

SELECTION BIAS AND REPRESENTATIVENESS OF SURVEY SAMPLES: THE EFFECTIVENESS OF MIXING MODES AND SAMPLING FRAMES

Background and research questions

Nowadays mixed-mode approaches are used to deal with the non-coverage issue and the nonresponse error in sample surveys. In literature there are many examples of surveys that mix web, telephone and F2F modes, adopting concurrent or sequential designs in experimental or non experimental studies. The issues of interest are various, e.g. experiments on question format, differences in nonresponse and coverage, social desirability bias, and data quality estimates. Moreover, the studies often use the same sampling frame for the mixed-mode survey, and show contrasting findings on the differences between the samples. Within this context, drawing on our previous work (Bartoli, Respi, and Fornea, 2018), we apply a mixed-mode survey design to different sampling frames (landline phones list and online panel). The problem with telephone coverage is exacerbated because households with landline are not equally represented throughout the Italian population. We hypothesize that this source of bias, combined with nonresponse error, could be reduced adopting approaches that use different sampling frames. This poster aims to assess the representativeness of samples from a mixed-mode survey design (web-landline) and a telephone survey (calling mobile and landline phone numbers), comparing their estimates to the observed values from registered voters' records, and to the socio-economic characteristics of the Italian population.

Method

To assess the representativeness of the samples, we first compare the estimated voting behaviour from the two survey designs to the observed voting behaviours ("true" values) in the last political elections. We conduct bivariate analyses and use the mean absolute error, the largest absolute error, and the percentage point differences, as accuracy metrics. In addition, we also compare the employment status and education of our respondents to those of the Labour Force Survey respondents, calculating (as accuracy metric) the percentage point error for the modal category of the benchmark.

Data

We use data from 6 telephone and web surveys conducted in Italy (period March 2018 - January 2019) on landlines or mobile phones owners and on members of the Italian online panel Opinione.net. We designed a sequential mixed-mode survey (a Computer Assisted Web Interview - CAWI survey followed by a Computer Assisted Telephone Interview - CATI survey, using two different sampling frames) and a survey with two different sampling frames (a Computer Assisted Mobile phone Interview - CAMI survey followed by a CATI survey). The same research institute commissioned all the surveys to Demetra opinioni.net s.r.l., that equally divided each sample (about 1,000 panelists/interviewees) between the two survey designs: half respondents to the mixed-mode design, and half to the telephone survey. Both designs adopt a quota sampling: quotas were defined to be proportional to the gender within age, and geographic area of residence distributions of the Italian population (we used administrative data from dati.istat website as benchmark). The questionnaires used for each survey are different, but they are all about attitudes and political behaviour, and include a question with the same question wording on vote in the last national elections (i.e. "In the national elections of 4 March 2018, which party did you vote for?"), that we used in our analyses. In our previous work (Bartoli, Respi, and Fornea, 2018) we used a question on vote taken in 2014, while in this study we focus on vote taken in 2018. Our surveys were carried out in the months immediately after the political elections. We speculate that the "memory effect" should not have occurred, thus removing a potential source of bias from the estimates of voting behaviour.

TABLE 1. OVERVIEW OF THE TWO STUDY DESIGNS.

Mode	Type of survey	
	mixed-mode CAWI + CATI	telephone CAMI + CATI
Design	sequential	sequential
Sampling frame	online panel + landline phones list	unknown (RDD) + landline phones list
Sampling method	quota sampling (gender within age, and geographic area of residence)	quota sampling (gender within age, and geographic area of residence)

TABLE 2. OVERVIEW OF THE SURVEYS

Month	Survey	INITIAL SAMPLE		FINAL SAMPLE		COOPERATION RATE
		CAWI	CATI	CAWI	CATI	
MARCH 2018	CAWI	6,823	371	371	5	66
	CAMI-CATI	20,506	257	257	4	4
APRIL 2018	CAWI	4,778	250	250	5	57
	CAMI-CATI	16,444	503	503	3	3
MAY 2018	CAWI	11,186	250	250	2	75
	CAMI-CATI	13,230	505	505	4	4
SEPTEMBER 2018	CAWI	4,169	231	231	5	5
	CAMI-CATI	5	274	274	3	52
DECEMBER 2018	CAWI	14,633	501	501	3	3
	CAMI-CATI	12,218	318	318	3	3
JANUARY 2019	CAWI	5	306	306	68	68
	CAMI-CATI	16,356	615	615	4	4

We also use secondary data as benchmarks: the registered voters' database (2018), and the Labour Force Survey (2017).

Results: representativeness of voting behaviour

Results from the analyses on the magnitude of bias in the estimates of voting behaviour show some differences between the CAMI - CATI and the CAWI - CATI designs.

In particular, Table 3 focuses on the mean absolute error, and shows that the CAWI - CATI design performs better than the CAMI - CATI one in all the six surveys, when representing voting behaviour.

TABLE 3. MEAN ABSOLUTE ERROR FOR THE QUESTION ABOUT VOTING BEHAVIOUR OF EACH SURVEY SAMPLE.

	CAMI-CATI	CAWI-CATI
March 2018	3,75	3,48
April 2018	4,42	3,93
May 2018	3,04	1,81
September 2018	2,95	2,78
December 2018	4,84	3,31
January 2019	3,52	3,48

Then, we focus on the category of the question about voting behaviour with the largest absolute error for each survey sample. Table 4 reports the results.

Table 4. Largest absolute error for each survey sample.

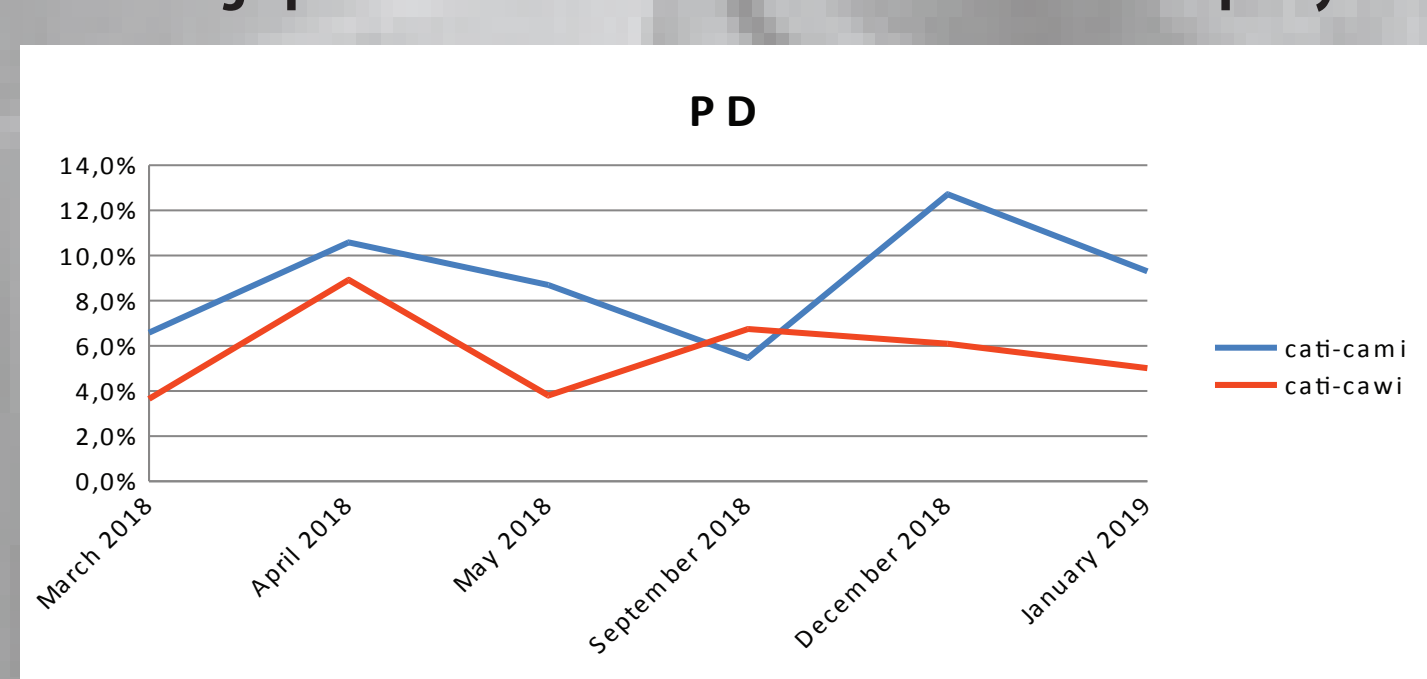
	CATI-CAMI	CATI-CAWI
March 2018	-7,3	-5,8
April 2018	10,6	8,9
May 2018	8,7	3,8
September 2018	-6,2	-7
December 2018	12,7	-7,6
January 2019	9,3	-8,5

Three main findings stand out:

- the values of the largest absolute error are smaller for the CAWI - CATI than for the CAMI - CATI survey design;
- the CAMI - CATI samples tend to systematically over-represent people voting 'Partito Democratico';
- the CAWI - CATI samples tend to systematically under-represent people voting 'Forza Italia'.

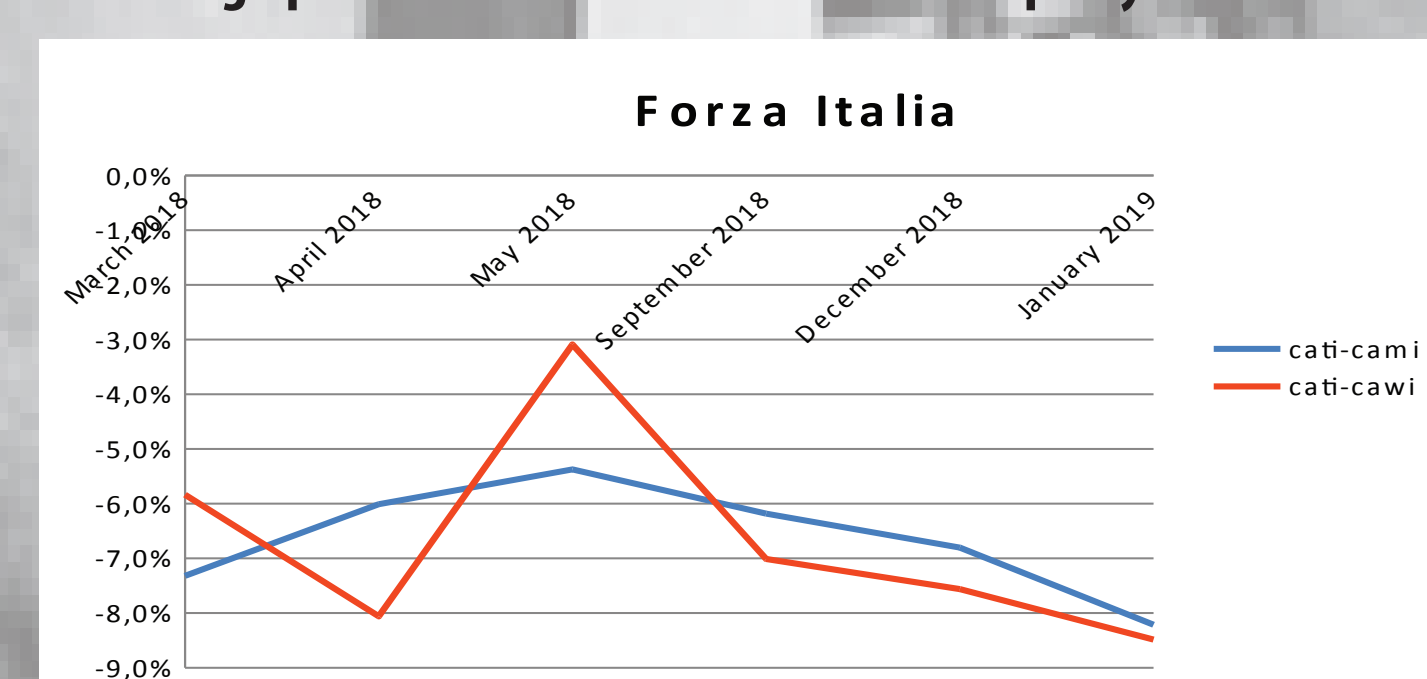
Lastly, looking at the percentage point differences for the main parties (Graphs 1-4), we find some patterns.

Graph 1. Percentage point differences for 'Partito Democratico' party.



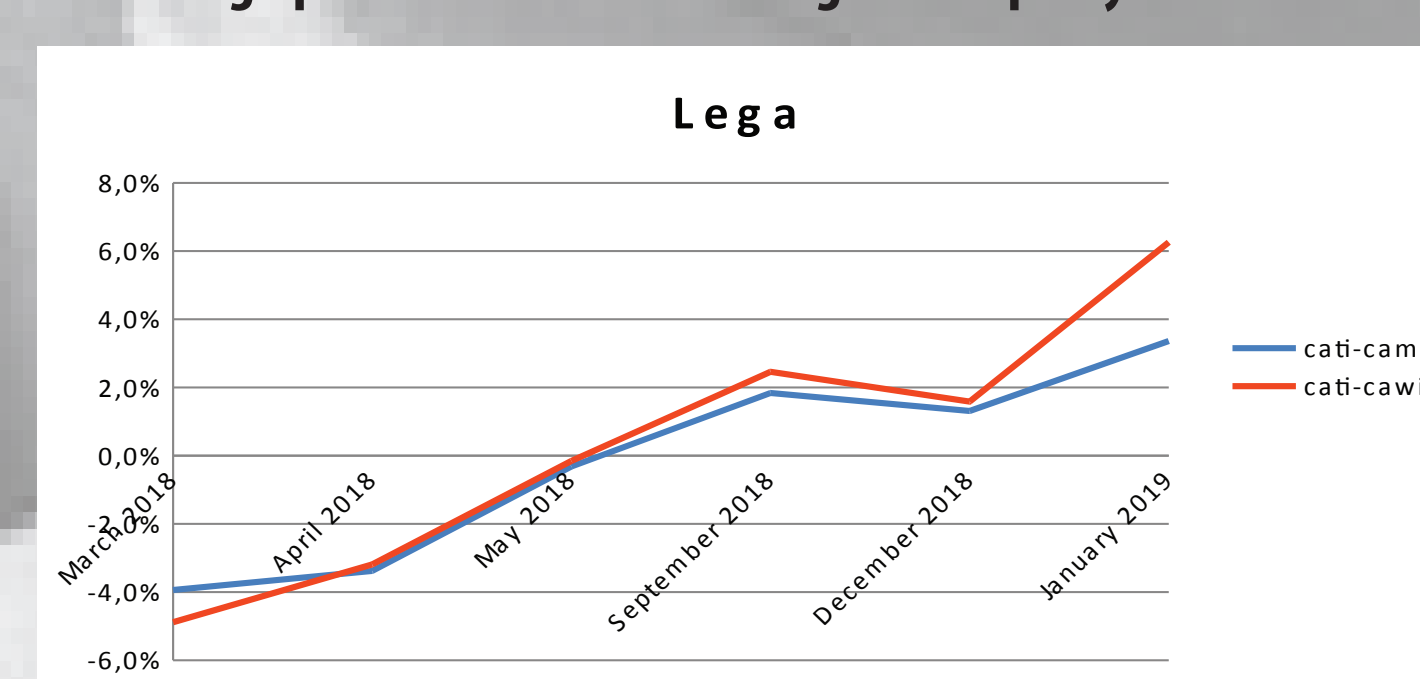
The 'Partito Democratico' party is always over-represented. The differences are systematically higher for the CAMI - CATI than for the CAWI - CATI survey design, except for the 'September 2018' survey where the CAMI - CATI sample performs better than the CAWI - CATI one. The telephone survey trend shows a higher variability than the mixed-mode survey series.

Graph 2. Percentage point differences for 'Forza Italia' party.



The 'Forza Italia' party is always under-represented in both the survey designs. Four out of six CAWI - CATI samples are more biased than the telephone ones.

Graph 3. Percentage point differences for 'Lega Nord' party.



The 'Lega Nord' party shows an interesting pattern for both survey designs: 'Lega Nord' voters are under-represented in the first three surveys, whereas, since September 2018, they are always over-represented. Overall, the percentage differences systematically increased for both the survey designs from March 2018 to January 2019.

Graph 4. Percentage point differences for 'Movimento 5 Stelle' party.

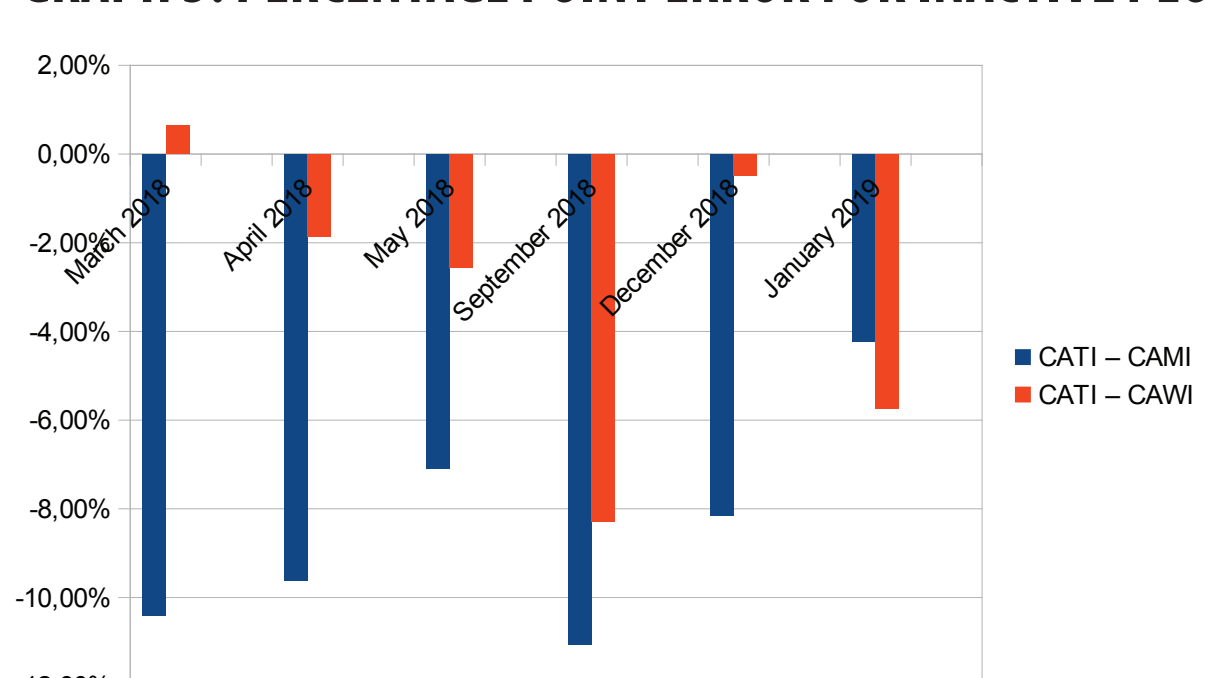


The 'Movimento 5 Stelle' party does not show a clear pattern throughout the surveys. We can say that the CAWI - CATI respondents are more likely to be 'Movimento 5 Stelle' voters (except for the survey carried out in September 2018) than the CAMI - CATI interviewees. However, the most recent survey (i.e. January 2019) reports very low (and equal) values of the percentage point differences for both the mixed-mode and the telephone samples.

RESULTS: REPRESENTATIVENESS OF SOCIO-DEMOGRAPHIC CHARACTERISTICS

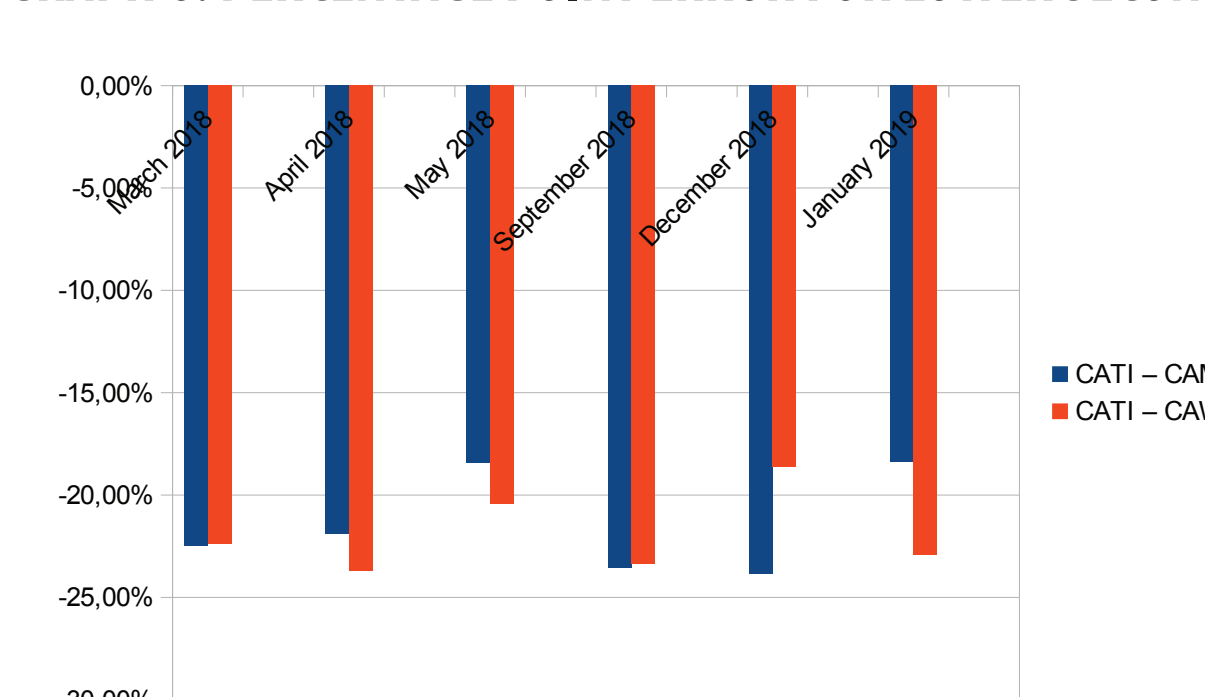
The analysis on bias in the estimates of the socio-economic characteristics of our survey samples reports high values for both the CAWI - CATI and the CAMI - CATI survey designs. Graphs 5 and 6 show the percentage point error for the modal categories of the employment status (i.e. inactive people) and education (i.e. lower secondary education). Graph 5 shows the CAMI-CATI vs CAWI-CATI percentage point error for people in employment.

GRAPH 5. PERCENTAGE POINT ERROR FOR INACTIVE PEOPLE.



The CAMI - CATI samples systematically underrepresent inactive people (Graph 5). The CAWI - CATI survey design performs better in representing this category. Indeed, in comparison with the telephone series, the magnitude of the error is always smaller, but the last survey carried out in January 2019.

GRAPH 6. PERCENTAGE POINT ERROR FOR LOWER SECONDARY EDUCATION.



Graph 6 shows that both telephone and mixedmode survey samples are not representative of people who have not gone beyond lower secondary education: all the samples underrepresent this category and the differences with the general population are very marked (at least -18 percentage points).

Conclusions

We compared the estimates from six telephone and mixedmode surveys with benchmark data and we assessed the sample representativeness. We focused our analyses on voting behaviour, employment status, and education of respondents. The results are consistent with those from our previous work (Bartoli, Respi, and Fornea, 2018), and show that mixing both modes and sampling frame, as in the CAWI - CATI survey design, is a more effective strategy in reducing selection bias. In particular, the following main findings stand out.

- The CAWI - CATI survey design performs better than the CAMI - CATI one in representing the overall voting behaviour:
 - the value of the mean absolute error is lower in all the surveys;
 - the magnitude of the largest absolute error is smaller in all the surveys.
- When looking at the parties selected by the respondents, we identified four main patterns:
 - the CAMI - CATI samples tend to systematically over-represent people voting 'partito democratico';
 - the CAMI - CATI samples under-represent people voting 'Movimento 5 Stelle';
 - both the CAMI - CATI and the CAWI - CATI samples under-represent people voting 'Forza Italia';
 - there are no differences between the two survey designs when representing people voting 'Lega Nord', but all the differences increased over time.
- The CAWI - CATI survey design performs better than the CAMI - CATI one in representing the employment status of the Italian population. Indeed, the magnitude of the percentage point error for inactive people is always smaller, but in the last survey.
- Both the CAWI - CATI and the CAMI - CATI survey designs end up not representing people who have not gone beyond lower secondary education.

REFERENCES

- ATKINSON, IONINA RAE, ALEX N. ADAMS, AND R. MICHAEL ALVAREZ. 2014. "NONRESPONSE AND MODE EFFECTS IN SELF- AND INTERVIEWER-ADMINISTERED SURVEYS." POLITICAL ANALYSIS 22: 304-320.
- DE LEEUW, EDITH, JOOP HOX, AND ANNETTE SCHEPPELZEL. 2011. "MODE EFFECT OR QUESTION WORDING? MEASUREMENT ERROR IN MIXED MODE SURVEYS." JSM (JOINT STATISTICAL MEETINGS) PROCEEDINGS - SECTION ON SURVEY RESEARCH METHODS.
- KOLENINOV, STANISLAV, AND COURTNEY KENNEDY. 2013. "MODE EFFECT ANALYSIS AND ADJUSTMENT IN A SPLIT-SAMPLE MIXED-MODE WEB-CATI SURVEY." JSM (JOINT STATISTICAL MEETINGS) PROCEEDINGS - SECTION ON SURVEY RESEARCH METHODS.
- LUGTIG, PETER, GERTY J.L.M. LENSVELT, MULDER, RENCO FRIEDERS, AND ASSIN GREIF. 2011. "ESTIMATING NONRESPONSE BIAS AND MODE EFFECTS IN A MIXED-MODE SURVEY." INTERNATIONAL JOURNAL OF MARKET RESEARCH 51 (5): 669-686.
- REVILLA MELANIE. 2015. "COMPARISON OF THE QUALITY ESTIMATES IN A MIXED-MODE AND A UNIMODE DESIGN: AN EXPERIMENT FROM THE EUROPEAN SOCIAL SURVEY." QUALITY AND QUANTITY 49: 1219-1238.
- SALA, EMANUELE, AND ROBERTO LILLINI. 2015. "UNDERCOVERAGE BIAS IN TELEPHONE SURVEYS IN EUROPE: THE ITALIAN CASE." INTERNATIONAL JOURNAL OF PUBLIC OPINION RESEARCH. DOI: 10.1093/IJOPRE/D033.
- VANUIJENHUYZE, JOBBE T. A., AND MELANIE REVILLA. 2013. "EVALUATING RELATIVE MODE EFFECTS ON DATA QUALITY IN MIXED-MODE SURVEYS." SURVEY RESEARCH METHODS 7 (3): 157-168.
- FEARER, DAVID S., JON A. KRISNICK, LINCHAI CHANG, HAROLD S. JAVITZ, MATTHEW S. LEVENDOSKY, ALBERTO SIMPSON, AND RUI WANG. 2011. "COMPARING THE ACCURACY OF RDD TELEPHONE SURVEYS AND INTERNET SURVEYS CONDUCTED WITH PROBABILITY AND NON-PROBABILITY SAMPLES." PUBLIC OPINION QUARTERLY 75 (4): 709-47.
- BARTOLI B., RESPI C., FORNEA M. (2018). "REPRESENTATIVENESS OF SURVEY SAMPLES: MIXING MODES AND SAMPLING FRAMES". POSTER PRESENTATION AT THE CONFERENCE GENERAL ONLINE RESEARCH, COLOGNE (GERMANY).