Methods for treating bias deriving from different sources of nonsampling errors in mixed mode social surveys

Authors:
Claudia De Vitiis, Alessio Guandalini, Francesca Inglese, Marco D. Terribili

Keywords: mixed mode, selection effect, measurement error, multiple imputation, propensity score

1. Introduction

The mixed mode (MM), i.e. the use of different collection techniques in the same survey, is a relatively new approach that ISTAT, as well as other NSI, is adopting especially for social surveys. Its use is expected to spread both to contrast declining response rates and to reduce the total cost of the surveys. The use of different data collection techniques, in fact, helps in contacting different types of respondents in the most suitable way for each of them.

The use of mixed techniques provide, therefore, a gain both in population coverage and response rate. However, it introduces other issues that must be addressed at different levels: in the design phase by defining the best collection instruments to contain the measurement error and an appropriate sampling design; in the estimation phase by assessing and treating the bias effects due to the introduction of MM, in order to ensure the accuracy of the estimates. The surveys based on MM must be designed, in fact, keeping in mind the constraint that the produced estimates must be consistent and comparable with the analogue ones obtained in the previous survey editions, for ensuring that changes in the time series are exclusively due to real changes of the observed phenomenon.

Mode effect refers strictly to measurement error differences due to the mode of survey administration. When the mixed modes are assigned not randomly, a *selection effect* can generally occur and appropriate inference methods to evaluate mode effect are necessary. The two types of error, selection effect and measurement differences across modes, are confounded in mixed mode surveys especially when modes are administrated in sequential way.

The application of sampling and estimate methods for mixed mode surveys must be oriented by the type of survey techniques adopted and by the fact that they are used concurrently or sequentially:

- When the mixed techniques are used in a sequential way, the application of methods to disentangle the measurement error from the error due to the self-selection of the sample is advisable. If the mix mode is a sequential CATI/CAPI, it is likely that the measurement error is usually quite mild; but, when the CAWI technique is used jointly with techniques using the intermediary of an interviewer, the measurement error can be not negligible because the CAWI is always self-administered.
- Surveys that use mixed mode concurrently but not randomly, usually assign modes according to contact characteristics of the population units. In this case it is necessary to estimate the measurement error but also the selection effect caused by the dissimilar distribution of the populations from which the samples are selected.

In ISTAT several situations have occurred so far.

- Mixed mode used primarily to address coverage issues of previously only CATI surveys (“University and High school graduates' vocational integration” surveys, sequential CAWI/CATI); “Victimization survey”, concurrent CATI/CAPI).
- Mixed mode used primarily to reduce survey cost whereas expanding population coverage, through the introduction of CAWI technique in traditionally PAPI or CATI surveys. (“Multipurpose survey on households: Citizens and leisure time” - 2015”, CAWI/PAPI); “Multipurpose survey on households: Aspects of daily life - 2016” (sequential CAWI/CATI with a control mono mode sample PAPI); “Private copies survey” (concurrent CAWI/CATI).

For estimating the selection and measurement effects, for a global evaluation of data quality, the use of appropriate models is required. For this purpose the availability of *mode insensitive* auxiliary information, obtainable from registers or administrative archives, is a crucial issue.

Another aspect to be considered is that the inference process in mixed mode surveys is more cumbersome due to other sources of non-sampling error, such as total nonresponse which has to be treated together with the specific effects introduced by MM.

The confounding of selection and measurement effects is a central theme of the causal inference literature (Morgan and Winship, 2009) that offers two distinct covariate adjustment models for disentangling confounded effects and for obtaining unbiased estimates of population parameters. This models requires covariates that capture, alternatively, selection effect or measurement effect (Vannieuwenhuyze *et al*., 2014). Important information can derive also from the data collection phase (*paradata*, such as interviewer observations and call record data, or survey questions asking for mode preference) (Vandenplas *et al*. 2016).

The focus of the present work is the illustration of the experimentation plan for the treatment of mode effect in the CAWI/PAPI "Citizen and leisure time" (CLT) survey, together with the description of the state of play and the discussion of the open issues. Through the linkage of survey data with administrative data we exploit the auxiliary variables to define mixed mode models.

This is the first time ISTAT addresses this complex issue facing all the related theoretical implications. In this context, moreover, a project has been just launched for the editing of a methodological volume of which a brief summary and the general provisional index is outlined in the annex.

The paper is organized as follows: in section 2 the theoretical framework of the MM evaluation is outlined; section 3 describes the context of CLT survey, while section 4 illustrates the experimental phase and section 5 underlines some further aspects to be considered in the future.

2. Theoretical framework

The aim of mixed mode (MM) surveys is to combine the data collected with different techniques in a single data set. In this context, the accuracy of inference is ensured only if the assumption of equivalence of measurement across modes holds (Hox *et al.*, 2015). Mixed mode designs introduce, in fact, a data integrity problem.

Mode effects on measurement can be distinguished in two different categories, selection effect, named also mode choice, and measurement effect: the former, due to nonrandom mode assignment, causes only a shift in the response distribution, resulting in a difference in the mean of variables between survey modes; the latter, relating on the question answer process, can affect the correlation structure among variables (Hox *et al.*, 2015). Selection effects and measurement effects should both be investigated on the sample of respondents resulting from the survey modes.

Traditionally, in combining data from multiple response modes, statistical inference methods assume that the measurement effects are ignorable. Ignorability of measurement effects cannot be evaluate through standard statistical tests for the differences in parameter estimates, based on the data collected through different modes, without accounting for mode choice.

Disentangling selection and measurement effects requires auxiliary information from registers or collected by the survey itself. Almost all methods to distinguish these effects depend on covariates that are assumed to be mode insensitive and to fully explain selection effects.

The problem of the selection effect can be faced by referring to Propensity Score (PS) approach (Rosenbaum and Rubin, 1983; Rubin, 2006). With adjustments based on PS, the confounding effects of the selection mechanism are mitigated. Propensity score approach is adopted in observational studies by achieving a balance of covariates between comparison groups. In MM surveys propensity score can be interpreted as the probability of mode assignment conditional on observed covariates. Different PS adjustment methods can be used: matching on propensity score, subclassification (stratification) on propensity score, covariate adjustments using the propensity score (D'Agostino, 1998; Austin, 2011; Lee and Valliant, 2008). The limits of this approach are the mode selection ignorability assumption, the risk of finding unbalanced groups and the violation of the overlap assumption (Vandenplas *et al*., 2016).

Alternative inference methods, that take into account the mode choice in measurement effect adjustment, are based on multiple imputations (Rubin, 1987). Under a multiple imputation approach, mode effect is conceptualized as a missing data problem. Following this approach, mode specific data are imputed, taking mode choice under control, to obtain the complete sample under each alternative mode. These mode-specific data are used to estimate mode specific population means. In presence of measurement effects, the mode specific data estimates may then be adjusted to an internal standard, the best mode, or to an external standard if available. In controlling the mode choice it is possible to define different imputation models considering ignorable and nonignorable mode choice assumption (Park *et al*, 2016, Suzer-Gurtekin, *et al*., 2012).).

Additionally, in multimode data collection, it is necessary to consider the response process that generates a further selection effect due to total nonresponse. In order to assess the total nonsampling error, effects introduced by MM should be treated together with nonresponse selection effect. The mixed mode models and the response model are the basic elements to combine data from mixed mode surveys with different designs. In concurrent MM design, response model can be inserted into the mixed mode models more easily, while in sequential MM design this operation is more complicated. Moreover, in the latter design, the distribution of sampling units to subsequent modes depends on the results of the response process in the preceding modes. In this case, the sequence of modes (mode response path) should be considered in the response model (Cobben *et al*., 2006).

3. The context of the experimentation

3.1. The CLT survey

The sample survey "Citizens and leisure time" collects information about recreational and cultural activities carried out by citizens in their free time, such as sports, reading, cinema, music, the Internet, the recreational activities, social relations and other important issues for the quality of life of people. The survey, conducted for the first time in 1995, is in its fourth year and is part of an integrated system of social surveys (surveys multipurpose household survey).

The survey involves a selected sample of about 24.000 households (of which a set of about 18.000 respondent households are interviewed, 38.000 individuals), spread in nearly 850 Italian municipalities of different demographic size through a two stage sample design. The sample of households was selected from the centralized municipal register.

The information is collected with a mixed technique: an online questionnaire that can be self-compiled by respondents (CAWI technique) or, alternatively, direct interviews with a questionnaire on paper, administered by an interviewer (PAPI technique). To fill in the online questionnaire, sample households use the credentials given in the inviting letter sent by ISTAT. If the family did not complete the questionnaire over the Internet, at the end of the period provided for online filing, a municipal interviewer address personally the same questionnaire to all its members.

The overall response rate was 68.3%, where CAWI was chosen by 20.9% and PAPI by 79.1% of them.

3.2 Results of the linkage to administrative DB

After the survey, in order to obtain external auxiliary information, not affected by mode effect, to be used for the analysis and treatment of mixed mode effect (and nonresponse), administrative data were linked to the data set of selected sample individuals (respondent and nonrespondent). The linkage was performed through the fiscal code, a code identifying univocally each person living in Italy. The administrative data base is the DB of the Archimede Project, (Integrated archive of economic and demographic micro data, Garofalo, 2014) built for expanding ISTAT information provided by administrative archives to producing longitudinal paths and cross-sectional collections of micro data to be made available to different users This objective is achieved through the exploitation of administrative database information contents integrated into ISTAT platform SIM (Integrated Micro data System).

In the following table the numerical results of the linkage and the response phase through the choice of the collection technique are reported. These numbers show that the CAWI mode is still chosen by a limited number of respondents, while the linkage gave in general quite a good result (95,4% of linked units), uniformly distributed between CAWI and PAPI, but not between respondent and not respondent. This last evidence can be explained in part by the fact that the majority of not linked respondents are new members of the *de facto* families found at the moment of the interview, not previously identifiable in the selection register.

Table 1. Response, linkage and selection rates for CLT survey after the linkage to Archimede DB

RESPONSE AND MODE CHOICE	LINKAGE				Total	
	LINKED		NOT LINKED			
		%		%		%
NOT RESPONDENT	18.209	32,7%	316	11,8%	18.525	31,7%
RESPONDENT	37.495	67,3%	2.359	88,2%	39.854	68,3%
CAWI	*7.862*	*21,0%*	*464*	*19,7%*	*8.326*	*20,9%*
PAPI	*29.633*	*79,0%*	*1.895*	*80,3%*	*31.528*	*79,1%*
Total	55.704	100%	2.675	100%	58.379	

The auxiliary variables considered are some demographic (territory, sex, age, citizenship and marital status) and other social economic (educational level, income from tax files, occupational status and type of occupation). Some of those auxiliary variables present missing values in the administrative DB, especially the educational level and the occupational status. In the following scheme the general informative situation after the linkage is summarized: with Y the generic vector of survey variables is named and with X_k the generic auxiliary variable deriving from administrative DB, k=1,…,P. The black dots show data presence while the white ones absence.

Figure 1. The informative situation for CLT survey after the linkage to Archimede DB

Linkage	**Response**	**Mode**	**Y**	**X_1**	**…**	**X_p**
Linked	Respondent	CAWI	●	○ / ●	…	○ / ●
Linked	Respondent	PAPI	●	● / ○	…	● / ○
Linked	Nonrespondent		○	○ / ●	…	○ / ●
Not linked	Respondent	CAWI	●	○		
Not linked	Respondent	PAPI	●	○		
Not linked	Nonresponse		○	○		

From the situation represented in the scheme two main general issues emerge: a set of sample units were not found in the DB and missing data are reported for some X_k in the administrative DB for the set of linked units.

Both these issues are originated by "process reasons" which we can expect would develop well in the future: efforts can be done to link more units through a thorough search of the necessary linking codes, while the quality of administrative DB will improve in the future with more accurate procedures for the definition of variables.

4. The mode effect treatment for the CLT survey

4.1 The methods

For the treatment of mode effect in the CAWI/PAPI "Citizen and leisure time" survey, we are experimenting different methods:

- the Propensity Score Subclassification method (Rosenbaum and Rubin, 1983; Austin, 2011; Vandenplas, 2016)) and the Standard Multiple Imputation (Kolenicov and Kennedy, 2014), both assuming mode selection ignorability;
- the Parametric Fractional Imputation (PFI) method (Kim, 2011; Park *et al*., 2016) that considers also nonignorability of mode selection.

The multiple imputation approach proposed by Suzer-Gurtekin, Heeringa and Valliant (2012) is under study. This approach uses multiple imputation selection model (Heckman, 1979) to take into account nonignorability of mode selection.

For the implementation of methods based on multiple imputations, the PAPI mode is taken as a reference survey mode, because it was the technique used for all previous occasions of the survey. Therefore the imputation produces counterfactual values for the survey variable y as if the CAWI respondent had responded with the PAPI technique (Vannieuwenhuyze *et al*., 2014).

Given the operational context, the choice of the methods to be implemented for the analysis and treatment of mode-effects has to deal with several issues:

- missing values on covariates, so that method for incomplete data analysis should be considered (Ibrahim *et al*., 2005);
- evaluation of the non ignorability hypothesis (MAR assumption);
- the total nonresponse which results downstream of the selection process;

For the moment, assuming the MAR hypothesis, we concentrate our analysis to the set of respondent with a complete set of a reduced number of covariates X. Besides we leave the total nonresponse issue to a further phase of the study.

It is well known that the Complete-Case (CC) analysis can be biased when the data are not missing completely at random (MCAR). For handling missing-data problems Available-Case analysis (AC) can be used considering methods (Ibrahim et al. 2005) that incorporate the incomplete cases in the analysis both in the case of ignorable (MAR) and nonignorable missing covariates (NMAR).

4.2 The models

We focus first the attention on the Parametric Fractional Imputation (PFI) method which addresses in a comprehensive way all the components of the mode effect evaluation. Following the notation in Park *et al*. (2016), we started to study the structural model $f(y|\mathbf{X})$, the measurement error model $g(y_{\mathrm{CAWI}}|y)$ and the choice model $P(M = \mathrm{CAWI}|\mathbf{X}, y)$ for the CLT survey.

We consider two dichotomous survey variables: “internet access” (1=yes, 0=no) and “sport activity in the last month” (named “sport activity”, 1=yes, 0=no). In the following of our study we will extend the analysis also including polytomous variables.

The structural models f for the two independent variables y (internet access and sport activity) is a binomial logistic model at individual level and it is defined as follow:

Model 1	y ~ region + sex + age class + marital status + education + citizenship + occupation type + income class + income source

The measurement model g on y (Model 2) is a linear model and is still defined at individual level. In the model g, the covariate is the predicted value estimated from the structural model f.

The choice model is defined at household level. As in CLT the survey units are the household, the choice of the survey mode depends on household although, obviously, is the result of the combination of individual features. Therefore, the choice model $P(M = \mathrm{CAWI}|\mathbf{X}, y)$ for the case of nonignorability is defined as follow:

Model 3	*Survey mode* ~ region + # components + household income + # components by sex + # components by age groups + # components by marital status + # components by education + # income earners + # components by income sources + y

This expression can be directly used when y is an household variable but needs to be reformulated for individual variables.

When considering ignorability, the choice model reduces to $P(M = \text{CAWI}|\text{X})$, then the survey variable y can be removed from the Model 3.

4.2 The phase of the methods implementation

The structural Model 1 and the measurement Model 2 are used together with choice Model 3 to define the imputation model $f(y_{PAPI}|\mathbf{X}, y_{CAWI})$ for the PFI method. The imputation model, obtained through the Bayes formula, predicts imputed values (unobserved response PAPI) for the CAWI respondents. The parameters of the models and the predicted values are estimated via an iterative algorithm. In each iteration several imputed values, with their fractional weights, are created for each CAWI respondent. Each fractional weight is computed as the expected value of the imputation model.

Standard Multiple Imputation must be adapted to the context of mixed mode CLT survey, as described in Kolenicov and Kennedy (2014).

For both methods based on multiple imputation, the purpose is deriving an unbiased estimate of the interest parameter using the conditional expectation of the imputed values for the CAWI respondents.

The purpose of the implementation of the Propensity Score Subclassification method, conversely, is to correct the selection effect through the calculus of weights which adjust the distribution of the CAWI and PAPI respondents. This approach implies:

- an estimate of the propensity score model (Model 3) parameters;
- the definition of subclassification (strata) of CAWI and PAPI respondents based on propensity score;
- the validation of the balancing assumption;
- for each balanced subclassification, the calculus of weighs that equate the weighted proportion of CAWI respondents with the proportion of PAPI respondents in the same stratum.

5. Further aspect

A relevant aspect that our study will have to consider concerns variance. Any adjustment for mode effects has, in fact, its own level of uncertainty. When a mode effect adjustment is incorporated into an estimate, the standard error of that estimate increases so, even if there is evidence that the adjustment makes the survey estimate more accurate, it may not necessarily reduce the mean square error of the estimate if the standard error is greatly inflated.

Moreover, the study phase of mode effect can be heavily time and resource consuming as all of the statistical modeling and analysis takes time, lengthening the duration between the end of the field period and the release of the survey findings. For this and the previous reason, it is a crucial issue in the planning of the survey design to prevent the measurement effect as much as possible, so as to reduce the need for adjustments or limit them to the selection effect treatment.

6. Issues for the discussion:

- Restricting to our survey and data analysis context, in treating the mode effect:
  1) how to deal with incomplete auxiliary information: what are the implications of an imputation? can we insert the Available-Case analysis (AC) in the multiple imputation approach?
  2) how to deal with total nonresponse in sequential MM? how can we combine the response model with the MM model?
- In general, when planning a MM survey:
  3) what actions is it possible to take to make the propensity score approach useful to derive weight adjustments?
  4) which particular practice is possible to adopt to design more efficient samples that assign modes according to contact characteristics of the population units?

References

Austin P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46:399-424, 2011. Taylor & Francis Group, LLC.

Cobben F., Schouten B. and Bethlehem J. (2006). A model for statistical inference based on mixed mode interviewing. Proceedings of Q2006 – European conference on Quality in Survey Statistics.

D'Agostino, R. B. (1998). Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Statistics Medicine. Volume 17, Issue 19, 15 October 1998, Pages 2265-2281

Garofalo, G. (2014). Il Progetto ARCHIMEDE obiettivi e risultati sperimentali. Istat Working Paper (2014).

Heckman, J. J. (1979). Sample selection bias as specification error. Econometrica, 47: 153-161.

Hox, J., De Leeuw, E.D., Klausch, T. (2015). Mixed Mode Research: Issues in Design and Analysis. International Conference on total survey error: improving quality in the era of big-data, Baltimora, September 2015.

Ibrahim, J. G., Chen, MH., Lipsitz, S. R. and Herring, A. H. (2005). Missing data methods for generalized linear models: a comparative review. Journal of the American Statistical Association. March 2015, Vol. 100, No. 469.

Kim, J. K. (2011). Parametric fractional imputation for missing data analysis. Biometrika 98 119– 132. MR2804214

Kolenicov, S. and Kennedy, C. (2014). Evaluating three approaches to statistically adjust for mode effects. Journal of Survey Statistics and Methodology 2 126–158.

Lee, S., and Valliant, R. (2008). "Post-Survey Weighting Methods Using Propensity Scores: A Review," chapter 8 in Advances in Telephone Survey Methodology, John Wiley.

Morgan S. L. and Winship C. (2009). Counterfactuals and Causal Inference: Methods and Principles for Social Research. New York: Cambridge University Press, 2007, 319 pages.

Rosenbaum, P. R. and Rubin, D. B. (1983) The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika, Vol. 70, No. 1. (Apr., 1983), pp. 41-55.

Rubin, D.B. (1987), Multiple Imputation for Nonresponse in Surveys, New York: John Wiley & Sons, Inc.

Rubin D. B. (2006). Matching sampling for causal effects. New York, NY: Cambridge University Press.

Park, S., Kim, J. K., and Park, S. (2016). An imputation approach for handling mixed-mode surveys. The annals of Applied Statistics. Vol. 10, No. 2. 1063-1085. 2016.

Suzer-Gurtekin, Z. T., Heeringa, S. G. and Valliant, R. (2012). Investigating the bias of alternative statistical inference methods in sequential mixed-mode surveys. Section on Survey Research Methods – JSM 2012.

Vandenplas, C., Loosveldt, G., and Vannieuwenhuyze, J. T. A. (2016). Assessing the use of mode preference as a covariate for the estimation of measurement effects between modes. A sequential mixed mode experiment. Method, data, Analyses. Vol. 10(2), 2016, pp. 119-142.

Vannieuwenhuyze J. T. A., Loosveldt G. and Molenberghs G. (2014). Evaluating mode effects in mixed mode survey data using covariate adjustment models. Journal of Official Statistics, Vol. 30, No. 1, 2014, pp. 1-21.

ANNEX – The project for the volume "Mixed-mode effects in ISTAT social surveys using the web"

Orietta Luzi

Summary

In the context of social surveys, the increasing internet penetration in the Italian population is encouraging ISTAT to extend the use of self-administered Web Interviews (CAWI) in current statistical processes, in combination with other data collection modes (e.g. CAPI, CATI), in order to ensure better trade-off between costs and total survey errors (and specifically nonsampling errors - coverage error, nonresponse error and measurement error) with respect to single mode surveys.

However, it is well known from the plentiful literature on this subject that multi-mode designs may introduce other forms of bias. In particular, data collected through different modes may have quality problems due to (a confounding of) selection effects and measurement effects (measurement error) caused by mode differences.

For these reasons, survey researchers have to decide if a mixed mode approach is appropriate for their specific survey context, and how to efficiently design their data collection strategy in order to minimize mode effects on final results and to fully take advantage from the specific features of the different techniques.

This publication is intended to support ISTAT researchers in these activities, focusing on mixed-mode strategies involving the use of the CAWI technique. To this aim, an overview of concepts and main issues in mixed mode data collection strategies is provided. A broad discussion is devoted to the prevention of the biasing effects potentially associated to multimode data collection, e.g. those related to the questionnaire design, the field organization, the sample design. Methodological solutions for the diagnosis and treatment of mode effect(s) are also presented, with experimental applications of some approaches to selected social survey data. The preliminary Index of the publication is reported below.

This volume is intended as the first of a series of publications where the main results of research and applications in the area of mixed mode data collection will be periodically described.

General Index

1 Introduction

2 Mixed mode data collection

- 2.1 Definitions
- 2.2 Mixed mode effects
- 2.3 Pilot surveys and tests

3 Methods to prevent measurement effect

- 3.1 Survey design
- 3.2 Sample design
- 3.3 Questionnaire design

4 Methods for diagnosis and treatment of mode effect

- 4.1 General inferential framework
- 4.2 Methods for the evaluation and treatment of selection effect
- 4.3 Methods for the treatment of measurement error
- 4.4 Treatment of nonresponse in mixed mode surveys

5 Explorative analyses

6 Case studies

7 Conclusions and future research

Reference