Omissions by Design in a Survey: Is This a Good Choice when using Structural Equation Models?

Paula C. R. Vicente
Lusófona University, COPELABS, Intrepid Lab, Lisbon, Portugal
p951@ulusofona.pt

ARTICLE INFO

Original Scientific Article

Article history:
Received May 2024
Revised September 2024
Accepted September 2024

JEL Classification
C38, C63

Keywords:
Omissions by design
Structural Equation Model
Survey

UDK: 303.425
DOI: 10.2478/ngoe-2024-0018

Cite this article as: Vincente, P. C. R. (2024). Omissions by Design in a Survey: Is This a Good Choice when using Structural Equation Models? *Naše gospodarstvo/Our Economy, 70*(3), *83-91*. DOI: 10.2478/ngoe-2024-0018

Abstract

Missing observations can arise due to the effort required to answer many questions in long surveys and the cost required to obtain some responses. Implementing a planned missing design in surveys helps reduce the number of questions each respondent needs to answer, thereby lowering survey fatigue and cutting down on implementation costs. The three-form and the two-method design are two different types of planned missing designs. An important consideration when designing a study with omissions by design is to know how it will affect statistical results. In this work, a simulation study is conducted to analyze how the usual fit measures, root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), comparative fit index (CFI), and Tucker-Lewis index (TLI) perform in the adjustment of a Structural Equation Model. The results revealed that the CFI, TLI, and SRMR indices exhibit sensitivity to omissions with small samples, low factor loadings and large models. Overall, this study contributes to our understanding of the importance of considering omissions by design in market research.

Introduction

Incomplete or missing data is a common challenge in numerous studies across various fields, such as market research. Missing observations can arise not only due to the effort required to answer many questions in long surveys but also because of the cost required to obtain some responses. According to some authors, omissions represent one of the most significant statistical issues in research, with the usual practice being the exclusion of nonresponses from data modelling. Implementing a planned missing data design can help mitigate this problem by reducing the number of survey questions that each respondent must answer without reducing the total number of survey questions. The effort and cost that respondents must spend to complete the survey are thus also minimized, which increases the

quantity and quality of data collected in a study. In addition, planned missing designs allow to implementation of a survey while minimizing its overall cost. Two different types of planned missing design include the three-form design and the two-method design (Rioux, Lewin, Odejimi & Little, 2020; Graham, Taylor, Olchowski & Cumsville, 2006; Graham, Hofer & Mackinnon, 1996).

Researchers may be concerned that planning for a missing design could bias the results of the analysis. However, it is important to emphasize that this type of design can be handled appropriately by modern missing data techniques, such as Full Information Maximum Likelihood (FIML) and Multiple Imputation (MI) without erroneous analysis results (Little, Jorgensen, Lang & Moore, 2013; Enders, 2010; Azar, 2002; Arbuckle, 1996).

Structural Equation Models (SEM) include both measurement and structural models. The measurement model focuses on the relationship between latent and observed variables, while the structural model focuses on the relationship between latent variables (Bollen, 1989). This study considers a measurement model, which is crucial for examining a multidimensional concept and developing a scale.

When fitting a Structural Equation Model (SEM) to the data, multiple measures could be utilized to assess the compatibility of the theoretical model with the observed data, with the Root Mean Square Error of Approximation (RMSEA), Standardized Root Mean Square Residual (SRMR), Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) being the most commonly used (Hair, Black, Babin, Anderson & Tatham, 2006; Bandalous & Finney, 2010; Nye, 2022).

According to Rioux, Lewin, Odejimi &Little (2020) and Little, Jorgensen, Lang & Moore (2013), an important consideration when designing a study with a planned missing design is to know how this will affect statistical results. In this way, Wu & Jia, 2021, Moore, Lang & Grandfield (2020), Rioux, Lewin, Odejimi & Little (2020), Rhemtulla & Hancock, 2016, Schoemann, Miller, Pornprasertmanit & Wu (2014) studies examined the impact of a planned missing data design on parameter estimate bias, standard error bias, model convergence, and power analysis.

In addition, Lawes, Schultze & Eid (2020) and Rhemtulla & Hancock (2016) in their studies highlighted the cost benefits of omissions by design. The importance of the size of the model and sample size has been emphasized by Kenny, Kanishan & McCoach, (2015) and Kenny & McCoach, (2003).

Given the importance of fit indices in assessing the adjustment of a SEM, this discussion explores how the usual fit measures behave when intentional omissions are incorporated into analyses conducted using a SEM.

Planned Missing Data Designs

Two different types of planned missing data designs include the three-form design and the two-method design (Rioux, Lewin, Odejimi & Little, 2020; Graham, Taylor, Olchowski & Cumsville, 2006). For these designs, the critical point is the random assignment of the questions to the respondents because the missing data mechanism must be known.

According to Rubin (1976), there are three types of missing data mechanisms: Missing Completely at Random (MCAR), Missing at Random (MAR) and Missing Not at Random (MNAR). MCAR is present when the probability of missing data on a variable is not related to observed or missing values for any of the variables. MAR happens when the probability of missing data on a variable is related to some other measured variable (or variables) in the study. Finally, MNAR occurs when the probability of missing data on a variable is related to the missing values themselves.

In a planned missing design, when a random assignment is done, the non-responses from a three-form design and a two-method design follow a Missing Completely at Random (MCAR) mechanism (Enders, 2010).

Three-Form Design

The three-form design is a particular example of a planned missing data design. With this design, the survey's questions are divided into four groups: X, A, B, and C. After completing the questions in the X block, each participant is randomly assigned to complete two of the remaining blocks (A, B, and C). As a result, among

all participants, one-third answered questions in set XAB, one-third answered XAC, and one-third answered XBC, rather than responding to questions in the four groups (see Table 1). Through this approach, the overall number of questions asked is not decreased, but the number of questions answered by each participant is decreased from 100% to 67%.

According to Graham, Taylor, Olchowski, and Cumsville (2006), questions in the X group are answered by all survey respondents, hence these should include the most crucial questions for the investigation.

Although in a more conservative approach, the questions in group X should be answered by all participants, most of the time are sociodemographic questions. According to Lang & al (2020) and Moore, Lang & Grandfield (2020), if there are 13 or more questions, the X block could contain scale and sociodemographic items randomly assigned. According to Schoemann, Miller, Pornprasertmanit & Wu (2014), each group may not always have the same number of questions. Different applications of a three-form design can be seen in Fürst (2020) and Roche & al. (2019).

Table 1

The three-form design

Question Set					
Form	*% of sample*	*X*	*A*	*B*	*C*
1	1/3	O	O	O	NA
2	1/3	O	O	NA	O
3	1/3	O	NA	O	O

Note: O-observed value, NA-not available

Source: Author's work

Two-Method Design

The two-method design is useful in situations in which researchers face a choice between two measures of a construct: i. a more expensive, more intrusive, or time-consuming measure, designated gold standard measure, and ii. a biased measure that is inexpensive, non-intrusive, or timesaving. The least expensive measure will be observed for all participants, but only a small proportion (e.g., 1/3) of participants will be randomly selected to receive the expensive measure (see Table 2). As such, given the same budget with a two-method design, it is possible to have more participants than in a complete data design using only the expensive measure (Wu & Jia, 2021; Lawes, Schultze & Eid, 2020).

Examples of applications can be found in Graham, Taylor, Olchowski & Cumsville (2006) and Garnier-Villareal, Rhemtulla & Little (2014). Such a situation can happen in market studies, when it is required to have expensive measures, including tests or observations of consumer behaviour.

The basic idea of the two-method design is that the subsample containing data on all measures can be used to usefully define the construct under study and to control the systematic bias of the cheap method (Figure 1). Conversely, having participants with only data on the expensive measures allows for an increase in the total sample size and the statistical power, while keeping the costs low.

Randomly assigning participants creates a missing data pattern that satisfies the MCAR assumption. Nevertheless, if the participants are selected according to a certain characteristic (e.g., to be an expert or their age), then the data omission mechanism is considered MAR (Wu & Jia, 2021).

Table 2

Two-method design

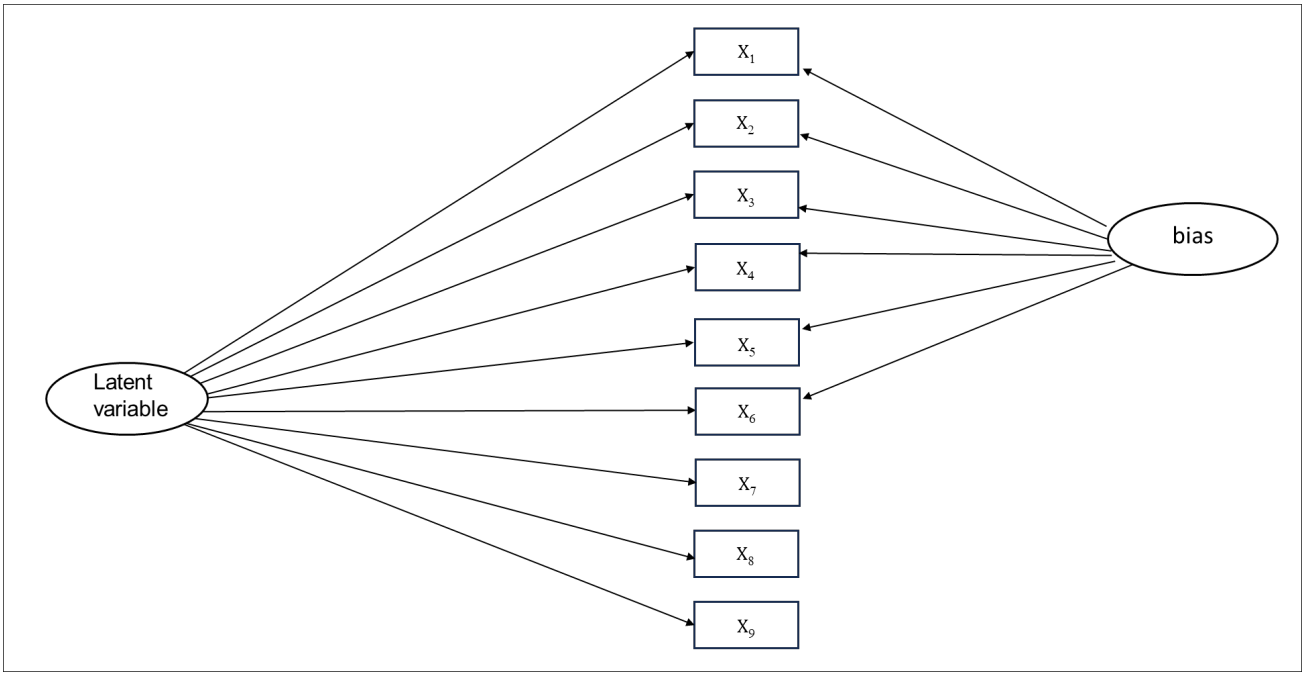
	Question set	
% of sample	*Cheap*	*Expensive*
1/3	O	O
2/3	O	NA

Note: O-observed value, NA-not available

Source: Author's work

Figure 1

Two-method design structure with three gold-standard measures (X_7 to X_9) and six (X_1 to X_6) cheap measures



Source: Author's work

Methods

Fit Measures

Quantifying the fit of a SEM is important to determine the best model and can be done using different fit indices, based on distinct criteria (Yuan, 2015). Although several fit indices are available for this purpose, the most utilized ones are RMSEA, SRMR, CFI, and TLI. On the other hand, most researchers agree that multiple fit indices should be used to ensure the reliability of model fit (Bandalous & Finney, 2010; Hair, Black, Babin, Anderson & Tatham, 2006). Given how often these fit indices are used, it is important to comprehend how well they perform when there are intentional omissions.

RMSEA (Steiger, 1990) and SRMR (Jöreskog & Sörbom, 2022) are absolute fit indices, and a better fit is indicated by lower values. When their values are below 0.05, it is considered a good fit (Hu & Bentler,1999; Kaplan, 2009; Kline, 2023).

The algebraic expression for RMSEA is

$$\text{RMSEA} = \sqrt{max\left\{\left(\frac{F(\boldsymbol{S},\boldsymbol{\Sigma}(\hat{\theta}))}{df} - \frac{1}{n-1}\right), 0\right\}} \qquad (1)$$

where F is the minimum value of a discrepancy function, **S** is the observed matrix and **∑** is the model implied matrix.

SRMR is calculated by

$$\text{SRMR} = \sqrt{\frac{\sum_{i=1}^{p}\sum_{j=1}^{i}\left[{}^{(s_{ij}-\hat{\sigma}_{ij})}/_{(s_{ii}s_{jj})}\right]^2}{{}^{p(p+1)}/_{2}}} \qquad (2)$$

with s_{ij} an element of **S**, $\hat{\sigma}_{ij}$ an element of $\boldsymbol{\Sigma}$ and p the number of observed variables.

CFI (Bentler, 1990) and TLI (Tucker & Lewis, 1973) are incremental fit indices, with higher values indicating a better fit. When obtained values are equal or exceed 0.95, a good fit is considered to have existed (Hu & Bentler,1999).

CFI and TLI have the following expressions

$$\text{CFI} = 1 - \frac{max[(\chi_t^2 - df_t), 0]}{max[(\chi_0^2 - df_0), 0]} \qquad (3)$$

and

$$\mathrm{TLI} = \frac{\left(\chi_0^2/df_0\right) - \left(\chi_t^2/df_t\right)}{\left(\chi_0^2/df_0\right) - 1} \quad (4)$$

χ_0^2 , χ_t^2 and df_0, df_t represent, for the baseline model and for the model under analysis, the chi-square statistic and the degrees of freedom.

Simulation Conditions

This work examines the impact of nonresponses on the SEM fit indices RMSEA, SRMR, CFI, and TLI due to different planned missing designs, three-form designs, and two-method designs. Three different sample sizes (small, medium, and large) of 200, 500, and 1000 observations are considered. Three different model sizes with 9 (small model), 18 (medium model), and 36 (large model) indicators. In Figure 2, a measurement model containing 9 indicators, is displayed.

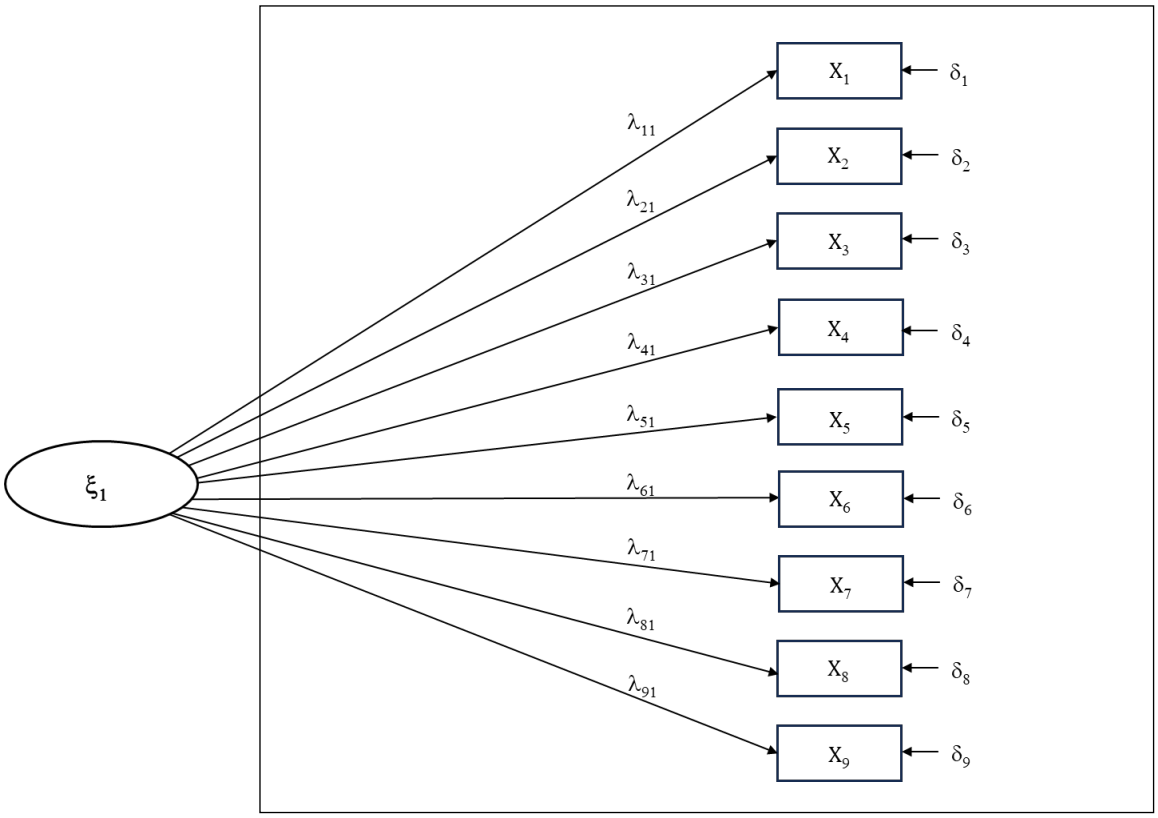
A measurement model is defined by

$$\boldsymbol{X} = \boldsymbol{\Lambda_X \xi} + \boldsymbol{\delta} \quad (5)$$

where $\boldsymbol{X}$, $\boldsymbol{\xi}$ and $\boldsymbol{\delta}$ are, respectively, the vectors of the indicators (observed variables), latent variables and residual terms, $\boldsymbol{\Lambda_X}$ is the matrix of factor loadings, and it is assumed that residual terms are normally distributed with mean zero and variance-covariance diagonal matrix $\boldsymbol{\Theta_\delta}$. Consequently, the generated data have a normal distribution. The parameters of the model are the factor loadings (λ_{ij}; $i = 1, \ldots, 9$; $j = 1$) and the factor covariance (ϕ_{11}), that is assumed to be 1. Two different values for factor loadings, 0.4 and 0.8 (low and high values), are considered. The variances of the residual terms are obtained by $1 - \lambda^2$.

Figure 2

Measurement model with 1 latent variable (ξ_1) and 9 indicators (X_1 to X_9). The λ_{ij} are the factor loadings and δ_i are the residual terms



Source: Author's work

Estimating a SEM involves determining the parameter values that result in a variance–covariance matrix, **∑**, that best approximates the empirical variance–covariance matrix, **S**, obtained from the data (Bollen,

1989). This process is guided by a discrepancy function (F), a mathematical procedure used to minimize the difference between the observed matrix (**S**) and the model implied matrix **∑**. The most used discrepancy function in SEM is based on the Maximum Likelihood (ML), which finds the parameter estimates that maximize the likelihood of observing the sample data, assuming the model is correct. For handling with omissions, the FIML method is applied. The key difference between FIML and traditional ML is that FIML computes the likelihood for each case based only on the variables that are observed for that case. This approach ensures that no information is lost, as it uses all available data points, even when some variables are missing, to estimate the model parameters. FIML relies on the MAR assumption (Enders, 2010).

Datasets with complete observations and missing observations were generated according to two different planned missing designs. The percentage of missing data remained the same in all generated datasets, regardless of the number of indicators in the model, and that used by Schoemann, Miller, Pornprasertmanit & Wu (2014).

A total of one thousand samples were generated at random, considering all possible combinations of the manipulated variables, which included both complete and incomplete datasets. The number of replications utilized corresponds with the guidelines established by McNeisch, An & Hancock (2018) and Schoemann, Miller, Pornprasertmanit & Wu (2014). The simulation study used the simsem package in R to manipulate parameter values, introduce missing data, and facilitate a summary of simulation results (Pornprasertmanit, Miller, Schoemann, & Jorgensen, 2021).

Results and Discussion

The results of the proposed simulation study are presented in Table 3. The displayed values correspond to the mean value of the indices of interest derived from the 1000 datasets. Larger sample sizes produce improved outcomes, with lower values for RMSEA and SRMR, and higher values for CFI and TLI. Similarly, the obtained values improve with higher factor loadings, except for the RMSEA index. For this index the values are similar, and all of them are acceptable (under 0.05, the cutoff).

The RMSEA index did not seem to be affected by missing data, which has likewise been noted by Davey, (2005), Hoyle, (2012), and Zhang & Savalei (2023). They investigated the possibility of a bias in RMSEA values and determined that elements such as the missing data mechanism have an important role in deciding whether biases occur.

For the SRMR index, the obtained values are larger with larger models, and worse with a three-form design than with a two-method design, for all sample sizes. For a small sample size and low factor loadings, the values are above the acceptable (0.05), with 9 indicators and data from a three-form design, with 18 indicators and with omissions by design, and with 36 indicators even for complete data. According to Jia, Moore, Kinai, Crowe, Schoemann & Little (2014), when the data have omissions by design, other factors beyond sample size play an important role. Kenny, Kanishan & McCoach, (2015) have shown that sample size and large models' impact in the adjustment of the model.

The CFI and TLI indices show values lower than the acceptable (0.95) in small sample sizes for large models (36 indicators) with low factor loadings, even with complete data. A deviation in CFI values was indicated by Zhang & Savalei (2023) due to using FIML for handling omissions. McNeish & Wolf (2021) and Hancock & Mueller (2011) have indeed pointed out that, in small sample sizes, the impact of the factor loading values measurements in the evaluation of model adjustment should be considered.

Vicente (2023) found similar behaviour for all the fit measures, RMSEA, SRMR, CFI, and TLI, in a simulation study with data from a three-form design with distinct models with and without considering model misspecification. Shi, Lee & Maydeu-Olivares (2019), Moshagen (2012) and Fan, Thompson & Wang (1999) found that in the presence of small sample sizes and large models have worse values for the considered fit indices. Finally, in their work, Cangur & Ercan (2015) showed that the fit index with the worst performance was SRMR.

Table 3
Obtained fit indices for a model with 9, 18 and 36 indicators

	Number of indicators		*9*			*18*			*36*		
	Factor loading	*Sample size*	*a*	*b*	*c*	*a*	*b*	*c*	*a*	*b*	*C*
RMSEA	*0.4*	*200*	0.016	0.017	0.016	0.013	0.016	0.015	0.018	0.025	0.018
		500	0.009	0.010	0.010	0.007	0.007	0.007	0.007	0.004	0.007
		1000	0.006	0.007	0.007	0.005	0.005	0.005	0.004	0.004	0.004
	0.8	*200*	0.016	0.017	0.016	0.014	0.016	0.014	0.018	0.025	0.018
		500	0.009	0.009	0.010	0.007	0.008	0.007	0.007	0.009	0.007
		1000	0.006	0.006	0.007	0.005	0.005	0.005	0.004	0.004	0.004
SRMR	*0.4*	*200*	0.042	**0.052**	0.047	0.05	**0.062**	**0.053**	**0.055**	**0.070**	**0.057**
		500	0.026	0.032	0.030	0.032	0.038	0.034	0.035	0.029	0.036
		1000	0.019	0.023	0.021	0.022	0.027	0.024	0.024	0.029	0.025
	0.8	*200*	0.018	0.023	0.020	0.022	0.028	0.025	0.024	0.033	0.025
		500	0.011	0.014	0.012	0.014	0.017	0.016	0.015	0.019	0.016
		1000	0.008	0.010	0.009	0.010	0.012	0.011	0.010	0.013	0.011
CFI	*0.4*	*200*	0.971	0.963	0.967	0.974	0.960	0.967	**0.946**	**0.902**	**0.944**
		500	0.99	0.987	0.987	0.99	0.988	0.990	0.989	0.994	0.998
		1000	0.995	0.994	0.993	0.996	0.994	0.995	0.996	0.994	0.995
	0.8	*200*	0.997	0.997	0.997	0.997	0.995	0.996	0.993	0.985	0.992
		500	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.998	0.998
		1000	1	0.999	0.999	1	0.999	0.999	0.999	0.999	0.999
TLI	*0.4*	*200*	0.999	0.994	1	0.987	0.969	0.981	**0.945**	**0.896**	**0.942**
		500	1	0.999	1	0.997	0.996	0.998	0.992	0.996	0.991
		1000	1	1	1	0.999	0.998	0.999	0.998	0.996	0.998
	0.8	*200*	0.999	0.999	0.999	0.998	0.996	0.997	0.992	0.985	0.992
		500	1	1	1	1	0.999	1	0.999	0.998	0.999
		1000	1	1	1	1	1	1	1	1	1

Notes: Column a - results for a complete case; Column b – results from a three-form design; Column c – results from a two-method design
Source: Author's work

Conclusion

This study explores the effect of omissions by design in the fit indices RMSEA, SRMR, CFI and TLI, the most popular fit measures used to evaluate the adjustment of a SEM. The use of the missing data analysis procedure FIML technic makes omissions by design possible (Rioux, Lewin, Odejimi & Little, 2020; Enders, 2010) since it includes intentionally missing data designs. The designs under considerations are a three-form design and a two-method design (Graham, Taylor, Olchowski & Cumsville, 2006). The aim of implementing these approaches is to improve data quality, reduce respondent fatigue, and minimize overall cost. Such designs are underutilized by researchers and can be very useful in areas like market research (Enders, 2010).

Our simulation study revealed that the CFI, TLI, and SRMR indices exhibit sensitivity to non-responses, particularly in scenarios involving small sample sizes, low factor loadings, and a high number of indicators. However, they perform worse when the data are from a three-form design than from a two-method design. The RMSEA index yields the best results in the presence of nonresponses caused by a planned missing data design. Overall, this study contributes to our understanding of the importance of considering omissions by design in market research.

This work has, however, some limitations, namely that the data generated are cross-sectional. In future work, it will be interesting to consider longitudinal data and other models, such as the latent growth curve model.

References

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In Marcoulides, G. A., & Schumacker, R. E. (eds.), *Advanced structural equation modelling*, 243–277. Erlbaum.

Azar, B. (2002). Finding a solution for missing data. *Monitor on Psychology*, *33*,70.

Bandalous, D. L., & Finney, S. J. (2010). Factor analysis: Exploratory and confirmatory. In Hancock, G. R. & Mueller, R. O. (eds.), *The reviewer's guide to quantitative methods in the social sciences*. New York, NY: Routledge. DOI: https://doi.org/10.4324/9780203861554-15

Bentler, P. M. (1990). Comparative Fit Indexes in Structural Equation Models. *Psychological Bulletin*, 107, 238–246. DOI: https://doi.org/10.1037/0033-2909.107.2.238

Bollen, K. (1989). *Structural Equations with Latent Variables*. New York, USA: Wile

Cangur, S., & Ercan, I. (2015). Comparison of model Fit Indices Used in Structural Equation Modeling Under Multivariate Normality, *14* (1), 152–167. DOI: https://doi.org/10.22237/jmasm/1430453580

Davey, A. (2005). Issues in evaluating model fit with missing data. *Struct. Equ. Model. Multidiscip. J., 12*, 578–597. DOI: https://doi.org/10.1207/s15328007sem1204_4

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

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods and model specification on structural equation modelling fit indexes. *Struct. Equ. Model. Multidiscip. J.,* 6, 56–83. DOI: https://doi.org/10.1080/10705519909540119

Fürst, G. (2020). Measuring creativity with planned missing data. *The Journal of Creative Behavior*, *54*(1), 150–164. DOI: https://doi.org/10.1002/jocb.352

Garnier-Vilarreal, M., Rhemtulla, M., & Little, T.D. (2014). Two-method planned missing designs for longitudinal research. *International Journal of Behavioral Development*, *38*(5), 411–422. DOI: https://doi.org/10.1177/0165025414542711

Graham, J., Taylor, B., Olchowski, A., & Cumsville, P. (2006). Planned missing data designs in psychological research. *Psychological Methods*, *11*, 323–343. DOI: https://doi.org/10.1037/1082-989x.11.4.323

Graham,J., Hofer,S. & Mackinnon,D. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research*, *31*, 197–218.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (6th ed). Columbus, Oh: Pearson.

Hancock, G. R. & Mueller, R. O. (2011). The reliability paradox in assessing structural relations within covariance structure models. *Educ. Psychol. Meas*., *71*, 306–324. DOI: https://doi.org/10.1177/0013164410384856

Hoyle, R. H. (2012). *Handbook of Structural equation Modelling*. New York, USA: Guilford Press.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55. DOI: https://doi.org/10.1080/10705519909540118

Jia, F., Moore, E. W. G., Kinai, R. Crowe, K. S., Schoemann, A. M., & Little, T. D. (2014). Planned missing data designs with small sample sizes: How small is too small? *Int. J. Behav. Dev*., *38*, 435–452. DOI: https://doi.org/10.1177/0165025414531095

Kaplan, D. (2000). *Structural Equation Modeling – Foundations and Extensions*. Sage Publications. DOI: https://doi.org/10.4135/9781452226576

Kenny, D. A., Kanishan, B. & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, *44*, 486–507. DOI: https://doi.org/10.1177/0049124114543236

Kenny, D. A. & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modelling. *Structural Equation Modeling*, *10*, 333–351. DOI: https://doi.org/10.1177/0049124114543236

Kline, R. B. (2010). *Principles and Practice of Structural Equation Modeling* (3rd ed.). Guilford Press, New York.

Lang, K. M., Moore, E. W. G., & Grandfield, E. M. (2020). A novel item-allocation procedure for the three-form planned missing design. *MethodsX*, *7*, 100941. DOI: https://doi.org/10.1016/j.mex.2020.100941

Lawes, M., Schultze, M., & Eid, M. J. A. (2020). Making the most of your research budget: Efficiency of a three-method measurement design with planned missing data. *Assessment*, *27*(5), 903–920. DOI: https://doi.org/10.1177/1073191118798050

Little, T. D., Jorgensen, T. D., Lang, K. M., & Moore, W. G. (2013). On the Joys of Missing Data. *Journal of Pediatric Psychology*, *39*(2), 151–162. DOI: https://doi.org/10.1093/jpepsy/jst048

McNeish, D., & Wolf, M. G. (2021). Dynamic Fit Index Cutoffs for Confirmatory Factor Analysis Models. *Pshycological Methods*. DOI: https://doi.org/10.1080/00223891.2017.1281286

Moore, E. W. G., Lang, K. M., & Grandfield, E. M. (2020). Maximizing data quality and shortening survey time: Three-form planned missing data survey design. *Psychology of sport & Exercise*, *51*, 1–12. DOI: https://doi.org/10.1016/j.psychsport.2020.101701

Moshagen, M. (2012). The model size effect in SEM: Inflated goodness of fit statistics are due to the size of the covariance matrix. *Struct. Equ. Model. Multidiscip. J.*, *19*, 86–98. DOI: https://doi.org/10.1080/10705511.2012.634724

Nye, C. D. (2022). Reviewer Resources: Confirmatory Factor Analysis. *Organ. Res. Methods*. Online publishing. DOI: https://doi.org/10.1177/10944281221120541

Pornprasertmanit, S., Miller, P., Schoemann, A., & Jorgensen, T. D. (2021). *simsem: SIMulated Structural Equation Modeling (version 0.5-16) [R package]*. Retrieved from http://simsem.org/

Rhemtulla, M., & Hancock, G. R. (2016). Planned Missing Data Designs in Educational Psychology Research. *Educational Psychologist*, 51(3-4), 305–316.

Rioux, C., Lewin, A., Odejimi, O. A., & Little, T. D. (2020). *International Journal of Epidemiology*, 1702–1711. DOI: https://doi.org/10.1093/ije/dyaa042

Rubin, D. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. DOI: https://doi.org/10.1093/biomet/63.3.581

Schoemann, A. M., Miller, P., Pornprasertmanit, S, & Wu, W. (2014). Using Monte Carlo simulations to determine power and sample size for planned missing designs. *International Journal of Behavioural Development*, *38*(5), 471–479. DOI: https://doi.org/10.1177/0165025413515169

Shi, D., Lee, T., & Maydeu-Olivares, A. (2019). Understanding the Model Size Effect on SEM Fit Indices. *Educational and Psychological Measurement*, *79*(2), 310–334. DOI: https://doi.org/10.1177/0013164418783530

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*, 173–180. DOI: https://doi.org/10.1207/s15327906mbr2502_4

Tucker, L. R., & Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*, 1–10. DOI: https://doi.org/10.1007/bf02291170

Vicente, P. C. R. (2023). Evaluating the Effect of Planned Missing Designs in Structural Equation Model Fit Measures. Psych, *5*, 983–995. DOI: https://doi.org/10.3390/psych5030064

Wu, W., & Jia, F. (2021). Applying planned missingness designs to longitudinal panel studies in developmental science: An overview. *Child & Adolescent Development*, 2021 (175), 35–63. DOI: https://doi.org/10.1002/cad.20391

Yuan, K.-H. (2005). Fit indices versus test statistics. *Multivariate Behavioural Research*, *40*(1), 115–148. DOI: https://doi.org/10.1207/s15327906mbr4001_5

Zhang, X., & Savalei, V. (2023). New computations for RMSEA and CFI following FIML and TS estimation with missing data. *Psychol. Methods*, *28*, 263–283. DOI: https://doi.org/10.1080/10705511.2019.1642111

Načrtovane opustitve opazovanj v raziskavi: Ali je to dobra izbira pri uporabi modelov strukturnih enačb?

Izvleček

Manjkajoča opazovanja se lahko pojavijo zaradi napora, potrebnega za odgovarjanje na številna vprašanja v dolgih anketah, in stroškov, ki so potrebni za pridobitev nekaterih odgovorov. Izvajanje načrtovane zasnove manjkajočih opazovanj v anketah pomaga zmanjšati količino vprašanj, na katera mora odgovoriti vsak anketiranec, s čimer se zmanjša utrujenost anketirancev in zmanjšajo stroški izvajanja. Načrt s tremi oblikami in načrt z dvema metodama sta dve različni vrsti načrtovanih manjkajočih opazovanj. Pomemben vidik pri načrtovanju raziskave z načrtovanimi opustitvami je vedeti, kako bo to vplivalo na statistične rezultate. V tem članku je izvedena simulacijska študija, da bi analizirali, kako se običajna merila ustreznosti, kvadratni koren povprečne kvadrirane napake ocen (angl. root mean square error of approximation - RMSEA), standardizirani kvadratni koren povprečja kvadriranih ostankov (angl. standardized root mean square residual - SRMR), primerjalna mera prileganja (angl. comparative fit index - CFI) in Tucker-Lewisov indeks (TLI), obnesejo pri prilagoditvi modela strukturnih enačb. Rezultati so pokazali, da indeksi CFI, TLI in SRMR kažejo občutljivost na izpuste pri majhnih vzorcih, nizkih faktorskih obremenitvah in velikih modelih. Na splošno ta študija prispeva k našemu razumevanju pomena upoštevanja opustitev opazovanj pri načrtovanju tržnih raziskav.

Ključne besede: opustitve opazovanj pri načrtovanju, model strukturnih enačb, anketa