Assessing and adjusting bias deriving from mode effect in mixed mode social surveys

Claudia De Vitiis, ISTAT, devitiis@istat.it
Alessio Guandalini, ISTAT, alessio.guandalini@istat.it
Francesca Inglese, ISTAT, fringles@istat.it
Marco D. Terribili, ISTAT, terribili@istat.it

Abstract

The mixed mode (MM), i.e. the use of different collection techniques in one survey, is a relatively new approach for ISTAT, especially for social surveys. It is adopted both to contrast declining response and coverage rates and to reduce the cost of the surveys. Nevertheless, mixed mode introduces several issues that must be addressed both at the design phase and at the estimation phase, by assessing and treating the bias effects (mode effect) due to the use of MM, in order to ensure the accuracy of the estimates. Mode effect refers strictly to measurement error differences due to the mode of survey administration, but, when modes are assigned not randomly, a selection effect can generally occur and appropriate inference methods to evaluate mode effect are needed because the two types of error are confounded. Disentangling selection and measurement effects requires auxiliary information that are assumed to be mode insensitive, acquired from registers or collected by the survey itself. The focus of this work is the experience in the evaluation and treatment of MM effect in the experimental situation of ISTAT survey "Aspects of daily life - 2017", a sequential web/PAPI survey for which a control single mode sample PAPI was planned to make an assessment of mode effect on two independent samples with different techniques. Methods to assess the impact of MM on the quality of the estimate, the representativeness of the two samples and models to evaluate the measurement error and selection effect in the MM sample are experimented.

Keywords: mixed mode, selection effect, measurement effect

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 (de Leeuw, 2005). 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, allowing a gain both in population coverage and response rate. However, it introduces a bias, named mode effect, that must be addressed at different levels: in the design phase by defining the best collection instruments to contain the measurement error; 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 quality of the produced estimates, that 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.

Mixed mode simultaneously generates nonresponse error (selection effects) and measurement error (measurement effects). Selection effects occur when different types of respondents choose different modes to complete the survey. The occurrence of a selection effect is in itself not a problem but its occurrence makes using a MM design valuable. Measurement effects refer to the influence of a survey mode on the answers respondents give, such that one person would give different answers in different modes. Put differently, measurement effects are caused by differences in measurement errors. These errors may originate from differences in, among others, interviewer effects and social desirability, primacy and recency effects, recall bias, acquiescence, etc. (de Leeuw, 2005).

The major problem of MM designs is that selection effects and measurement effects are confounded, especially when modes are administrated in sequential way. Differences (or similarities) between the outcomes of modes can be caused by differences between the respondents or by differences in measurement error. 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.

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.

Exclusive focus on MM survey data precludes evaluation of selection effects and measurement effects separately. This problem is avoided by an extension of MM data with data of a comparable single mode survey (Vannieuwenhuyze et al., 2010).

The focus of the present work is the illustration of the experimentation plan for the treatment of mode effect in the web/PAPI "Multipurpose Survey on Aspect of daily life" (ADL) survey. Through the linkage of survey data with administrative data we exploit the auxiliary variables to define mixed mode models. The final goal is to make an assessment of the introduction of the mixed mode and define an estimation strategy for the future editions of the survey.

The paper is organized as follows: in section 2 the survey context is outlined; section 3 describes the analyses carried out, while section 4 outlines some conclusions.

2. Survey context

The "Multipurpose Survey on Aspect of daily life" (ADL survey) involves yearly 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 through a two stage sample design. The sample of households is selected from the centralized municipal register. In the 2017 edition of the ADL survey a mixed mode technique was introduced for the first time: a web technique has been added to the traditionally used PAPI technique in a sequential design. In order to analyse the impact of the mixed mode on the estimation of the parameters of interest, a survey design was made in which the sample of each municipality was randomly divided into two sub-samples: to the first one, of larger size, the mixed web/PAPI technique has been administered sequentially (mixed mode, MM design); to the second, only the PAPI interview has been proposed (single mode SM, control sample).

Through the linkage of the sample units (respondent and non-respondent) with administrative data, sociodemographic auxiliary variables are obtained and utilized to define mixed mode models. The linkage was performed through the individual code.

3. The analysis of the mode effect

3.1. Outline

The objective of the presented analyses is to evaluate first the impact on the estimates of the survey of the introduction of mixed mode design with respect to the previous single mode design and subsequently to analyse in depth the reasons that determine significant differences in the estimates obtained with the two designs. For this purpose, the study is developed on several levels of analysis: the first level is based on the comparison between the two samples SM and MM; the second one on the mode effect (selection and measurement) of the samples of respondents web and PAPI in the MM design.

In the analysis, the issues described above are affected also by total response. The analysis and treatment of total nonresponse in MM survey is a complex operation due the particular way in which the response process is developed. In fact, the distribution of the sample of PAPI respondents of the follow-up phase depends on the results of the response that is realized in the first phase with the web technique.

In the first level of analysis, tests were performed on the differences in the estimates calculated on the two sample, SM and MM. Subsequent analyses were conducted to study the bias caused by the total nonresponse in the two samples. To this end, auxiliary variables acquired from archives on individuals have been redefined at the household level. The analyses of the total response rates and the indicators of representativeness were conducted in order to identify differences (especially in terms of magnitude of the bias) that could explain, in part, the differences in the estimates of the survey produced with the SM and MM samples. The different composition of samples determined by differences in the total nonresponse could, in fact, contribute to generate differences in the estimates.

The analysis of the mode effect in the MM sample was carried out taking into account the complexity of the problem and an appropriate theoretical reference context. Methods that make the samples of respondents web and PAPI comparable, as propensity score (Rosenbaum and Rubin, 1983), have been used to study the selection effect and the measurement effect of some target variables of the survey.

3.2. The comparison between the single and the mixed mode survey

Test of differences between estimates

To evaluate the differences between the estimates of the main parameters of interest of the survey, obtained with the mixed and the single mode samples, hypothesis tests were carried out (Martin and Lynn, 2011). The hypothesis tests concerned the following estimates: Satisfaction for life (Satisfaction); Health conditions (Health); Valuation of the economic situation compared to the previous year (EcoSit); Reading books in the last 12 months (Books); Frequency of seeing friends (Friends); Habit to smoke (Smoke). The differences between the estimated proportions with the two samples for each response item were studied through the test of the differences in proportions (t-test), while the independence between the response distributions were evaluated as a whole through the Chi-square test. Among these variables, the difference for Satisfaction, Books and Friends resulted significant.

Test of differences between response rates

Having so highlighted some significant differences between estimates obtained from MM and SM surveys, we moved to analyse the impact of nonresponse in the two samples in order to understand the role of the different response processes.

To assess whether the response rate distributions are independent from the individual structural variables, the hypothesis of independence between the response and the variables was tested, distinctly in the two samples. The χ^2 test for the samples SM and MM show that the structural variables considered at household level, typology by number of components and age, income class (5 quintiles: € 11.955, 20.892, 30.028, 46.119), nationality (Italian or not), (below/equal/above high school diploma), geographical area and type of municipality, effectively influence the response in both samples. Analyzing the geographical areas and the type of municipalities the response rates results significantly different in the two samples only in the North-West and in the central Municipalities of the metropolitan area.

The analysis, carried out on respondents only, is aimed at testing whether the fact of responding to SM or MM is independent of structural variables. What emerges is that the distribution by geographical area is significantly different in the two samples; on the contrary, the distribution of respondents by type of municipality, household typology, income class and nationality is not statically different in the two samples.

Analogous analysis is also carried out on the PAPI component of the two samples. The result shows that the distribution by geographical area and income class of the respondents to PAPI is not independent from whether they were selected for the PAPI or web/PAPI samples, differently from the distribution by type of municipality, household typology and nationality.

Analysis of total nonresponse bias

To assess the overall quality of respondents' samples in terms of bias, indicators of representative response, known as *R*-indicators and unconditional partial *R*-indicators, were used. These indicators are based on a measure of the variability of the response propensity and describe how the sample of respondents to a survey reflects the population of interest with respect to certain characteristics. Essentially, they measure how much the sample of respondents in a survey deviates from the representative response; furthermore, the unconditional partial *R*-indicator can be seen as a measure of the contribution of each variable to the representative response (Schouten et al., 2011).

For the analysis of the total nonresponse bias in both (independent) SM and MM samples, *R*-indicators were calculated as: ($R(\rho_x) = 1 - 2S(\rho_x)$ and $\hat{R}(\hat{\rho}_x) = 1 - 2\hat{S}(\hat{\rho}_x)$) where ρ_x is the response propensity estimated thought a logistic regression model and $S(\rho_x)$ is the standard deviation of ρ_x; unconditional partial *R*-indicator is $\hat{P}_u(Z, \hat{\rho}_x) = \hat{S}_B(\hat{\rho}_x|Z)$, being Z an auxiliary variable and $\hat{S}_B$ the between variance.

The auxiliary variables utilized in the response model are: household typology, income class, higher educational level and geographical area.

The table below shows the values of the *R*-indicator and its estimate for single mode (SM) and mixed mode (MM) samples.

Table 1. R-indicators in SM and MM samples

***R*_Indicator**	**SM sample**	**MM sample**
$R(\rho_X)$	0.81195	0.85227
$\hat{R}(\hat{\rho}_X)$	0.81397	0.85376

As response is said to be representative if all the response propensities in the sample are equal, that is when the *R*-indicator is equal to 1, from table 1 emerges that the

MM sample of respondents deviates less from the representative response with respect to the SM sample, 0.85376 in the first and to 0.81195 in the second.

In the following analysis, the *R*-indicator is calculated on the basis of the estimated response propensity through response models defined for each geographical area (North, Center, South and Islands).

Table 2. *R*-indicators in SM and MM samples in the geographical area

***R*_Indicator**	**SM sample**		**MM sample**	
	$R(\rho_X)$	$\hat{R}(\hat{\rho}_X)$	$R(\rho_X)$	$\hat{R}(\hat{\rho}_X)$
North	0.84654	0.84977	0.84043	0.84295
Center	0.75239	0.74822	0.84160	0.83563
South and Islands	0.83956	0.84012	0.90717	0.91357

Table 2 shows that while for the North the values of the *R*-indicators are similar for the two samples, for the other geographical areas they are very different. The response in these cases is more representative when the MM survey is adopted. It seems that, although the web response rates are much lower in the South and Islands, in the MM survey the sample of respondents better reflects the population of interest with respect to certain characteristics used in the models.

The contribution to the representative response of some variables, as household typology, income class and geographical area, is greater in the MM survey respect to SM survey. For example, the value assumed by unconditional partial *R*-indicator for geographical area is equal to 0.00245 and 0.00116 respectively in the SM and MM samples.

In order to show an example of measure of the total bias due to nonresponse, in table 3 the estimated prevalences of one auxiliary variable, the income class, in both SM and MM samples (direct estimates on respondents) are reported in comparison with the estimates obtained on the theoretical sample as a benchmark value (SM+MM). The numbers highlight that the SM sample is overall slightly more biased than the MM sample: this example confirms that the nonresponse bias is different in the two samples and therefore that for an assesment of the mode effect based on a control sample (SM) it is necessary to make the two samples comparable.

Table 3. ***Income class bias in*** **SM and MM samples**

Income class	Benchmark estimate	SM sample estimate	MM sample estimate	SM Absolute bias	MM Absolute bias
I	17.5%	15.1%	15.4%	2.4%	2.1%
II	16.1%	15.6%	15.8%	0.5%	0.3%
III	18.5%	18.6%	19.0%	0.1%	0.5%
IV	22.1%	23.1%	22.5%	1.0%	0.4%
V	25.8%	27.5%	27.3%	1.8%	1.5%
Total				5.8%	4.9%

3.3. The analysis of mode effect in the MM sample based on propensity score

In MM sample difference in the estimates of the parameters of interest of the survey - calculated on the samples of respondents web and PAPI - can be determined either by the different composition of the samples or by differences in measurement errors (Hox et al., 2015).

Moving to get an assessment of selection and measurement effect in the MM sample, a Propensity score stratification adjustment methods was used (Rosenbaum and Rubin, 1983). Propensity score (PS) 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. With adjustments based on PS, the confounding effects of the selection mechanism are mitigated.

The application of this approach implied: an estimation of the propensity score model parameters; the definition of subclassification (strata) of web and PAPI respondents based on propensity score; the validation of the balancing assumption, through a chi-square test of the independence between the mode choice and each of the covariates; for each balanced group, the calculus of weighs that equate the weighted proportion of web respondents with the proportion of PAPI respondents in the same stratum.

A logit regression model was used where the binary response variable is the mode choice web/PAPI. The parameters resulted significant for the following auxiliary variables: geographic region, type of municipality, household typology, income class and higher educational level. For eight out of ten of the deciles of the distribution of

the predicted probabilities the independence hypothesis was accepted for all variables.

For each balanced group *k,* a correction factor, or weight, of the selection effect has been calculated as $w_k = \frac{n_{k,papi}/n_{papi}}{n_{k,web}/n_{web}}$ (Vandenplas et al., 2016), being $n_{k,T}$ the number of respondents to the mode T in the group *k*. This corrector allows an overall evaluation of the mode effect in the balanced classes (Table 4): the selection effect is obtained, following Vandenplas (2016), as the difference between the weighed and unweighted estimates of the respondents to the web mode, while the measurement effect is obtained as the difference between the weighed estimate of the web respondents and the unweighted estimate of PAPI respondents.

Table 4. Selection and measurement effect estimated through PS for Books and Friends

Target variable	Web mean	Weighted Web mean	PAPI mean	Selection effect	Measurement effect
Reading books (last 12 months)					
No	0.479	0.397	0.602	-0.082	-0.123
Yes	0.417	0.522	0.320	0.105	0.097
NR	0.045	0.037	0.031	-0.008	0.014
Frequency of seeing friends					
Everyday	0.102	0.082	0.190	-0.020	-0.089
Sometimes a week	0.235	0.243	0.255	0.008	-0.020
Once a week	0.190	0.202	0.179	0.012	0.011
Sometimes a month	0.189	0.209	0.166	0.020	0.023
Sometimes a year	0.131	0.139	0.081	0.009	0.050
Never	0.052	0.042	0.053	-0.009	-0.002
No friends	0.017	0.017	0.018	-0.001	-0.001
NR	0.025	0.021	0.010	-0.004	0.014

The table shows that for these two variables, which resulted to be the most affected by mode effect, both selection and measurement effect seem to be present, especially for some items.

4. Conclusions and future plans

The analyses highlighted that the two samples, SM and MM, produce generally different estimates. The MM sample is more “representative” of the population with respect to the main sociodemographic variables, but produces both selection and measurement effect for some of the considered target variables. The strategy for the future editions of the survey, which is carried out yearly with MM design, should be based both on prevention of measurement effect and use of calibration method such

as those proposed in Buelens et al. (2015), which seems to allow to stabilise the mode effect in repeated surveys.

5. References

Buelens, B., and Van den Brakel, J. A. (2015). Measurement error calibration in mixed-mode. Sociological methods & Research, 2015, Vol. 4483) 391-426.

de Leeuw, E. 2005. To mix or not to Mix Data Collection Modes in Surveys. Journal of Official Statistics 21(2), 233-55.

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.

Martin, P. and Lynn, P. (2011). The effects of mixed mode survey designs on simple and complex analyses. Centre for Comparative Social Surveys. Working Paper Series. Paper n.04, November 2011.

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.

Schouten, B., Shlomo, N. and Skinner, C. (2011). Indicators for Monitoring and Improving Representativity of Response. Journal of Official Statistics 27: 231–253.

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. (2010). A Method for Evaluating Mode Effects in Mixed-mode Surveys. *Public Opinion Quarterly*, Volume 74, Issue 5, 1 January 2010, Pages 1027–1045,https://doi.org/10.1093/poq/nfq059

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.