Assessing and adjusting bias due to mixed mode in Aspect of Daily Life Survey

**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 NSIs, is adopting especially for social surveys. Its use is spreading both to contrast declining response rates and to reduce the total cost of surveys ([1] de Leeuw, 2005). Using 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 in population coverage and response rate. However, it introduces a bias, named mode effect, that must be faced 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 use of MM, in order to ensure the accuracy of the 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: it is in itself not a problem as its occurrence makes using a MM design valuable. Measurement effects refer to the influence of a survey mode on the given answers, such that one person would give different answers in different modes. Put differently, measurement effects are caused by differences in measurement errors originated from differences, for example, in interviewer effects and social desirability, recall bias, etc. The main critical issue of MM designs is that selection and measurement effects are confounded, especially when modes are administered in sequential way. Differences between the outcomes of modes can be caused by differences between the respondents or by differences in measurement error: method for disentangling the two effects are needed and separate assessment should be carried out.

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 an independent control single mode (SM) PAPI sample was planned to make an assessment of the introduction of the mixed mode. The ADL survey is part of the integrated system of Multipurpose Surveys on households and collects information about recreational and cultural activities in free time, such as sports, reading, cinema, music, the Internet, social relations, issues for the quality of life of people. ADL survey is based on a sample of about 24.000 households, selected through a two stage sample design (municipalities/households) from the centralized municipal register. The aim of the analyses presented in the next sections is to evaluate first the impact on the survey estimates of the introduction of MM 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 samples. The study is developed on several levels of analysis: the first level makes the comparison between SM and MM samples; the second level assesses the mode effect (selection and measurement) in the MM sample.

# 2. the analysis of mode effect

In the first level of analysis, tests have been performed on the differences in the estimates calculated on the SM and MM samples, through Chi-square test to determine if there is a significant difference between the distribution of the answer with respect to the data collection mode and T-test to determine if the difference, between proportions of individuals for each item is significant with respect to the data collection mode ([2] Martin and Lynn, 2011). The results show that the two samples yield significant different estimates.

Subsequent analyses have studied the bias caused by the total nonresponse in the two samples in order to identify dissimilarities that could explain, at least partly, differences in the estimates of the survey produced with the SM and MM samples. The unlike composition of samples, determined by differences in the total nonresponse could, in fact, contribute to generate differences in the estimates. Through the linkage of survey data with administrative data, auxiliary variables are acquired to define models in mixed mode. The auxiliary variables linked on individuals have been redefined at the household level (household typology per number of components and age, income class, higher educational level (below/equal/above high school diploma) and geographical area). Tests on differences between response rates in the two samples using a Z-test have been carried out: response rates differences are significantly different when the household presents some characteristics, such as mixed nationality, lower income class. To assess the overall bias of respondents' samples, the indicators of representative response have been used, known as *R*-indicators and unconditional partial *R*-indicators ([3]Schouten et al., 2011). These indicators are based on a measure of the variability of the response propensity and describe how the sample of respondents to a survey represents 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. The expression of *R*-indicators are given by and where is the response propensity estimated through a logistic regression model and is the standard deviation of . Finally, to evaluate both selection and measurement effects, a method proposed by Vannieuwenhuyze et al. ([4], 2010) has been adopted, based on the probability distributions of the survey categorical variables, estimated from the two comparable respondents’ samples (SM and MM), in which SM sample (only PAPI technique) is taken as a benchmark. In order to make the SM and MM samples comparable, a calibration procedure is adopted, separately for the two samples, starting from the sampling weights.

In the second level of analysis, mode effect in the MM sample has been assessed taking into account an appropriate theoretical reference context. Methods that make the samples of respondents web and PAPI comparable, called propensity score (PS) ([5] Rosenbaum and Rubin, 1983), have been used to study the selection effect and the measurement effect of some target survey variables. With adjustments based on PS, the confounding effects of the selection mechanism are mitigated. The application of this approach implies: an estimation of the propensity score model parameters; the definition of sub-classification (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 items of the covariates; for each balanced group, calculation of correction factors that equate the weighted proportion of web respondents with the proportion of PAPI respondents in the same stratum. A Logit regression model have been used, in which 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 deciles of the predicted probability distribution the independence hypothesis have been accepted for all variables. For each balanced group *k*, a correction factor of the selection effect has been calculated as ([6] Vandenplas et al., 2016), being the number of respondents to the mode *T* in the group *k*. The propensity score, as all weighing methods, can be used to correct the selection effect. In the analyses, this last method based on the assumption of measurement error invariance over time are compared with method that not consider this assumption. In repeated sequential mixed-mode surveys, the assumption is not very sustainable, because the composition of the respondents by mode can change in the subsequent editions of the survey, leading to variations in the total measurement error. To control this aspect, an application of the interesting method proposed by Buelens and Van den Brakel ([7] 2011), has been carried out, aiming at keeping the measurement error constant on the various survey occasions, based on a calibration procedure that takes into account fixed levels of mode proportions.

# 3. Main results

We present here some of the main results of the analyses carried out in the experimentation plan outlined in paragraph 2. The analyses conducted on the response representativeness show important results at national level: *R*-indicator values are equal to 0.81195 and 0.85376, respectively for SM and MM samples; MM sample of respondents deviates less from the representative response with respect to the SM sample. In Table 1 the results of *R*-indicators calculated on the basis of response models defined at geographical area level are shown: while for the North the values are similar for the SM and MM samples, for the other areas they are very different, despite the lower web response rate observed for the South and Island.

Table 1. *R*-indicators in SM and MM samples in the geographical area

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Geographical area | SM sample | | MM sample | |
|  |  |  |  |  |
| North | 0.84654 | 0.84977 | 0.84043 | 0.84295 |
| Center | 0.75239 | 0.74822 | 0.84160 | 0.83563 |
| South and Island | 0.83956 | 0.84012 | 0.90717 | 0.91357 |

The following tables 2 and 3 show, for the variable “*reading books in the last 12 months*”, the results of the estimation of selection and measurement effects obtained through the application of the method proposed by Vannieuwenhuyze et al. ([4], 2010) and the propensity score ([6] Vandenplas et al., 2016).

Table 2. Estimate of the selection and measurement effects for “reading books in the last 12 months” variable comparing SM and MM respondent’s samples

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Category | Selection effect | Measurement effect |
| Reading books  in the last  12 months | No | 0,1478 | -0,0727 |
| Yes | -0,1767 | 0,0416 |
| NR | 0,0288 | 0,0311 |

Table 3. Estimate of the selection and measurement effects for “reading books in the last 12 months” variable in MM respondent’s sample

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variable | Category | Weighted Web mean | Web  mean | PAPI mean | Selection effect | Measurement effect |
| Reading books (last  12 months) | No | 0.485 | 0.451 | 0.618 | 0.034 | -0.132 |
| Yes | 0.432 | 0.508 | 0.347 | -0.075 | 0.085 |
| NR | 0.043 | 0.041 | 0.035 | 0.002 | 0.007 |

The two results are quite similar and show the existence of both a selection and a measurement effect in the same direction.

Finally, Table 4 shows the comparison of the estimates for “Reading books” deriving from the application of different methods. These methods are based on calibration procedures with respect to distributions of the same socio-demographic totals (age class, sex, educational level) at geographical area level, but differ for other constraints or for the sampling weights used in the procedure: 1) only socio-demographics; 2) socio-demographics and observed fixed levels of mode proportions by six municipal typologies; 3) socio-demographics and hypothesized fixed levels of mode proportions by six municipal typologies; 4) socio-demographics with sampling weights corrected for the web selection effect through correction factors .

Table 4. Estimate of “reading books in the last 12 months” variable with different methods

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Category | Estimate (%) | | | |
| Meth. 1 | Meth. 2 | Meth. 3 | Meth. 4 |
| Reading books (last  12 months) | No | 59,82 | 58,88 | 58,54 | 59,81 |
| Yes | 36,51 | 37,47 | 37,76 | 36,46 |
| NR | 3,67 | 3,65 | 3,70 | 3,73 |

What emerges from the table is that the two calibrations including the constraints with respect to fixed level of mode proportions (methods 2 and 3) determine a difference in the estimate of about one percentage point.

The results presented in this abstract would need a significance assessment, based on tests or replication methods, which are in progress.

# References

1. E. de Leeuw, To mix or not to Mix Data Collection Modes in Surveys, Journal of Official Statistics 21(2) (2005) pp. 233-55.
2. P. Martin and P. Lynn, The effects of mixed mode survey designs on simple and complex analyses. Centre for Comparative Social Surveys. Working Paper Series. Paper n.04 (2011).
3. B. Schouten, N. Shlomo, and C. Skinner, Indicators for Monitoring and Improving Representativity of Response. Journal of Official Statistics 27 (2011), pp. 231–253.
4. J. T. A Vannieuwenhuyze, G. Loosveldt and G. Molenberghs, A Method for Evaluating Mode Effects in Mixed-mode Surveys. Public Opinion Quarterly, 74(5) (2010), pp. 1027–1045.
5. P. R. Rosenbaum and D. B. Rubin, The Central Role of the Propensity Score in Observational Studies for Causal Effects, Biometrika, 70 (1) (1983), pp. 41-55.
6. C.Vandenplas, G. Loosveldt and J. T. A. Vannieuwenhuyze, 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 10(2) (2016), pp. 119-142.
7. B. Buelens and J. A. Van den Brakel, Measurement error calibration in mixed-mode, Sociological methods & Research, 4483 (2015), pp 391-426.