practice. The loading parameter λ had the same value for all items in a given condition. τ_i was 0 for every item. The variance of e_i was $1 - \lambda^2$, such that x_i had a mean of 0 and a variance of 1.

Figure 2 The CFA model for data generation



In addition to X_1–X_{10}, AVs were generated that fell into three categories, according to Collins et al. (2001). First, A was the AV that correlated with X_1–X_{10} and directly caused missingness. Incorporating A into the CFA model as an AV met the MAR assumption, thus leading to unbiased estimates. Second, B_1–B_{10} were the AVs that correlated with X_1–X_{10}, but were not the direct causes of missingness. Although B_1–B_{10} did not directly cause missingness, including them might still improve the estimation accuracy (e.g., Collins et al., 2001; Graham, 2003). Third, C_1–C_{10} were the AVs that were independent of X_1–X_{10} and did not cause missingness. Therefore, C_1–C_{10} were considered ineffective AVs.

3.2 Manipulated conditions

3.2.1 Correlation with AVs

The simulation design systematically varied the correlation between the AVs and latent factors (ρ_{AF}) and the correlation between the AVs and observed items (ρ_{AX}). The magnitude of correlations is important because it is directly related to the effectiveness of AVs (e.g., Collins et al., 2001; Graham, 2003). In the current design, ρ_{AX} is a function of ρ_{AF} or vice versa. Specifically, ρ_{AX} between x_i and AV_j is

$$\begin{aligned} \rho_{AX} &= cor\left(x_i, AV_j\right) \\ &= cor\left(\lambda_{ik} F_k + e_i, AV_j\right) \\ &= \lambda_{ik} cor\left(F_k, AV_j\right), \end{aligned} \tag{1}$$

assuming $cor(e_i, AV_j)$. Therefore, $\rho_{AX} = \lambda_{ik}\, \rho_{AF}$.

While both ρ_{AX} and ρ_{AF} can potentially determine the effectiveness of AVs, previous simulation studies typically varied ρ_{AX} and ρ_{AF} together while fixing λ_{ik} at given values. For example, Enders (2008) fixed λ_{ik} at 0.7 for all conditions and varied ρ_{AF} at 0.24, 0.40,

0.54, and 0.90, letting ρ_{AX} to be determined by equation (1). Such a design avoided the confounding effect due to λ_{ik} but made ρ_{AX} and ρ_{AF} confound. It remains unclear whether ρ_{AX} or ρ_{AF} is more important in determining the effectiveness of the AVs.

Understanding the role of ρ_{AX} and ρ_{AF} is important in practice because practitioners rely on the value of ρ_{AX} and ρ_{AF} to select effective AVs. While ρ_{AX} can be obtained by requesting a FIML-correlation matrix for the AVs and studied variables, ρ_{AF} is not directly observable and requires researchers' theoretical understanding of the variables. In order to distinguish the effects of ρ_{AX} and ρ_{AF}, the present study systematically manipulated the values of ρ_{AX} and ρ_{AF}, letting λ_{ik} be determined by equation (1).

To generate A and B_1–B_{10}, we manipulated ρ_{AX} at 0.2, 0.3, and 0.4 to represent realistic correlations for data in behavioural research. For each value of ρ_{AX}, ρ_{AF} varied at 0.3, 0.5, and 0.7. The value of λ was then determined based on equation (1). For example, when ρ_{AX} = 0.2, λ was 0.67, 0.40, and 0.29 corresponding to ρ_{AF} of 0.3, 0.5, and 0.7, respectively. Note that the two conditions with ρ_{AX} = 0.3 and 0.4, paired with ρ_{AF} = 0.3 led to non-positive definite covariance matrices for X_1–X_{10}, A, and B_1–B_{10} and were thus excluded. The resulting seven values for λ were presented in Table 2. Additionally, the correlations among A and B_1–B_{10} were fixed at .6 to ensure that all the data-generation matrices were positive-definite. C_1–C_{10} were independent and identically distributed random variables, and each was generated according to a standard normal distribution. Therefore, C_1–C_{10} were considered ineffective AVs.

3.2.2 The proportion of missing values

Missingness on X_1–X_{10} was determined by the value of A. The proportion of missing values (P%) varied at 10%, 25%, and 50%. For conditions with P% missing participants, participants whose scores of A smaller than the P^{th} percentile point of A had missing values on all X_1–X_{10}.

3.2.3 Sample size

The sample size was 100, 300, and 600. Therefore, the number of participants that had data on X_1–X_{10} was smaller due to missing values (e.g., 50 participants if 50% missingness existed for the sample size of 100). Parameter estimates from a large sample size of 100,000 were reported to approximate the population values. Therefore, any differences in the results between the small samples and the large sample reflect the impact of sampling fluctuation. The sample sizes of 300 and 600 were considered acceptable for a simple-structured CFA model with ten items. The available data were at least 150 with 50% of missingness. Although 150 may not be sufficient for a CFA with ten items, it is not uncommon in practice. The sample size of 100 could be too small, particularly with a considerable amount of missingness, but it provided insights regarding how small the inclusive strategy might lead to severe inadmissible solution issues and highly biased results.

A total of 63 simulation conditions were obtained (i.e., seven combinations of ρ_{AF} and ρ_{AX} × three proportions of missingness × three sample sizes). For each simulation condition, 1,000 replications were implemented.

3.3 Data analysis

For each generated dataset, a two-factor model was analysed with or without including AVs. Complete data before the generation of missing values were analysed using normal-theory maximum likelihood without including the AVs. Parameter estimates from the complete datasets were used as the baseline for comparison. The parameter of interest was the correlation between the two factors. For each dataset with missing values, the following five analyses were conducted. For the analyses with AVs, Graham's (2003) saturated model was used such that every AV was correlated with the residuals of X_1–X_{10} and that all AVs were correlated with each other.

1 No AVs were included in the analyses. This analysis was essentially the same as applying listwise deletion.

2 A was included. Because A was the correlate of X_1–X_{10} and the direct cause of missingness, the parameter estimation accuracy was expected to increase compared with analysis (1).

3 B_1–B_{10} were included. Including many correlates could potentially increase the parameter estimation accuracy.

4 C_1–C_{10} were included. Including ten ineffective AVs would not improve or degrade the estimation accuracy. Inadmissible solution rates, however, might increase when C_1–C_{10} were included, according to previous studies (e.g., Savalei and Bentler, 2009).

5 A_1, B_1–B_{10}, and C_1–C_{10} were included. Including all AVs mimicked the scenario where a researcher adds all the available variables in the analysis without selecting the AVs carefully. Including too many variables that are either effective or ineffective, the inadmissible solution rates may increase. However, whenever the algorithm successfully converged, we expected that the parameter estimation was more accurate than listwise deletion.

Rates of inadmissible solutions and average parameter estimates across the replications were summarised. A replication had inadmissible solutions if the model did not converge or converged to improper solutions. Improper solutions included negative variances of item residuals, factor correlations greater than 1 in absolute values, and standard errors not applicable. Replications with inadmissible solutions were excluded from further results summary. All data generation and analyses were implemented in R (R Core Team, 2019). The *lavaan* package was used for the CFA (Rosseel, 2012), and AVs were incorporated into the analyses by Graham's (2003) saturated model using the *semTools* package (Jorgensen et al., 2018).

4 Results

In this section, parameter estimates based on a large sample (N = 10,000) were first reported. These estimates were used to approximate population values under various conditions of missingness and the correlation with AVs. Then results from small samples (n = 100, 300, and 600) were reported. For small samples, inadmissible solution rates were summarised, followed by parameter estimation.

Table 1 Parameter estimates of latent factor correlation when sample size was 100,000

ρ_{AX}	ρ_{AF}	λ	*Com*	*LW*	*A*	*B*	*C*	*ALL*
				10% missingness				
0.2	0.3	0.67	0.298	0.281	0.300	0.297	0.281	0.300
0.2	0.5	0.40	0.296	*0.244*	0.297	0.290	*0.244*	0.297
0.2	0.7	0.29	0.301	*0.180*	0.293	0.281	*0.180*	0.292
0.3	0.5	0.60	0.300	*0.248*	0.302	0.294	*0.248*	0.302
0.3	0.7	0.43	0.300	*0.184*	0.302	0.287	*0.184*	0.302
0.4	0.5	0.80	0.305	*0.248*	0.303	0.295	*0.248*	0.303
0.4	0.7	0.57	0.296	*0.185*	0.297	0.281	*0.185*	0.296
					25% missingness			
0.2	0.3	0.67	-	*0.268*	0.299	0.295	*0.268*	0.299
0.2	0.5	0.40	-	*0.196*	0.285	0.273	*0.196*	0.285
0.2	0.7	0.29	-	*0.091*	0.301	0.271	*0.091*	0.299
0.3	0.5	0.60	-	*0.201*	0.293	0.279	*0.201*	0.293
0.3	0.7	0.43	-	*0.097*	0.304	0.277	*0.097*	0.305
0.4	0.5	0.80	-	*0.204*	0.298	0.284	*0.204*	0.297
0.4	0.7	0.57	-	*0.084*	0.289	0.263	*0.084*	0.289
					50% missingness			
0.2	0.3	0.67	-	*0.262*	0.306	0.296	*0.262*	0.306
0.2	0.5	0.40	-	*0.151*	0.290	0.263	*0.151*	0.288
0.2	0.7	0.29	-	*0.012*	0.315	0.278	*0.012*	0.318
0.3	0.5	0.60	-	*0.164*	0.299	0.276	*0.164*	0.299
0.3	0.7	0.43	-	*–0.003*	0.305	0.264	*–0.003*	0.304
0.4	0.5	0.80	-	*0.170*	0.304	0.283	*0.170*	0.304
0.4	0.7	0.57	-	*–0.003*	0.312	0.268	*–0.003*	0.311

Note: Values that were greater than 10% relative bias (i.e., > 0.33 or < 0.27) were italics; Com is for complete data; LW is listwise deletion; *A* was the direct cause of missingness and correlated with the studied variables; *B* represented the analytical conditions where B_1–B_{10} (i.e., correlates that are not the direct cause of missingness) were included; *C* represented the analytical conditions where C_1–C_{10} (independent random noises) were included.

4.1 Parameter estimation accuracy in large samples

Table 1 presents the parameter estimates for the factor correlation when the sample size was 100,000. When data were complete, the estimates were unbiased across all conditions. With incomplete data, listwise deletion led to substantial bias (> 10% relative bias) across all the conditions except for the condition with $\rho_{AX} = 0.2$, $\rho_{AF} = 0.3$, and 10% missingness, where the relative bias was 6% (i.e., 0.281 compared with the population value of 0.300). The bias resulting from listwise deletion increased when the proportion of missingness increased.

Given a value of ρ_{AX}, the bias due to listwise deletion increased when ρ_{AF} increased. For example, when ρ_{AX} was 0.2 under the conditions with 10% missingness, the estimates were 0.281, 0.244, and 0.180 when ρ_{AF} was 0.3, 0.5, and 0.7, respectively. Given a value of ρ_{AF}, the magnitude of ρ_{AX} and thus λ only slightly impacted the estimates for listwise deletion. For example, when ρ_{AF} was 0.7 under the conditions with 10% missingness, the estimates were 0.180, 0.184, and 0.185 when ρ_{AX} was 0.2, 0.3, and 0.4.

When missing values existed, including A as the AV resulted in accurate parameter estimates across all conditions. The highest relative bias was 5% among all conditions. Including B_1–B_{10} as the AVs also improved the estimation accuracy compared with the listwise deletion. However, the results obtained by including B_1–B_{10} could be less accurate than those obtained by having A, especially when more missing values existed. For example, four conditions (one condition with 25% missingness and three conditions with 50% missingness) led to relative bias greater than 10% in Table 1 when B_1–B_{10} were treated as the AVs. C_1–C_{10} were ineffective AVs because including them did not improve nor degrade the estimation accuracy compared with listwise deletion. With all the AVs included, results were almost the same as those obtained from analyses with A only.

4.2 Inadmissible solution rates in small samples

Table 2 presents the inadmissible solution rates for each simulation condition. The conditions with complete data provided the baseline inadmissible solution rates for comparison. When data were complete, ρ_{AX}, ρ_{AF}, and the proportion of missingness had no impact on the inadmissible solution rates. Therefore, only n and λ could change the inadmissible solution rates for complete data in the simulation design. For complete data, the highest inadmissible solution rates occurred under the conditions with λ being 0.29 and n being 100. When n increased to 600, all conditions had close to 0 inadmissible solution rates. When missing values existed, the inadmissible solution rates increased with higher proportions of missingness, regardless of how the missing data were handled. For listwise deletion, the inadmissible solution rates increased in general compared with the rates from complete data, especially for conditions with a small ρ_{AX} in conjunction with a large ρ_{AF}, which corresponded to a small λ. For example, with 25% missingness, $\rho_{AX} = 0.2$, $\rho_{AF} = 0.7$ and $n = 100$, the complete-data condition yielded 61.7% inadmissible solution rates and listwise deletion led to a higher rate of 79.2%.

Compared to the listwise deletion, including A as the only AV or including B_1–B_{10} as AVs resulted in lower, or almost the same (i.e., difference smaller than .1%) inadmissible solution rates. Including C_1–C_{10} as AVs and the listwise deletion yielded similar inadmissible solution rates (i.e., the difference was mostly within 1%, and the largest difference was 2.1% among all conditions) because C_1–C_{10} were simply random noises. Nevertheless, compared with the listwise deletion, when A, B_1–B_{10}, and C_1–C_{10} were all included as the AVs, the inadmissible solution rates decreased by at least 1% in 12 out of the 63 conditions, increased by at least 1% in 4 out of the 63 conditions (all the four conditions had a sample size of 100 and 50% missingness), and remained to be similar (difference smaller than 1%) for the other 47 conditions. The results implied that adding a certain number of extra random noises would not increase the inadmissible solution rates as long as the effective AVs were included unless the effective sample size was extremely small.

Table 2 Inadmissible solution rates (%) for the CFA model

ρ_{AX}	ρ_{AF}	λ	*n* = 100						*n* = 300						*n* = 600					
			Com	*LW*	*A*	*B*	*C*	*ALL*	*Com*	*LW*	*A*	*B*	*C*	*ALL*	*Com*	*LW*	*A*	*B*	*C*	*ALL*
10% missingness																				
0.2	0.3	0.67	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.2	0.5	0.40	9.0	19.3	16.0	17.2	19.0	17.6	0	0	0	0	0	0	0	0	0	0	0	0
0.2	0.7	0.29	60.5	69.5	67.1	69.5	70.7	69.5	13.0	30.2	22.7	26.1	31.0	25.7	0.2	2.9	2.3	3.0	3.0	3.3
0.3	0.5	0.60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.3	0.7	0.43	5.0	14.9	11.4	12.7	15.2	13.1	0	0	0	0	0	0	0	0	0	0	0	0
0.4	0.5	0.80	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.4	0.7	0.57	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25% missingness																				
0.2	0.3	0.67	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.2	0.5	0.40	10.7	37.4	31.7	34.1	37.2	36.6	0	0.5	0.3	0.3	0.6	0.4	0	0	0	0	0	0
0.2	0.7	0.29	61.7	79.2	74.2	76.7	78.8	77.8	11.7	46.9	34.9	39.5	47.1	39.0	0.5	15.4	9.7	11.5	15.1	10.9
0.3	0.5	0.60	0	0	0	.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.3	0.7	0.43	5.2	31.1	24.2	25.7	30.6	27.8	0	0.9	0.4	0.4	0.7	0.4	0	0	0	0	0	0
0.4	0.5	0.80	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.4	0.7	0.57	0	1.2	0.4	0.5	1.3	0.5	0	0	0	0	0	0	0	0	0	0	0	0
50% missingness																				
0.2	0.3	0.67	0	0.4	0.4	0.4	0.4	0.8	0	0	0	0	0	0	0	0	0	0	0	0
0.2	0.5	0.40	12.7	57.0	55.0	54.9	57.5	62.3	0	7.3	5.2	4.2	7.1	5.3	0	0.2	0.2	0.3	0.2	0.2
0.2	0.7	0.29	59.0	86.5	84.5	85.2	87.3	85.2	12.1	66.3	58.9	58.5	68.4	59.2	1.1	45.5	31.2	32.2	45.7	33.1
0.3	0.5	0.60	0	3.6	3.5	3.2	4.5	7.8	0	0	0	0	0	0	0	0	0	0	0	0
0.3	0.7	0.43	5.1	59.0	51.9	52.9	60.5	61.8	0	9.8	5.0	4.8	9.8	5.6	0	0.4	0.2	0.2	0.4	0.2
0.4	0.5	0.80	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.4	0.7	0.57	0	13.8	10.5	8.7	14.7	16.7	0	0.1	0	0	0	0	0	0	0	0	0	0

Note: Inadmissible solution rates included the replications with non-convergence, negative variances of residuals, factor correlations in absolute values greater than 1, and standard errors that were not applicable; Com is for complete data; LW is listwise deletion; *A* was the direct cause of missingness and correlated with the studied variables; *B* represented the analytical conditions where B_1–B_{10} (i.e., correlates that are not the direct cause of missingness) were included; *C* represented the analytical conditions where C_1–C_{10} (independent random noises) were included.

4.3 Parameter estimation accuracy in small samples

Tables 3, 4, and 5 present the means and empirical standard errors of factor correlation estimates across replications yielding admissible solutions for each condition with a sample size of 600, 300, and 100, respectively. Because analyses were conducted based on the small samples, the differences between Tables 1 and Tables 3–5 were consequences of sampling fluctuation. Under small samples, the means of factor correlation estimates showed similar patterns to those in Table 1.

First, small samples resulted in unbiased factor correlation estimates for complete data. The only exception was for the condition with $\rho_{AX} = 0.2$, $\rho_{AF} = 0.7$, $\lambda = 0.29$, and $n = 100$, where the estimates were biased due to the small sample size in conjunction with a small λ. Biased estimates existed for all conditions when missing data were handled by listwise deletion. Similar to the pattern in Table 1 when data were complete, given a value of ρ_{AX}, the bias due to listwise deletion increased when ρ_{AF} increased. Given a value of ρ_{AF}, however, the magnitude of ρ_{AX} and thus λ only slightly impacted the estimates for listwise deletion.

With A included as the only AV, the estimation accuracy improved substantially compared with listwise deletion. With A included, all conditions with 600 observations resulted in the relative bias smaller than 10%. When n was 300, the only condition that showed the relative bias greater than 10% had a λ of 0.29 and 50% missingness. When n was 100, more conditions had greater than 10% relative bias, especially for conditions with 50% missingness and conditions with a λ of 0.29. However, the bias obtained by including A was consistently smaller than the bias from listwise deletion. For example, with 50% missingness and 100 observations, listwise deletion resulted in average estimates ranging from –0.004 to 0.248, but including A resulted in average estimates from 0.133 to 0.272.

Including B_1–B_{10} under small samples also improved the estimation accuracy compared with listwise deletion, although the improvement was not as large as that from the analysis with A included.[1] With B_1–B_{10}, the bias increased when λ was smaller, n was smaller, and the proportion of missingness was larger. Nevertheless, the bias obtained by including B_1–B_{10} was consistently smaller than the bias from listwise deletion.

With C_1–C_{10} included, results remained almost the same as the results obtained from listwise deletion. When all A, B_1–B_{10}, and C_1–C_{10} were included as AVs, parameter estimation accuracy substantially improved compared with listwise deletion. Including all AVs provided almost identical results compared with the analyses with only A.

Note that when λ was too small, including A, B_1–B_{10}, or all AVs could still result in biased estimates, but they still resulted in improved estimation accuracy compared with the listwise deletion. For example, with 50% missingness and a λ of 0.29, including A led to mean estimates of 0.133, 0.209, and 0.275 for n of 100, 300, and 600, respectively, while listwise deletion only led to 0.040, –0.053, and –0.006.

As for the empirical standard errors, the inclusion of different types of AVs did not change the results substantially for most conditions. Notable differences (e.g., > 0.01 difference) only occurred under conditions where large standard errors were expected (i.e., conditions with very small sample size, low loadings, or a large proportion of missingness). Under such conditions, including A or all the AVs resulted in the smallest empirical standard errors, followed by including B_1–B_{10} only. Applying the listwise deletion or including C_1–C_{10} as the only AVs led to the largest empirical standard errors.

Table 3 Means and empirical standard errors of factor correlation estimates when sample size was 600

ρ_{AX}	ρ_{AF}	λ	*Com*		*LW*		*A*		*B*		*C*		*ALL*	
			Est	*SE*	*Est*	*SE*	*Est*	*SE*	*Est*	*SE*	*Est*	*SE*	*Est*	*SE*
10% missingness														
0.2	0.3	0.67	0.300	0.049	0.281	0.052	0.299	0.052	0.296	0.052	0.281	0.052	0.299	0.052
0.2	0.5	0.40	0.297	0.087	*0.244*	0.095	0.297	0.092	0.290	0.092	*0.244*	0.095	0.296	0.092
0.2	0.7	0.29	0.294	0.142	*0.179*	0.175	0.291	0.157	0.276	0.158	*0.179*	0.175	0.290	0.155
0.3	0.5	0.60	0.300	0.052	*0.246*	0.057	0.300	0.057	0.293	0.056	*0.246*	0.057	0.300	0.056
0.3	0.7	0.43	0.296	0.073	*0.182*	0.087	0.296	0.081	0.283	0.079	*0.182*	0.087	0.297	0.079
0.4	0.5	0.80	0.300	0.041	*0.246*	0.044	0.300	0.044	0.293	0.044	*0.246*	0.044	0.300	0.044
0.4	0.7	0.57	0.299	0.055	*0.182*	0.063	0.297	0.059	0.283	0.059	*0.182*	0.063	0.297	0.058
25% missingness														
0.2	0.3	0.67	-	-	*0.271*	0.056	0.301	0.056	0.297	0.056	*0.271*	0.056	0.301	0.057
0.2	0.5	0.40	-	-	*0.208*	0.105	0.300	0.101	0.286	0.100	*0.209*	0.105	0.299	0.100
0.2	0.7	0.29	-	-	*0.105*	0.214	0.300	0.180	0.273	0.180	*0.105*	0.214	0.300	0.180
0.3	0.5	0.60	-	-	*0.206*	0.065	0.297	0.064	0.283	0.063	*0.206*	0.065	0.296	0.063
0.3	0.7	0.43	-	-	*0.099*	0.109	0.302	0.096	0.279	0.095	*0.099*	0.109	0.304	0.093
0.4	0.5	0.80	-	-	*0.207*	0.050	0.298	0.051	0.285	0.049	*0.207*	0.050	0.297	0.05
0.4	0.7	0.57	-	-	*0.093*	0.073	0.298	0.069	0.273	0.064	*0.093*	0.073	0.299	0.063
50% missingness														
0.2	0.3	0.67	-	-	*0.259*	0.070	0.299	0.072	0.293	0.070	*0.259*	0.070	0.298	0.072
0.2	0.5	0.40	-	-	*0.167*	0.142	0.291	0.137	0.277	0.132	*0.167*	0.142	0.292	0.136
0.2	0.7	0.29	-	-	*–0.006*	0.296	0.275	0.252	0.246	0.238	*–0.011*	0.296	0.279	0.246
0.3	0.5	0.60	-	-	*0.168*	0.084	0.295	0.087	0.277	0.081	*0.168*	0.084	0.294	0.084
0.3	0.7	0.43	-	-	*–0.011*	0.146	0.299	0.127	0.261	0.117	*–0.011*	0.146	0.300	0.118
0.4	0.5	0.80	-	-	*0.170*	0.063	0.297	0.068	0.282	0.061	*0.170*	0.063	0.299	0.065
0.4	0.7	0.57	-	-	*–0.018*	0.094	0.295	0.090	0.255	0.076	*–0.018*	0.095	0.294	0.076

Notes: Values that were greater than 10% relative bias (i.e., > 0.33 or < 0.27) were iatlics; Com is for complete data; LW is listwise deletion; *A* was the direct cause of missingness and correlated with the studied variables; *B* represented the analytical conditions where B_1–B_{10} (i.e., correlates that are not the direct cause of missingness) were included; *C* represented the analytical conditions where C_1–C_{10} (independent random noises) were included.

Table 4 Means and empirical standard errors of factor correlation estimates when sample size was 300

ρ_{AX}	ρ_{AF}	λ	*Com*		*LW*		*A*		*B*		*C*		*ALL*	
			Est	*SE*	*Est*	*SE*	*Est*	*SE*	*Est*	*SE*	*Est*	*SE*	*Est*	*SE*
								10% missingness						
0.2	0.3	0.67	0.297	0.067	0.279	0.073	0.296	0.072	0.294	0.072	0.279	0.073	0.296	0.073
0.2	0.5	0.40	0.301	0.119	*0.248*	0.133	0.302	0.129	0.295	0.129	*0.248*	0.133	0.301	0.129
0.2	0.7	0.29	0.292	0.225	*0.187*	0.271	0.293	0.247	0.275	0.247	*0.187*	0.271	0.284	0.247
0.3	0.5	0.60	0.299	0.074	*0.246*	0.082	0.299	0.081	0.292	0.080	*0.246*	0.082	0.299	0.081
0.3	0.7	0.43	0.301	0.106	*0.186*	0.127	0.298	0.118	0.285	0.118	*0.186*	0.127	0.299	0.117
0.4	0.5	0.80	0.300	0.059	*0.245*	0.065	0.299	0.065	0.292	0.064	*0.245*	0.065	0.298	0.064
0.4	0.7	0.57	0.302	0.081	*0.188*	0.094	0.301	0.089	0.287	0.087	*0.188*	0.094	0.301	0.086
								25% missingness						
0.2	0.3	0.67	-	-	*0.266*	0.082	0.295	0.083	0.291	0.082	*0.266*	0.082	0.295	0.083
0.2	0.5	0.40	-	-	*0.215*	0.155	0.301	0.151	0.293	0.147	*0.215*	0.156	0.303	0.150
0.2	0.7	0.29	-	-	*0.087*	0.286	0.280	0.251	*0.248*	0.247	*0.086*	0.283	0.276	0.244
0.3	0.5	0.60	-	-	*0.208*	0.094	0.296	0.094	0.285	0.090	*0.208*	0.093	0.296	0.091
0.3	0.7	0.43	-	-	*0.101*	0.157	0.301	0.138	0.278	0.133	*0.101*	0.157	0.303	0.131
0.4	0.5	0.80	-	-	*0.206*	0.071	0.296	0.072	0.284	0.069	*0.206*	0.071	0.296	0.071
0.4	0.7	0.57	-	-	*0.097*	0.104	0.300	0.097	0.276	0.089	*0.097*	0.104	0.301	0.088
								50% missingness						
0.2	0.3	0.67	-	-	*0.254*	0.099	0.291	0.103	0.287	0.099	*0.255*	0.099	0.291	0.105
0.2	0.5	0.40	-	-	*0.174*	0.208	0.288	0.201	0.279	0.195	*0.172*	0.209	0.288	0.205
0.2	0.7	0.29	-	-	*–0.053*	0.386	*0.209*	0.368	*0.188*	0.358	*–0.039*	0.401	*0.213*	0.373
0.3	0.5	0.60	-	-	*0.169*	0.119	0.288	0.127	0.275	0.116	*0.169*	0.119	0.285	0.123
0.3	0.7	0.43	-	-	*–0.020*	0.209	0.279	0.189	*0.251*	0.172	*–0.021*	0.209	0.285	0.178
0.4	0.5	0.80	-	-	*0.174*	0.092	0.296	0.103	0.281	0.091	*0.174*	0.092	0.294	0.098
0.4	0.7	0.57	-	-	*–0.023*	0.136	0.282	0.126	*0.254*	0.106	*–0.023*	0.135	0.287	0.108

Notes: Values that were greater than 10% relative bias (i.e., > 0.33 or < 0.27) were iatlics; Com is for complete data; LW is listwise deletion; *A* was the direct cause of missingness and correlated with the studied variables; *B* represented the analytical conditions where B_1–B_{10} (i.e., correlates that are not the direct cause of missingness) were included; *C* represented the analytical conditions where C_1–C_{10} (independent random noises) were included.

Table 5 Means and empirical standard errors of factor correlation estimates when sample size was 100

ρ_{AX}	ρ_{AF}	λ	*Com*		*LW*		*A*		*B*		*C*		*ALL*	
			Est	*SE*	*Est*	*SE*	*Est*	*SE*	*Est*	*SE*	*Est*	*SE*	*Est*	*SE*
10% missingness														
0.2	0.3	0.67	0.300	0.121	0.283	0.129	0.299	0.129	0.298	0.129	0.283	0.129	0.299	0.130
0.2	0.5	0.40	0.300	0.234	*0.246*	0.263	0.292	0.254	0.284	0.256	*0.245*	0.262	0.292	0.256
0.2	0.7	0.29	*0.236*	0.373	*0.131*	0.413	*0.216*	0.396	*0.171*	0.398	*0.130*	0.409	*0.172*	0.421
0.3	0.5	0.60	0.296	0.131	*0.241*	0.144	0.290	0.144	0.284	0.143	*0.241*	0.145	0.290	0.146
0.3	0.7	0.43	0.290	0.202	*0.178*	0.238	0.285	0.219	0.272	0.216	*0.178*	0.238	0.282	0.216
0.4	0.5	0.80	0.294	0.102	*0.242*	0.109	0.293	0.110	0.285	0.109	*0.242*	0.109	0.290	0.110
0.4	0.7	0.57	0.296	0.132	*0.179*	0.153	0.291	0.146	0.283	0.142	*0.179*	0.153	0.294	0.141
25% missingness														
0.2	0.3	0.67	-	-	0.279	0.142	0.303	0.144	0.301	0.144	0.279	0.143	0.302	0.149
0.2	0.5	0.40	-	-	*0.215*	0.290	0.290	0.282	0.282	0.283	*0.214*	0.294	0.285	0.302
0.2	0.7	0.29	-	-	*0.061*	0.455	*0.175*	0.448	*0.122*	0.470	*0.055*	0.472	*0.138*	0.471
0.3	0.5	0.60	-	-	*0.200*	0.166	0.280	0.168	0.271	0.166	*0.200*	0.168	0.275	0.175
0.3	0.7	0.43	-	-	*0.083*	0.287	0.272	0.267	*0.252*	0.255	*0.086*	0.285	*0.257*	0.267
0.4	0.5	0.80	-	-	*0.205*	0.126	0.285	0.128	0.278	0.124	*0.205*	0.127	0.280	0.129
0.4	0.7	0.57	-	-	*0.103*	0.192	0.295	0.176	0.279	0.163	*0.102*	0.192	0.292	0.167
50% missingness														
0.2	0.3	0.67	-	-	*0.248*	0.170	0.272	0.179	0.273	0.175	*0.249*	0.175	0.277	0.215
0.2	0.5	0.40	-	-	*0.183*	0.365	*0.238*	0.375	*0.249*	0.361	*0.173*	0.375	0.281	0.430
0.2	0.7	0.29	-	-	*0.040*	0.510	*0.133*	0.526	*0.178*	0.493	*0.045*	0.540	*–0.005*	0.593
0.3	0.5	0.60	-	-	*0.168*	0.223	*0.249*	0.235	*0.257*	0.230	*0.166*	0.228	*0.240*	0.281
0.3	0.7	0.43	-	-	*–0.009*	0.356	*0.205*	0.380	*0.224*	0.345	*0.007*	0.362	*0.177*	0.422
0.4	0.5	0.80	-	-	*0.154*	0.153	*0.241*	0.172	*0.248*	0.160	*0.155*	0.157	*0.229*	0.206
0.4	0.7	0.57	-	-	*–0.004*	0.258	*0.252*	0.266	*0.254*	0.219	*–0.012*	0.266	*0.218*	0.277

Notes: Values that were greater than 10% relative bias (i.e., > 0.33 or < 0.27) were italics; Com is for complete data; LW is listwise deletion; *A* was the direct cause of missingness and correlated with the studied variables; *B* represented the analytical conditions where B_1–B_{10} (i.e., correlates that are not the direct cause of missingness) were included; *C* represented the analytical conditions where C_1–C_{10} (independent random noises) were included.

Although loading parameters were not the focal parameters in the current simulation study, we reported the results for loading estimates in Tables A1–A4 in the appendix. In general, the loading estimates were less biased (i.e., relative bias smaller than 10% in most conditions) compared with the factor correlation estimates when the listwise deletion was employed or when C_1–C_{10} were included as AVs. Such bias could be reduced by including A or B_1–B_{10} as the AVs or by including all the available AVs.

5 Discussion

This study aims to increase the awareness of a particular missing data pattern where some participants skip all studied variables but provide information for AVs. Despite the increasing understanding of missing data analysis, this missing data pattern is typically handled by listwise deletion, which can result in substantially biased parameter estimates if the remaining participants form a non-random sample. The simulation results evidenced the potential gains in parameter estimation accuracy when effective AVs were included in the analyses. We recommend that one should not simply delete the non-responses from further analysis because such practice is equivalent to applying listwise deletion.

A challenge lying in the application of AVs is the selection of effective AVs, particularly when a large number of candidate AVs exist. Results based on linear regression models from Collins et al. (2001) suggested that, with 25% missingness, AVs that correlated with the studied variables smaller than 0.4 could be safely ignored. However, in CFA analyses, a ρ_{AX} of 0.4 can be demanding because it requires both ρ_{AF} and λ to be high. For example, a ρ_{AF} of 0.6 and a λ of 0.6 yield only a ρ_{AX} of 0.36. The present study systematically varied ρ_{AX} and ρ_{AF} and showed that AVs with a small ρ_{AX} could also be effective if the size of ρ_{AF} is relatively large. Intuitively, because latent factor scores can be considered entirely missing, the factor correlation estimate is more accurate when including AVs that correlate highly with the latent factor scores. Given a value of ρ_{AX}, a larger ρ_{AF} is accompanied by a smaller λ, which leads to a higher bias when the listwise deletion is applied. If the sample size is enough, including the AVs can largely improve the estimation accuracy. Even if the sample size is small, including the AVs that have high ρ_{AF} still improve the estimation accuracy greatly, compared with listwise deletion.

Our results showed that including AVs with ρ_{AF} as low as 0.3 can substantially increase the estimation accuracy when at least 25% missingness exists. With 10% missingness, ρ_{AF} needs to be 0.5 in order for the AVs to contribute to a noticeable improvement in estimation accuracy. Although the influence of ρ_{AF} on the effectiveness of AVs is not that important when ρ_{AF} is held constant, ρ_{AX} can be easily obtained by requesting a FIML correlation matrix among the studied variables and AVs. When selecting AVs from many candidates, we suggest that one can follow a two-step procedure. First, examine the values of ρ_{AX} estimated by FIML because a very small ρ_{AX} (i.e., < 0.2) suggests either the loading or ρ_{AF} is extremely small. Therefore, we can quickly remove the AVs that have ρ_{AX} smaller than 0.2. Second, for the remaining candidate AVs, careful consideration of ρ_{AF} is critical. Because the value of ρ_{AF} is not provided by the data, the second step thus requires researchers' substantive understanding of the variables. We recommend selecting AVs that are, in theory, correlated with the investigated latent factors in the model of interest. Note that new methods have been

proposed to improve the practice of selecting AVs (e.g., Raykov and Marcoulides, 2014), which are beyond the scope of the present study.

Another difficulty of the application of inclusive strategy is that including ineffective AVs can potentially increase inadmissible solution rates, particularly when too many AVs exist (Savalei and Bentler, 2009). However, there exists no definite answer regarding the maximum number of auxiliary variables researchers should include. After identifying a set of auxiliary variables that are potentially helpful, researchers can conduct sensitivity analyses to investigate whether a set of auxiliary variables results in non-convergence. If non-convergence occurs, a smaller set of auxiliary variables could then be considered. Previous simulation studies focused on analyses where only ineffective AVs (i.e., random noises) were included (e.g., Collins et al., 2001; Enders, 2008; Graham, 2003) but did not consider the analyses in which both effective and ineffective AVs were included. Our simulation results showed that when both effective and ineffective AVs are included in the saturated model, the parameter estimates are as accurate as those from including the effective AVs only, and the inadmissible solution rates may decrease compared with the analyses with no AVs at all and the analyses with only the ineffective AVs.

The use of AVs has several other limitations that require further investigation. First, Savalei and Bentler (2009) pointed out the potential under-identification issue of the saturated model under conditions with large model sizes and many AVs. Savalei and Bentler (2009) also illustrated an awkward feature of the saturated model. That is, the saturated model can result in a non-positive definite error covariance matrix, leading to results that may not make sense. Second, while Graham (2003) did not find any type of auxiliary variable that could degrade the estimation accuracy, Thoemmes and Rose (2014) brought up a situation where the inclusion of AVs can increase bias in missing data analysis. Thoemmes and Rose (2014) showed that bias could be introduced when an auxiliary variable is a collider variable (Pearl, 2000). A collider AV is the type of AV that introduces a covariance between a dependent variable and the missingness of that dependent variable and thus introduces MNAR when the collider AV is included in the saturated model. However, Thoemmes and Rose (2014) did not propose methods for detecting collider AVs in practice, which requires further investigation.

The current simulation design is limited in several aspects. For example, generated data were continuous and followed multivariate normal distributions. Behavioural research typically involves Likert-type items or non-normally distributed data. More complicated data distributions can result in lower estimation accuracy in general and higher inadmissible solution rates, which requires more replications in the simulation design. Our simulation is also limited in how missing data were generated. Collins et al. (2001) have shown that different missing data generation methods (e.g., whether the probability of missingness and an AV has a linear or nonlinear relationship) can result in different levels of bias. Additionally, although the results of fit indices were not reported, the authors of the current study found that RMSEA, CFI, and TLI were generally not influenced by the choice of AVs mainly because the current study only analysed correctly specified models. It is expected that under misspecified models, the choice of AVs could impact model goodness of fit. Despite the limitations, the overarching goal of the present study is to raise the awareness of the investigated missing data pattern where a portion of participants skips all studied variables but provides data to AVs. We hope that methodologists will incorporate the investigated missing data pattern into their simulation studies in the future because of its prevalence in behavioural research.

References

Bollen, K.A. (1989) *Structural Equations with Latent Variables*, John Wiley & Sons, http://dx.doi.org/10.1002/9781118619179.

Collins, L.M., Schafer, J.L. and Kam, C.M. (2001) 'A comparison of inclusive and restrictive strategies in modern missing data procedures', *Psychological Methods*, Vol. 6, No. 4, pp.330–351, https://doi.org/10.1037/1082-989X.6.4.330.

Curtin, T.R., Ingels, S., Wu, S. and Heuer, R. (2002) *Base-Year to Fourth Follow-up Data File User's Manual*, National Education Longitudinal Study of 1988, NCES 2002-323, National Center for Education Statistics.

Enders, C.K. (2008) 'A note on the use of missing auxiliary variables in full information maximum likelihood-based structural equation models', *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 15, No. 3, pp.434–448, https://doi.org/10.1080/10705510802154307

Graham, J.W. (2003) 'Adding missing-data-relevant variables to FIML-based structural equation models', *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 10, No. 1, pp.80–100, https://doi.org/10.1207/S15328007SEM1001_4.

Graham, J.W., Hofer, S.M., Donaldson, S.I., MacKinnon, D.P. and Schafer, J.L. (1997) 'Analysis with missing data in prevention research', in Bryant, K., Windle, M. and West, S. (Eds.): *The Science of Prevention: Methodological Advances from Alcohol and Substance Abuse Research*, pp.325–366, American Psychological Association, https://doi.org/10.1037/10222-010.

Graham, J.W., Taylor, B.J., Olchowski, A.E. and Cumsille, P.E. (2006) 'Planned missing data designs in psychological research', *Psychological Methods*, Vol. 11, No. 4, pp.323–343, https://doi.org/10.1037/1082-989X.11.4.323.

Jorgensen, T.D., Pornprasertmanit, S., Schoemann, A., Rosseel, Y., Miller, P., Quick, C. et al. (2018) *Package 'semTools'* [online] https://cran.r-project.org/web/packages/semTools/semTools.pdf (accessed 1 December 2020).

Little, R.J.A. and Rubin, D.B. (2002) *Statistical Analysis with Missing Data*, John Wiley & Sons, https://doi.org/10.1002/9781119013563.

Muthén, L.K. and Muthén, B.O. (1998–2017) *Mplus User's Guide*, 8th ed., Muthén & Muthén, Los Angeles, CA.

Pearl, J. (2000) *Causality: Models, Reasoning, and Inference*, Cambridge University Press, https://doi.org/10.1017/S0266466603004109.

R Core Team (2019) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria [online] https://www.R-project.org/ (accessed 1 December 2020).

Raykov, T. and Marcoulides, G.A. (2014) 'Identifying useful auxiliary variables for incomplete data analyses: a note on a group difference examination approach', *Educational and Psychological Measurement*, Vol. 74, No. 3, pp.537–550, https://doi.org/10.1177/0013164413511326

Rosseel, Y. (2012) 'lavaan: An R package for structural equation modeling', *Journal of Statistical Software*, Vol. 48, No. 2, pp.1–36 [online] https://www.jstatsoft.org/article/view/v048i02 (accessed 1 December 2020).

Rubin, D.B. (1976) 'Inference and missing data', *Biometrika*, Vol. 63, No. 3, pp.581–592, https://doi.org/10.1093/biomet/63.3.581.

Savalei, V. and Bentler, P.M. (2009) 'A two-stage approach to missing data: theory and application to auxiliary variables', *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 16, No. 3, pp.477–497, https://doi.org/10.1080/10705510903008238.

Schafer, J.L. (1997) *Analysis of Incomplete Multivariate Data*, Chapman and Hall/CRC, https://doi.org/10.1201/9780367803025.

Schafer, J.L. (2003) 'Multiple imputation in multivariate problems when the imputation and analysis models differ', *Statistica Neerlandica*, Vol. 57, No. 1, pp.19–35, https://doi.org/10.1111/1467-9574.00218.

Schafer, J.L. and Graham, J.W. (2002) 'Missing data: our view of the state of the art', *Psychological Methods*, Vol. 7, No. 2, pp.147–177, https://doi.org/10.1037/1082-989X.7.2.147.

Thoemmes, F. and Rose, N. (2014) 'A cautious note on auxiliary variables that can increase bias in missing data problems', *Multivariate Behavioral Research*, Vol. 49, No. 5, pp.443–459, https://doi.org/10.1080/00273171.2014.931799.

Yuan, K-H., Tong, X. and Zhang, Z. (2015) 'Bias and efficiency for SEM with missing data and auxiliary variables: two-stage robust method versus two-stage ML', *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 22, No. 2, pp.178–192, https://doi.org/10.1080/10705511.2014.935750.

Notes

1 Analyses with a smaller number of *B* variables (starting from including one *B* variable to all the 10 *B* variables) were also implemented to understand the estimation accuracy when only a few correlates were included in the analyses. Results (not reported in Tables 3–5) found that, the more *B* variables included, the more accurate parameter estimates were obtained.

Appendix

Results for loading parameter estimates

Tables A1–A4 present the means and empirical standard errors of estimates for the first loading parameter. Because all items had the same population loading value, results for the other loading estimates were similar and were thus not reported.

Unlike the bias in factor correlation estimates, relative bias in loadings was in general smaller than 10%. Tables A1–A4 show a consistent pattern that listwise deletion, as well as the analytical conditions with C_1–C_{10} included as the AVs, resulted in a small level of bias in loading estimates. Such bias could be reduced by including A or B_1–B_{10} as the AVs, or by including all the available AVs.

Table A1 Parameter estimates of the first loading when sample size was 100,000

ρ_{AX}	ρ_{AF}	λ	*Com*	*LW*	*A*	*B*	*C*	*ALL*
					10% missingness			
0.2	0.3	0.67	0.668	0.664	0.669	0.668	0.664	0.668
0.2	0.5	0.40	0.395	0.381	0.393	0.392	0.381	0.395
0.2	0.7	0.29	0.279	0.261	0.279	0.277	0.261	0.279
0.3	0.5	0.60	0.599	0.585	0.599	0.597	0.585	0.599
0.3	0.7	0.43	0.430	0.403	0.430	0.426	0.403	0.430
0.4	0.5	0.80	0.798	0.787	0.798	0.797	0.787	0.798
0.4	0.7	0.57	0.575	0.547	0.575	0.572	0.547	0.575
					25% missingness			
0.2	0.3	0.67	-	0.656	0.663	0.663	0.656	0.665
0.2	0.5	0.40	-	0.382	0.401	0.398	0.382	0.401
0.2	0.7	0.29	-	0.261	0.294	0.290	0.261	0.293
0.3	0.5	0.60	-	0.578	0.603	0.598	0.578	0.602
0.3	0.7	0.43	-	*0.375*	0.417	0.412	*0.375*	0.423
0.4	0.5	0.80	-	0.781	0.800	0.797	0.781	0.800
0.4	0.7	0.57	-	0.524	0.571	0.565	0.524	0.572
					50% missingness			
0.2	0.3	0.67	-	0.659	0.669	0.667	0.659	0.669
0.2	0.5	0.40	-	0.371	0.401	0.394	0.371	0.399
0.2	0.7	0.29	-	*0.258*	0.308	0.297	*0.258*	0.296
0.3	0.5	0.60	-	0.566	0.601	0.593	0.566	0.598
0.3	0.7	0.43	-	*0.357*	0.420	0.408	*0.357*	0.427
0.4	0.5	0.80	-	0.776	0.802	0.798	0.776	0.803
0.4	0.7	0.57	-	*0.499*	0.572	0.561	*0.499*	0.569

Notes: Values that had greater than 10% relative bias (i.e., > 0.33 or < 0.27) were italics; Com is for complete data; LW is listwise deletion; *A* was the direct cause of missingness and correlated with the variables of interests; *B* represented the analytical conditions where B_1–B_{10} (i.e., correlates that are not the direct cause of missingness) were included; *C* represented the analytical conditions where C_1–C_{10} (independent random noises) were included.

Table A2 Means and empirical standard errors of loading estimates when sample size was 600

ρ_{AX}	ρ_{AF}	λ	*Com*		*LW*		*A*		*B*		*C*		*ALL*	
			Est	*SE*	*Est*	*SE*	*Est*	*SE*	*Est*	*SE*	*Est*	*SE*	*Est*	*SE*
								10% missingness						
0.2	0.3	0.67	0.665	0.029	0.660	0.032	0.665	0.031	0.664	0.031	0.660	0.032	0.665	0.029
0.2	0.5	0.40	0.401	0.058	0.390	0.063	0.402	0.062	0.400	0.062	0.390	0.063	0.401	0.058
0.2	0.7	0.29	0.288	0.082	0.273	0.096	0.292	0.087	0.289	0.088	0.273	0.096	0.291	0.080
0.3	0.5	0.60	0.599	0.036	0.585	0.039	0.599	0.039	0.597	0.039	0.585	0.039	0.599	0.036
0.3	0.7	0.43	0.429	0.054	0.404	0.063	0.429	0.061	0.426	0.061	0.404	0.063	0.429	0.054
0.4	0.5	0.80	0.799	0.017	0.788	0.019	0.799	0.018	0.797	0.019	0.788	0.019	0.799	0.017
0.4	0.7	0.57	0.570	0.038	0.541	0.043	0.570	0.042	0.566	0.042	0.541	0.043	0.570	0.038
								25% missingness						
0.2	0.3	0.67	-	-	0.656	0.035	0.664	0.034	0.663	0.034	0.656	0.035	0.666	0.029
0.2	0.5	0.40	-	-	0.376	0.070	0.396	0.069	0.393	0.069	0.376	0.070	0.397	0.056
0.2	0.7	0.29	-	-	0.264	0.116	0.290	0.100	0.288	0.102	0.262	0.113	0.289	0.080
0.3	0.5	0.60	-	-	0.576	0.043	0.599	0.043	0.595	0.042	0.576	0.043	0.598	0.035
0.3	0.7	0.43	-	-	*0.382*	0.070	0.426	0.067	0.420	0.066	*0.382*	0.070	0.428	0.054
0.4	0.5	0.80	-	-	0.783	0.022	0.801	0.021	0.798	0.021	0.783	0.022	0.801	0.018
0.4	0.7	0.57	-	-	0.520	0.049	0.569	0.048	0.563	0.046	0.520	0.049	0.570	0.037
								50% missingness						
0.2	0.3	0.67	-	-	0.654	0.045	0.664	0.046	0.662	0.045	0.654	0.045	0.667	0.030
0.2	0.5	0.40	-	-	0.373	0.095	0.400	0.093	0.396	0.090	0.373	0.095	0.401	0.058
0.2	0.7	0.29	-	-	0.266	0.147	0.301	0.135	0.291	0.130	0.263	0.144	0.290	0.082
0.3	0.5	0.60	-	-	0.565	0.054	0.598	0.055	0.592	0.053	0.565	0.054	0.598	0.033
0.3	0.7	0.43	-	-	*0.372*	0.098	0.433	0.091	0.422	0.087	*0.372*	0.098	0.430	0.053
0.4	0.5	0.80	-	-	0.773	0.029	0.798	0.029	0.795	0.027	0.773	0.029	0.800	0.018
0.4	0.7	0.57	-	-	*0.500*	0.063	0.571	0.062	0.559	0.058	*0.501*	0.063	0.571	0.040

Notes: Values that had greater than 10% relative bias were italics; Com is for complete data; LW is listwise deletion; *A* was the direct cause of missingness and correlated with the variables of interests; *B* represented the analytical conditions where B_1–B_{10} (i.e., correlates that are not the direct cause of missingness) were included; *C* represented the analytical conditions where C_1–C_{10} (independent random noises) were included.

Table A3 Mean loading estimates across replications and empirical standard error of the factor correlation when sample size was 300

ρ_{AX}	ρ_{AF}	λ	*Com*		*LW*		*A*		*B*		*C*		*ALL*	
			Est	*SE*	*Est*	*SE*	*Est*	*SE*	*Est*	*SE*	*Est*	*SE*	*Est*	*SE*
10% missingness														
0.2	0.3	0.67	0.668	0.040	0.664	0.043	0.669	0.043	0.668	0.043	0.664	0.043	0.668	0.040
0.2	0.5	0.40	0.398	0.084	0.384	0.094	0.396	0.091	0.394	0.091	0.384	0.094	0.398	0.084
0.2	0.7	0.29	0.294	0.119	0.284	0.135	0.302	0.133	0.301	0.134	0.285	0.134	0.300	0.114
0.3	0.5	0.60	0.599	0.050	0.585	0.054	0.599	0.054	0.597	0.054	0.585	0.054	0.599	0.050
0.3	0.7	0.43	0.422	0.077	0.396	0.086	0.422	0.083	0.418	0.083	0.396	0.086	0.422	0.077
0.4	0.5	0.80	0.799	0.024	0.788	0.027	0.799	0.027	0.797	0.027	0.788	0.027	0.799	0.024
0.4	0.7	0.57	0.569	0.055	0.540	0.063	0.569	0.061	0.566	0.061	0.540	0.063	0.569	0.055
25% missingness														
0.2	0.3	0.67	-	-	0.657	0.05	0.664	0.050	0.663	0.049	0.657	0.050	0.665	0.042
0.2	0.5	0.40	-	-	0.378	0.104	0.398	0.103	0.396	0.101	0.378	0.104	0.399	0.081
0.2	0.7	0.29	-	-	0.282	0.146	0.308	0.140	0.304	0.14	0.280	0.146	0.304	0.118
0.3	0.5	0.60	-	-	0.575	0.062	0.598	0.062	0.595	0.061	0.575	0.062	0.599	0.049
0.3	0.7	0.43	-	-	*0.386*	0.105	0.428	0.098	0.422	0.098	*0.385*	0.105	0.430	0.076
0.4	0.5	0.80	-	-	0.782	0.032	0.800	0.031	0.797	0.031	0.782	0.032	0.800	0.025
0.4	0.7	0.57	-	-	0.525	0.072	0.573	0.070	0.566	0.067	0.525	0.072	0.572	0.053
50% missingness														
0.2	0.3	0.67	-	-	0.652	0.062	0.661	0.063	0.660	0.062	0.652	0.062	0.662	0.042
0.2	0.5	0.40	-	-	0.376	0.135	0.400	0.136	0.398	0.13	0.375	0.134	0.401	0.082
0.2	0.7	0.29	-	-	0.313	0.195	0.345	0.172	0.333	0.169	0.310	0.184	0.307	0.115
0.3	0.5	0.60	-	-	0.568	0.079	0.600	0.080	0.594	0.078	0.569	0.079	0.603	0.052
0.3	0.7	0.43	-	-	*0.382*	0.138	0.434	0.133	0.427	0.128	*0.382*	0.139	0.433	0.073
0.4	0.5	0.80	-	-	0.773	0.042	0.797	0.043	0.794	0.04	0.773	0.042	0.801	0.025
0.4	0.7	0.57	-	-	*0.499*	0.098	0.564	0.095	0.557	0.09	*0.499*	0.099	0.571	0.055

Notes: Values that had greater than 10% relative bias were italic; Com is for complete data; LW is listwise deletion; *A* was the direct cause of missingness and correlated with the variables of interests; *B* represented the analytical conditions where B_1–B_{10} (i.e., correlates that are not the direct cause of missingness) were included; *C* represented the analytical conditions where C_1–C_{10} (independent random noises) were included.

Table A4 Mean loading estimates across replications and empirical standard error of the factor correlation when sample size was 100

ρ_{AX}	ρ_{AF}	λ	*Com*		*LW*		*A*		*B*		*C*		*ALL*	
			Est	*SE*	*Est*	*SE*	*Est*	*SE*	*Est*	*SE*	*Est*	*SE*	*Est*	*SE*
10% missingness														
0.2	0.3	0.67	0.661	0.071	0.656	0.076	0.661	0.076	0.660	0.076	0.656	0.076	0.661	0.071
0.2	0.5	0.40	0.407	0.152	0.396	0.167	0.406	0.163	0.406	0.160	0.395	0.168	0.407	0.150
0.2	0.7	0.29	*0.365*	0.172	*0.355*	0.182	*0.361*	0.176	*0.364*	0.177	*0.354*	0.175	*0.348*	0.172
0.3	0.5	0.60	0.598	0.088	0.586	0.095	0.598	0.095	0.597	0.095	0.585	0.095	0.598	0.088
0.3	0.7	0.43	0.426	0.146	0.410	0.158	0.430	0.155	0.428	0.153	0.411	0.160	0.431	0.142
0.4	0.5	0.80	0.799	0.043	0.788	0.048	0.799	0.047	0.797	0.047	0.788	0.048	0.799	0.043
0.4	0.7	0.57	0.570	0.093	0.542	0.108	0.570	0.103	0.566	0.103	0.542	0.108	0.57	0.093
25% missingness														
0.2	0.3	0.67	-	-	0.661	0.088	0.667	0.088	0.666	0.088	0.660	0.088	0.667	0.071
0.2	0.5	0.40	-	-	0.397	0.166	0.411	0.169	0.411	0.163	0.394	0.168	0.417	0.144
0.2	0.7	0.29	-	-	*0.371*	0.233	*0.387*	0.226	*0.374*	0.220	*0.364*	0.219	*0.336*	0.183
0.3	0.5	0.60	-	-	0.573	0.114	0.593	0.114	0.591	0.113	0.573	0.114	0.600	0.091
0.3	0.7	0.43	-	-	0.406	0.173	0.446	0.171	0.441	0.168	0.407	0.175	0.447	0.135
0.4	0.5	0.80	-	-	0.778	0.058	0.793	0.058	0.792	0.057	0.778	0.058	0.797	0.046
0.4	0.7	0.57	-	-	0.526	0.124	0.571	0.121	0.566	0.118	0.526	0.124	0.574	0.092
50% missingness														
0.2	0.3	0.67	-	-	0.658	0.112	0.664	0.115	0.664	0.115	0.658	0.114	0.663	0.073
0.2	0.5	0.40	-	-	0.426	0.200	0.439	0.195	0.438	0.198	0.425	0.198	0.428	0.147
0.2	0.7	0.29	-	-	*0.356*	0.216	*0.340*	0.224	*0.407*	0.220	*0.375*	0.227	*0.315*	0.217
0.3	0.5	0.60	-	-	0.567	0.148	0.587	0.154	0.589	0.147	0.570	0.148	0.605	0.086
0.3	0.7	0.43	-	-	0.428	0.208	0.459	0.212	0.458	0.197	0.421	0.199	0.452	0.136
0.4	0.5	0.80	-	-	0.771	0.075	0.787	0.076	0.788	0.076	0.771	0.077	0.799	0.047
0.4	0.7	0.57	-	-	*0.500*	0.168	0.555	0.164	0.55	0.158	*0.498*	0.168	0.573	0.094

Notes: Values that had greater than 10% relative bias were bolded; Com is for complete data; LW is listwise deletion; *A* was the direct cause of missingness and correlated with the variables of interests; *B* represented the analytical conditions where B_1–B_{10} (i.e., correlates that are not the direct cause of missingness) were included; *C* represented the analytical conditions where C_1–C_{10} (independent random noises) were included.