The challenge of modular matrix designs: How much information do we need to get reliable estimates?

Michael Ochsner

Swiss Foundation for Research in Social Sciences (FORS), and ETH Zürich, Switzerland.
E-mail: michael.ochsner@fors.unil.ch

Jessica M. E. Herzing

LINES/FORS, Institute of Social Sciences, University of Lausanne, Switzerland.
E-mail: jessica.herzing@unil.ch

Patricia J. Milbert

Swiss Foundation for Research in Social Sciences (FORS), Switzerland.
E-mail: patricia.milbert@fors.unil.ch

PLEASE DO NOT CITE WITHOUT AUTHOR PERMISSION.

Keywords: multiple imputation, split questionnaire, nonresponse, multiple matrix design, survey length, missing data, response burden

Introduction

In general, survey practitioners try to avoid long web survey questionnaires[1], as lengthy surveys can affect data quality. For example, previous research finds a negative impact of survey length on response rates (see for examples Crawford, Couper, and Lamis 2001; Cook et al., 2000; Galesic and Bosnjak, 2009; Kaplowitz, Lupi, Couper, and Thorp, 2012; Lynn, 2013; Walston et al., 2006), and response quality in web surveys (e.g., breakoff, satisficing, item nonresponse; see Couper, 2008, p. 298; Galesic, 2006; Peytchev and Peytcheva 2017). In addition, lengthy surveys lower respondents' convenience (see Couper, 2013) and respondents' survey engagement (see McCutcheon, 2014). Consequently, lengthy questionnaires can introduce nonresponse and measurement bias. To tackle the issue of lower data quality due to survey length, the convention for the maximum length of web surveys is around 20 minutes (Callegaro, Manfreda, and Vehovar, 2015, p. 102; Couper, 2008, p. 298).

Especially in general population surveys, researchers often conduct surveys that last longer than 20 minutes. Splitting lengthy questionnaires into parts is a possible solution to reduce the length of a web survey questionnaire and, thus, to the threats of nonresponse and measurement bias from low response rates and high response burden. Two types of split questionnaires can be distinguished, a matrix questionnaire design and a modular questionnaire design. In a matrix design, the questionnaire is split into question sets and sample units are randomly assigned to one or multiple question sets, but never to the whole questionnaire. Unlike matrix questionnaire designs, modular questionnaire designs split the source questionnaire into parts. Each questionnaire part is called module, so sample units receive the complete source questionnaire in several modules over time. A third alternative to long surveys is a combination of both the matrix and the modular designs, which we call modular matrix questionnaire design. In a first step, subsets of sample units receive a set of questions from the source questionnaire (matrix design). In a second step, sample units who answered the first set of questions receive a second survey, which covers the set of questions that were missing by design in the first survey (modular design).

Split questionnaire designs may reduce response burden and increase response rates and as a result,

[1] We do not differentiate between perceived and actual survey length.

might increase data quality. However, in split questionnaire designs not all respondents are asked all survey questions and not all respondents answer all survey questions. In conclusion, a diminution of response burden may be observed, but the number of answers per specific survey question is also reduced. In the literature, the creation and the consequences of split questionnaire designs are discussed (e.g., Lynn, 2013; Peytchev and Peytcheva 2017; West et al. 2015), but literature is sparse on the amount of information needed to get reliable results with data generated from surveys with split questionnaire designs. Therefore, this study focuses on the question: How much information is needed to get unbiased estimates from data that was generated from surveys with split questionnaire designs?

In a matrix design, the overall sample size for each survey variable is reduced due to the random assignment of each question set to the sample units. This results in a loss of information and hence, a loss of precision in survey estimates (see Merkouris, 2010). Yet, there is little experimental research, which evaluates the inferences obtained by conducting a full questionnaire versus a split questionnaire. Furthermore, little is known about whether data from a matrix questionnaire design can be enriched when it is combined with a modular questionnaire design and whether the available case method (use of available data) or multiple imputation of the missing data yield better estimates. Therefore, this study evaluates the quality of estimates and their standard errors of a modular matrix design with and without imputation by comparing it with the quality of estimates and their standard errors of a web survey and a f2f survey containing the full questionnaire. For this purpose, we compare the results of a substantive analysis with data from one survey using different survey designs. We apply the same substantive analysis to data collected via a) a face-to-face survey, b) a long web survey, c) a web survey with modular matrix design using only data from the first module, d) a web survey with modular matrix design using data from the first module with multiply imputed data, e) a web survey with modular matrix design using data from module 1 with multiple imputed data using information from module 2, f) a web survey with modular matrix design using data from module 1 and module 2, and g) a web survey with modular matrix design using data from module 1 and module 2 with multiply imputed data.

Study design

Using the Swiss data of the European Values Study (EVS) 2017, we investigate the effects of a modular matrix design on the results of a substantive analysis. In total 8,200 sample units were drawn from a Swiss register random sample (Stichprobenrahmen für Personen- und Haushaltserhebungen, SRPH). Separate samples were drawn for the f2f survey and the web survey. The sample of the f2f survey consisted of 1,400 individuals and the sample of the web survey consisted of 6,800 individuals. The sample units of the web survey were randomly assigned to one of the eight experimental conditions (for an illustration see table 1). A random subdivision was done on 2,000 individuals assigned to the full questionnaire surveys; half got a letter announcing a short survey duration (25 minutes) and half got an invitation announcing a realistic survey duration (45 minutes).

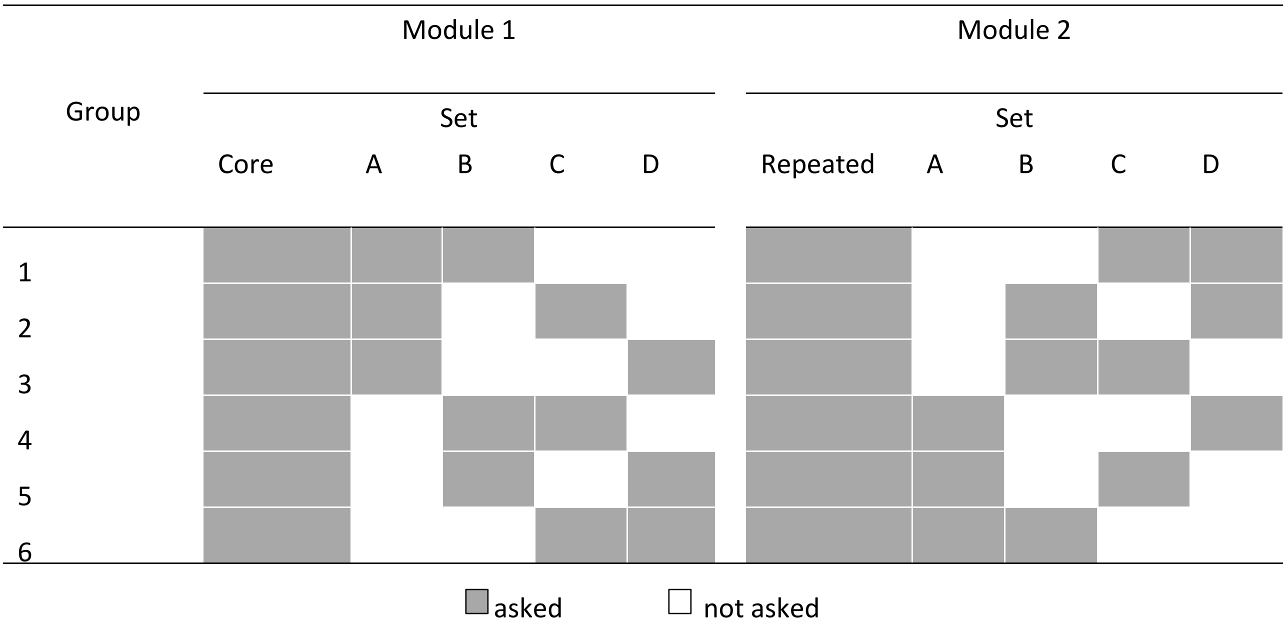
All web surveys are mixed-mode web/paper surveys (push-to-web design). Independent of the experimental condition (f2f or web surveys), all sample units received an unconditional 10 CHF postal voucher. The field duration for all experimental conditions lasted for four and a half months (15th September 2017 to 30th February 2018).

In matrix questionnaire designs, the source questionnaire (here a 60 minutes f2f or web survey) is either split randomly, statistically (i.e. according to the structure of correlations) or thematically into sets of questions. For the matrix design of the web survey, the questionnaire was split thematically into six sets of questions (for arguments in favor of a thematic split see Adigüzel & Wedel, 2008, p. 616), as the correlations found in the previous round of the EVS Switzerland (2008) were low or moderate and few variables had a high degree of correlation. Thus, a loss of information due to a thematic split instead of a statistical split is considered minimal.

Table 1. – Description of the study design and the response rates for each experimental condition.

Survey mode	Sample units	Questionnaire design	Announced survey length	Questionnaire order	Response rate (%)
F2f	1400	full-length	60 minutes	original	52.3%
Web and mail	6 x 800 = 4800	split design module 1 (3 sets)	25 minutes	original	44.5%
		+ module 2 (3 sets)	+ 15 minutes	original	35.1%
	1000	full-length	25 minutes	489 original	42.1%
				491 reversed	47.9%
	1000	full-length	45 minutes	490 original	40.6%
				496 reversed	41.9%

Note.- F2f = Face-to-face questionnaire, % = presented in percent. Source: EVS dataset MDS1.



Note. – Repeated = repeated questions from the core questionnaire of module 1 included questions on gender, importance in life, and personal health.

Figure 1. - Split questionnaire design with a modular matrix questionnaire design. Module 1 consists of the core set and two out of four sets of questions. Module 2 consists of a repeated set and two out of four sets of questions.

Five criteria were applied to split the questionnaire thematically regarding meaningfulness, comparability between experiments, avoidance of context effects, variables often analyzed together, and variables which improve the analytical power (see for examples Raghunathan & Grizzle, 1995; Rässler, Koller, & Mäenpää, 2002). Furthermore, the six sets of questions were split in two modules to reduce overall survey length (see figure 1 for an illustration), based on three pragmatic criteria. As a result, of these pragmatic and thematic criteria the source questionnaire was split into six parts consisting of a core set of questions, a repeated set of questions, and four questionnaire sets with different topics (for an illustration see figure 1). All possible combinations of two sets were randomly assigned to sample units in module 1. In module 2 sample units received the missing two sets of questions. Thus, by completing module 1 the respondents answered about 60 percent of the source questionnaire. By completing module 2, respondents answered the remaining 40 percent of the source questionnaire. The order of the questions is the same as in the source questionnaire (the missing items are skipped).

Analytic approach

Typically, survey researchers assume that information for all questions (or items) is collected from all respondents (see Chipperfield and Steel 2009). When using a matrix questionnaire designs, this assumption needs to be relaxed, as information on specific sets of questions is collected for subsets of respondents. However, in modular matrix designs, it is possible to have full information from respondents when respondents answer both module 1 and module 2 (equivalent to the full-length questionnaire). In both scenarios, one can deal with nonresponse by either using the available case method or an imputation method (for examples see Raghunathan and Grizzle 1995; Merkouris, 2010). Yet, it remains unclear which is the best approach for the handling of missing data in split questionnaire designs.

In the following, we evaluate the strategies to handle nonresponse –available cases method and multiple imputation–based on a model proposed by a substantive study on how church attendance is associated with attitudes towards homosexuality by Halman and van Ingen (2015). We use the European Value Study (EVS) 2017 data for Switzerland. The question on attitudes towards homosexuality was asked in the questionnaire set on religion (set B). Question set B occurs for respondent group 1, 4 and 5 in module 1 and respondent group 2, 3, and 6 in module 2 and hence, systematic missings occur in the split questionnaire scenarios (see figure 1). Furthermore, we include independent variables from the core questionnaire on respondents' behavior (church attendance, religious denomination) and socio-demographic characteristics (education, age, gender). Consequently, we have systematic missings on our dependent variable on justifying homosexuality and on two of the independent variables, i.e. importance of god in daily life and being a religious person (systematic unit nonresponse and non-systematic item nonresponse). Additionally, we have non-systematic missings in our independent variables educational level, age, gender, religious denomination, and church attendance (non-systematic unit and item nonresponse).

Regarding imputation methods for the split questionnaire designs, we imputed missing values using chained equations (see Kaplan and Su, 2016; Johnson and Young, 2011; White et al, 2010). In addition, as recommended by previous research, we include (at least) the same variables that we later use in our estimated model (von Hipple 2013, White et al., 2010). We only included the town size, a variable that has no missing values, as an auxiliary variable in the imputation model. We decided on the number of imputations for each experimental condition based on the approach proposed by Hippel (2018).

For the source questionnaire conditions (long web surveys and f2f survey), we imputed the missing values with the same procedure as in the modular matrix designs. However, we yet did not correct for potential interviewer effects in case of the f2f survey.

The modules and sets of questions were randomly assigned to sample units and the missing patterns do suggest that the data is missing at random (MAR). Hence, multiple imputation can produce unbiased estimates in our data set (Allison, 2002; Enders, 2010; Little & Rubin, 2002; Schafer & Graham, 2002).

Preliminary results

In the following analysis, scenario e) that was described in the introduction is not presented yet. The labels of the coefficient plots presented in Figure 2 and 3 are equivalent to the scenarios described in the introduction. We see in Figure 2 and 3 that the estimations of the f2f survey (A) and the long web survey (B) differ from each other. The effects of educational level and being an atheist point to different directs in case of the f2f survey and the long web survey. Furthermore, we see that the imputed data (D, G) compared to the complete case data (C, F) does not result in different estimates. In addition, the results presented show that the estimates of the mixed modular designs (F,G) and the matrix designs (C, D) have no clear pattern in their similarity to one of the surveys where the source questionnaire was used (A,B).

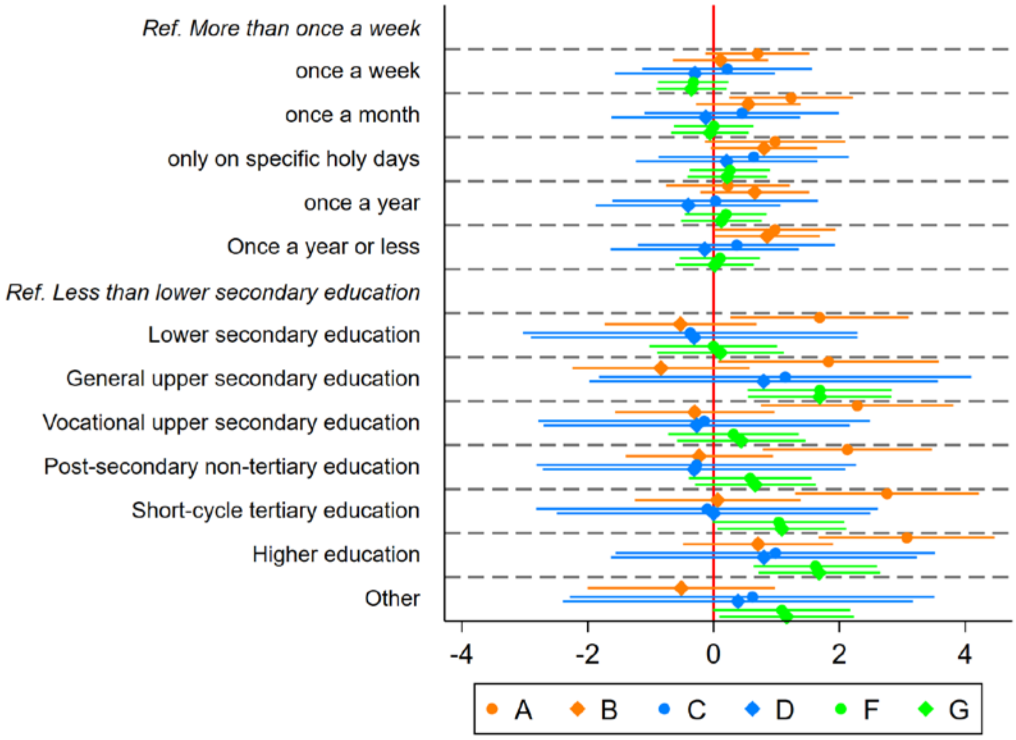


Figure 2. – Effects of church attendance and educational level on justification of homosexuality by A) f2f survey, B) long web survey, C) module 1, complete cases, D) module 1 with imputation, F) module 1 +2, complete cases, G) module 1 + 2 with imputation. Plot markers are coefficients and horizontal spikes are 95% confidence intervals.

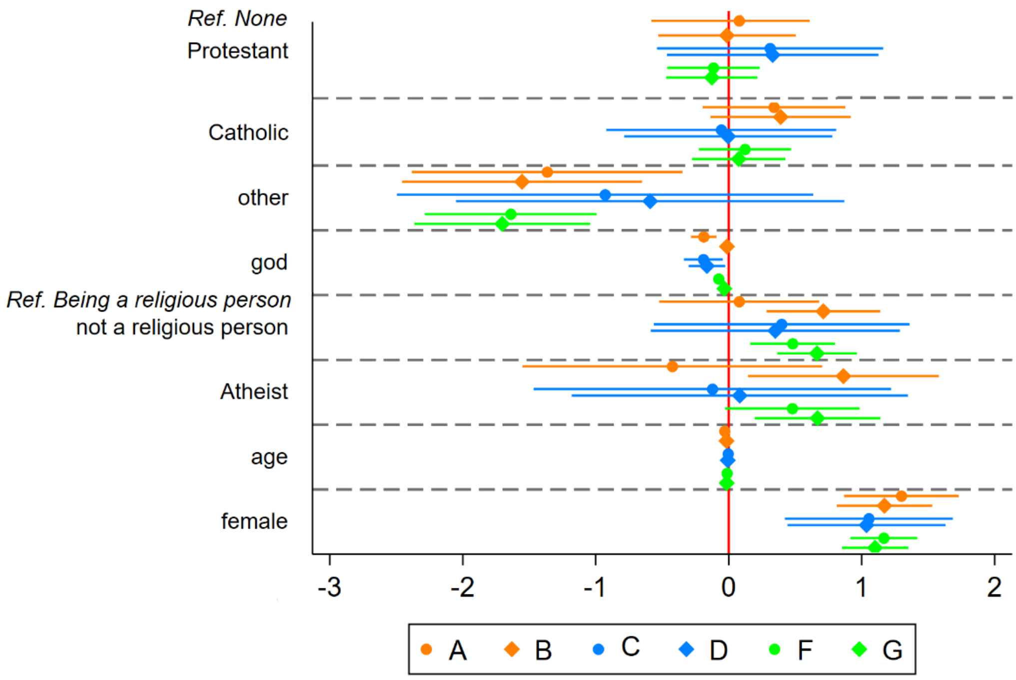


Figure 3. – Effects of denomination, importance of god, self-rated religiosity, age, and gender on justification of homosexuality by A) f2f survey, B) long web survey, C) module 1, complete cases, D) module 1 with imputation, F) module 1 +2, complete cases, G) module 1 + 2 with imputation. Plot markers are coefficients and horizontal spikes are 95% confidence intervals.

Conclusion and points of discussion

Intuitively, we expect larger differences in estimates between the complete case analysis and the imputed data. We were astonished to find these large differences between the f2f survey (A) and the long web survey (B). Interestingly, we did not find that more information (mixed modular design) result in similar estimates between the full surveys (A, B), the matrix designs (C, D), and the modular matrix designs (F, G).

We would appreciate a discussion on the selection of potential useful auxiliary variables for the imputation model based on theoretical (e.g., media consumption, political interest) or statistical arguments (e.g., correlations). In addition, it would be great to learn from other experience on how to handle potential interviewer effects for the f2f survey (e.g., clustered standard errors), as this should be used as one benchmark for the modular matrix designs. Furthermore, we have the feeling that this example (investigating effects on justifying homosexuality) might not be the best choice as an exemplary case and hence, it would be interesting to hear how others have decided on models to exemplify imputation techniques. Finally, we plan to consider different estimation scenarios, such as differences in the skewness of the dependent variable, a dependent variable based on factor analysis, whether the dependent variable has systematic missings or not, whether the independent variables have systematic missings or not and so forth.

References

Adigüzel, F. & Wedel, M. (2008). Split questionnaire design for massive data. *Journal of Marketing Research, 45*, 608-617.

Allison, P. D. (2002). Missing data. Thousand Oaks, CA: Sage Publications.

Bianchi, A., Biffignandi, S., & Lynn, P. (2017). Web-f2f mixed-mode design in a longitudinal survey: Effects on participation rates, sample composition, and costs. *Journal of Official Statistics*, 33(2), 385-408.

Callegaro, M., Manfreda, K. L., & Vehovar, V. (2015). *Web survey methodology*. London, UK: Sage.

Chipperfield, J. O. & Steel, D. G. (2009). Design and estimation for split questionnaire surveys. *Journal of Official Statistics*, 25, 227-244.

Converse, P. D., Wolfe, E. W., Huang, X., & Oswald, F. L. (2017). Response rates for mixed-mode surveys using mail and e-mail/web. *American Journal of Evaluation*, 29, 99 – 107.

Couper, M. P. (2013). Is the sky falling? New technology, changing media, and the future of surveys. *Survey Research Methods,* 7, 145–156.

Couper, M. P. (2008). *Designing effective web surveys*. New York, NJ: Cambridge University Press.

Cook, C.; Heath, F. & Thompson, R.L. (2000). A meta-analysis of response rates in web- or internet-based surveys. *Educational and Psychological Measurement*, 60, 821-836.

Crawford, S. D., Couper, M. P., & Lamias, M. J. (2001). Web surveys: Perception of burden. *Social Science Computer Review*, 19, 146-162.

de Leeuw, D. (2005). To mix or not to mix data collection modes in surveys*. Journal of Official Statistics*, 21, 233-255.

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

Galesic, M. (2006). Dropouts on the web: Effects of interest and burden experienced during an online survey. *Journal of Official Statistics*, 22, 313-328.

Galesic, M. & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly*, 73, 349-360.

Halman, L. & van Ingen E. (2015). Secularization and changing moral views: European trends in church attendance and views on homosexuality, divorce, abortion, and euthanasia. *European Sociological Review*, 31(5), 616-627.

Hippel, P. T. (2018). How many imputations do you need? A two-stage calculation using a quadratic rule. *Sociological Methods & Research*, online first.

Johnson, D. R. & Young, R. (2011). Toward best practices in analyzing datasets with missing data: comparisons and recommendations. *Journal of Marriage and Family*, 73(5), 926-945.

Kaplan, D. & Su, D. (2016). On matrix sampling and imputation of context questionnaires with implications for the generation of plausible values in large-scale assessments. *Journal of Educational and Behavioral Statistics,*41(1), 57-80.

Kaplowitz, M. D., Lupi, F., Couper, M. P., & Thorp, L. (2012). The effect of invitation design on web survey response rates. *Social Science Computer Review*, 30, 339-349.

Klausch, T., Schouten, B., & Hox, J. J. (2017). Evaluating bias of sequential mixed-mode designs against benchmark surveys. *Sociological Methods & Research*, 46, 456-489.

Little, R. J. A. & Rubin, D. B. (2002). Statistical analysis with missing data. New York, NJ: John Wiley.

Lynn, P. (2013). Alternative sequential mixed-mode designs: effects on attrition rates, attrition bias, and costs. *Journal of Survey Statistics and Methodology*, 1, 183-205.

McCutcheon, A. L. (2014, May). *Web surveys, online panels, and paradata: automating responsive design*. Census Bureau: NCRN Meeting Spring 2014. Retrieved June, 26th, 2018 from https://ecommons.cornell.edu/bitstream/handle/1813/36398/NCRN-Spring2014-SessionIII-McCutcheon.pdf;jsessionid=816ADE31BC73D036BAFCF962196F660A?sequence=2

Merkouris, T. (2010). An estimation method for matrix survey sampling. *Proceedings of the Section on Survey Research Methods: American Statistical Association, July*, *31*, 4880-4886.

Millar, M. M. & Dillman, D. A. (2011); Improving response to web and mixed-mode surveys. *Public Opinion Quarterly*, *75*, 249–269.

Peytchev, A., & Peytcheva, E. (2017). Reduction of measurement error due to survey length: Evaluation of the split questionnaire design approach. *Survey Research Methods*, 11, 361-368.

Walston, J.T., Lissitz, R.W., & Rudner, L.M. (2006), The influence of web-based questionnaire presentation variations on survey cooperation and perceptions of survey quality. *Journal of Official Statistics*, *22*, 271-29.

Raghunathan, T. E., & Grizzle, J. E. (1995). A split questionnaire survey design, *Journal of the American Statistical Association*, 90(429), 54-63.

Rässler, Susanne, Koller, F., & Mäenpää, C. (2002). A split questionnaire survey design applied to German media and consumer surveys, Diskussionspapiere, Friedrich-Alexander-Universität Erlangen-Nürnberg, Lehrstuhl für Statistik und Ökonometrie, No. 42b/2002, Universität Erlangen-Nürnberg, Lehrstuhl für Statistik und empirische Wirtschaftsforschung, Nürnberg. Retrieved June, 27th, 2018 from https://www.econstor.eu/handle/10419/29618

Roberts, C., & Vandenplas, C. (2017). Estimating components of mean squared error to evaluate the benefits of mixing data collection modes. *Journal of Official Statist*ics, 33(2), 303-334.

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

Von Hipple P. & Lynch, L. (2013). Efficiency gains from using auxiliary variables in imputations. Cornell University Library, arXiv e-print 1311.5249.

West, B. T., Ghimire, D., & Axinn, W. G. (2015). Evaluating a Modular Design Approach to Collecting Survey Data Using Text Messages. *Survey Research Methods*, 9(2), 111–123.

White, I. R., Daniel, R., & Royston P. (2010). Avoiding bias due to perfect prediction in multiple imputation of incomplete categorical variables. *Computational Statistics & Data Analysis*, 54, 2267-2275.