

ADAM RYBAK
Adam Mickiewicz University

Declarative and Behavioral Data in Predicting Respondents Survey Mode Preference

Abstract: Using previously established knowledge about survey mode preferences distribution in a population can be one of many ways of improving representativeness and quality of data gathered by survey research. Apart from mode preference existence and stability, the main problem concerns the question: which of the sources of information about mode preference could be treated as trustworthy. Because real observed choices are typically treated as better predictors of future choices than declaration only, this paper tries to answer the question: 'Are the declarations a good predictor of preference in comparison to real choices?' For this purpose, it uses combined data from 1) the 2015 Polish mixed-mode ESS experiment and 2) data from ESS8 in Poland. The multinomial logistic regression survey preference model includes socio-demographic variables accompanied by declaration/choice control variable. The results suggest significant differences between choices/declarations. Findings could be used to refine the contact strategies used in surveys.

Keywords: European Social Survey, survey mode preference, mixed-mode survey design, adaptive survey design.

Introduction

The analysis of survey mode preference is one of many approaches employed to deal with the continuously rising level of nonresponse error (and additionally with satisficing—[Smyth, Olson, and Kasabian 2014](#)). Although there are some definitional discrepancies—especially concerning the difference between observed and declared preference ([Al Baghal and Kelley 2016: 145](#))—it is usually described as a relatively stable propensity toward participation in the survey research in one mode (at the expense of others). It is worth mentioning that the time stability of preferences is sometimes questioned ([Al Baghal and Kelley 2016: 162](#); [Wardle and Robinson 2007](#)). Despite the evidence for the usability of the surveying approach based on mode preference predictions being still sparse, it is mentioned in more recent handbooks ([Dillman, Smyth, and Christian 2014](#)).

The two main strategies of gathering data about survey mode preference distribution in the population are asking respondents about their preference in the single-mode survey or giving them the choice of mode in the mixed-mode survey. Ostensibly, observing choice could look like the best way of predicting future behavior, but this strategy comes with a cost of reducing the response rate ([Medway and Fulton 2012](#)), thus counteracting the primary goal. On the other hand, asking about preferences in one mode leads to a significantly higher number of respondents selecting the mode of the question as preferred

(Groves and Kahn 1979; Tarnai and Paxson 2004), so this is not the perfect solution either.

Gathering data about the preference of both behavioral and declarative origin could lead to the creation of classification models, which can be used for allocating a sampled person for their probable preferred mode without asking. The usefulness of knowledge about mode preference distribution also manifests in the possibility of preliminary selection effect assessment. “Mixed-mode surveys are only advantageous over single-mode surveys if selection effects occur” (Vannieuwenhuyze 2013). The adequately designed mixed-mode survey can lead to a significant decrease in fieldwork costs. Centre of Sociological Research in the Institute of Philosophy and Sociology of Polish Academy of Sciences (Polish ESS Fieldwork Agency) declared that representative survey in mail mode would cost 40% of the analogous survey in face-to-face mode (Villard and Fitzgerald 2017). Nevertheless, the two major examples of mixed-mode survey experiments in central-eastern Europe led to the final unit-cost being very similar (or even higher) than in the corresponding single-mode face-to-face surveys (Ainsaar et al. 2013; Sztabiński 2018). However, both mixed-mode studies had a pioneering character in their countries in contrast to the refined methodology of main (f2f) surveys—in the next rounds achieved cost could be lower.

Rybak (2018) used data about mode choices in the Polish survey experiment to build a model assessing the basic socio-demographic indicators of preference. The main aim of that study was to determine who can be pushed to mail mode in Poland (this limitation comes from the assumption that the preferences are dependent on the local, cultural and infrastructural conditions). This study combines the Polish behavioral data and declaration data about survey mode preferences with the main aim of estimating how much both sources are similar in assessing the impact of preference predictors. Because real choices are assumed to be the better reflection of preference, this question can be transformed into: ‘Are the declarations a good predictor of preference?’

Hypotheses

Apart from the main question, however, comparing the data from both studies in more specific areas could be useful. The aforementioned first study found support for several hypotheses based on survey methodology literature (Diment and Garrett-Jones 2007; Groves and Couper 1998: 113, 147, 176; Haan, Ongena, and Aarts 2014; Millar, O’Neill, and Dillman 2009; Smyth, Olson, and Millar 2014). They can be used as hypotheses for this research:

H1: There will be a higher preference toward the mail mode among the female population.

H2: There will be a higher preference toward the Web mode among the male population.

It is worth mentioning that from the first study (2015) to the second study (2017), internet usage in Poland changed from 68% to 76% of the population (International Telecommunication Union 2019).

H3: There will be a higher preference toward an interviewer mode survey among residents of villages, and/or small cities, than the bigger ones.

H4: There will be a higher preference toward a postal mode survey among residents of bigger cities, than of small ones and/or villages.

Rybak's study also came to two findings contradictory to the previous research, which can also be used as hypotheses:

H5: There will be a curvilinear relationship between age and the mail mode preference with middle-aged people preferring it the most.

H6: There will be no significant relationship between age and the Web mode preference.

The last hypothesis is based on the central question of this study stated above. As the declarations about preferences are used as the indicator of preference (and future choices) (e.g. [Olson, Smyth, and Wood 2012](#)), so in two samples representative of the Polish population with just a 1.5-year gap, there should be no significant difference in findings.

H7: There will be no significant difference between behavioral and declarative data concerning survey preferences.

The use of basic socio-demographic predictors only is determined by the idea of creating the classification model that can be used to allocate respondents to mode before reaching them. This approach is called Adaptive Survey Design (e.g. [Tourangeau et al. 2016](#)). Furthermore, as the sampling frame most usually utilized in Poland (which is the PESEL database) contains information about gender, age, and place of residence, these variables are used in this study.

Data and Methods

Data for this study comes from two sources. The first one is the Mixed-mode experiment conducted parallelly to European Social Survey Round 7 in Poland by Paweł & Franciszek Sztabiński. Second is the national module attached to European Social Survey Round 8 in Poland. Stated experiment (**S1**) took place between 7th April and 30th September 2015 ([Sztabiński 2018](#); [Sztabiński and Sztabiński 2016](#)) and used an identical sampling scheme as main ESS7 study ([European Social Survey 2018a](#)). ESS8 (**S2**) in Poland took place between 7th November 2016 and 22nd February 2017 ([European Social Survey 2018b](#)). Both studies resulted in the creation of datasets based on samples representative for the Polish population.

S1 mixed-mode design was neither strictly sequential nor parallel. The respondent was first contacted by mail. The envelope contained the paper self-administered questionnaire, return envelope, examples of finding from previous ESS research, gift (a notepad with a magnet) and cover letter. The letter stated that the sampled person could answer by mail mode or access the Web questionnaire on the dedicated website. When no answer was given, reminders were sent, and if there was still no response, interviewers visited selected addresses. A sampled person then had the option to participate in a face-to-face interview with interviewers from ESS7 in Poland. If there was no contact or cooperation, a third letter was sent, this time containing only the paper questionnaire with the return envelope.

S2 was the single-mode face-to-face interview but contained an additional question about the preferred mode of participation: F2F, CATI, mail, or Web (also 'no preference'; 'do not want to participate at all'). As CATI preference was absent in the first study, these responses were excluded from further analysis. Because that could lead to the breach of the assumption of irrelevant alternatives, the test of Hausman-McFadden ([Long and Freese 2014: 407–9](#)) was conducted. Results proved to be statistically insignificant.

In the first study, 800 people were sampled, and 419 questionnaires were completed—the response rate was 55.9%. 164 (39.1%) people answered via mail mode, 24 (5.7%) by Web mode, and 231 (55.2%) by the mode with the interviewer present. In the second study, the number of sampled individuals was 2675, and 1694 questionnaires were completed—the response rate was 69.6%. 1213 (71.6%) respondents declared face-to-face mode preference, 96 (5.7%) mail mode preference and 190 (11.2%) Web mode preference. Frequencies for both studies are presented in [tables 1 and 2](#).

Table 1

Frequencies of mode choices in ESS7 mixed-mode experiment in Poland

Study 1 ESS7PL mixed-mode		
Choice	Frequency	Percentage
Face2face	231	55.2
Mail	164	39.1
Web	24	5.7
Total	419	100.0

Table 2

Frequencies of mode preference declarations in ESS8 in Poland

Study 2 ESS8PL		
Declaration	Frequency	Percentage
Face2face	1213	71.6
Mail	96	5.7
Web	190	11.2
Phone	27	1.6
Not interested	74	4.4
No preference	81	4.8
Refusal	2	0.1
No answer	11	0.7
Total	1694	100.0

The combined dataset, which included information about age, gender, size of settlement of residence, weights, and preference of each respondent from both studies, was created. Preferences were coded dependent on the source study as participation in one mode in the mixed-mode experiment (the behavioral indicator of preference) or as a declaration of preferring one of the possible modes of contact in the single-mode survey (the declarative indicator of preference).

The multinomial logistic regression model was computed in STATA 15, containing age, gender, size of settlement, and source dataset as independent variables and preference (face-to-face, mail, or Web) as the dependent variable. Face-to-face was treated as the base outcome. Age and urbanicity were treated as categorical variables because of a supposed non-linear relationship. Design weights from both surveys were used in the model. In the

tested models, which included interactions between variables, these interactions turned out to be not significant.

Results

The odds of mail mode preference in comparison to face-to-face were significantly higher for females, individuals 15–19 and 25–64 years old (as opposed to 75 and older) living in cities having more than 10k inhabitants (as opposed to the residents of villages) and for the observed, behavioral data. Therefore, there is clear support for *H1*, *H3* and *H4*, but no support for the *H7*. The odds for people 45–54 are the highest from each category and are significantly higher than the value of base-level (75+) so there is also some support for *H5*.

The odds of Web mode preference in comparison to face-to-face were higher for people living in cities having 10–19k inhabitants and cities with 50k inhabitants or more as opposed to residents of villages. They were also significantly higher for respondents 15–54 years old in comparison to people 75 and older. The ratio was also lower for behavioral data, but the difference was at the limit of statistical significance ($p < 0.051$). Therefore, there is clear support for the *H3*, unclear for *H7*, but no support for *H2* and *H6*.

McFadden's pseudo- R^2 of the model was 0.1824 and the regression was significant as a whole. Model excluding type-of-data variable had an R^2 value of 0.0938. The model is presented in [table 3](#).

Table 3

Multinomial logistic regression: Mail and Web survey mode preference relative risk ratios

Model (base=F2F)	Mail RRR	Web RRR
Female	2.149**	1.196
Age (base 75+)		
15–19	3.012*	21.56**
20–24	2.429	27.99**
25–34	2.669*	14.92**
35–44	2.907*	12.12**
45–54	3.758**	6.944*
55–64	2.903**	4.095
65–74	2.198	1.374
Urbanicity (base village)		
city <10k	1.330	0.969
city 10–19k	1.859*	2.069*
city 20–49k	1.950**	1.479
city 50–99k	2.456**	3.517**
city 100–199k	1.806*	2.744**
city 200–499k	3.186**	3.363**
city 500k+	3.547**	7.164**
Type of data (base declarative)		
behavioral data	9.302**	0.620'
Constant	0.011**	0.008**

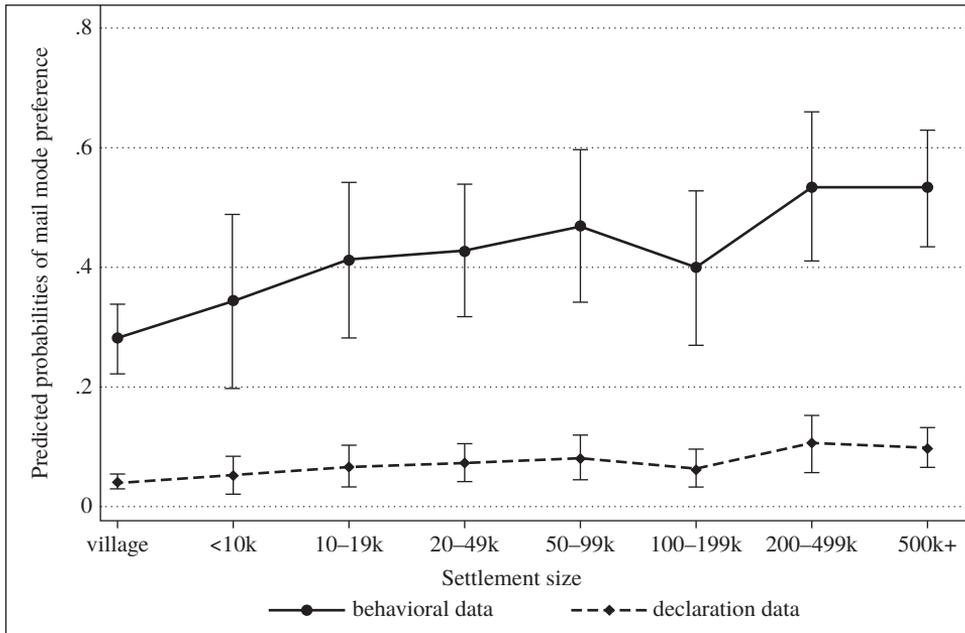
** $p < 0.01$,

* $p < 0.05$,

' $p < 0.051$

Predicted probabilities of mail mode preference for age and urbanicity categories were computed. The charts below show the probabilities with 95% confidence intervals for the variable “dataset” held at values “behavioral data” and “declaration data.” There are significant differences between the age and urbanicity levels, but the differences between values for different datasets are much more apparent.

Chart 1
Predicted probabilities—mail mode propensities for urbanicity categories.
 ‘Dataset’ variable held at ‘behavioral’ and ‘declaration’ level

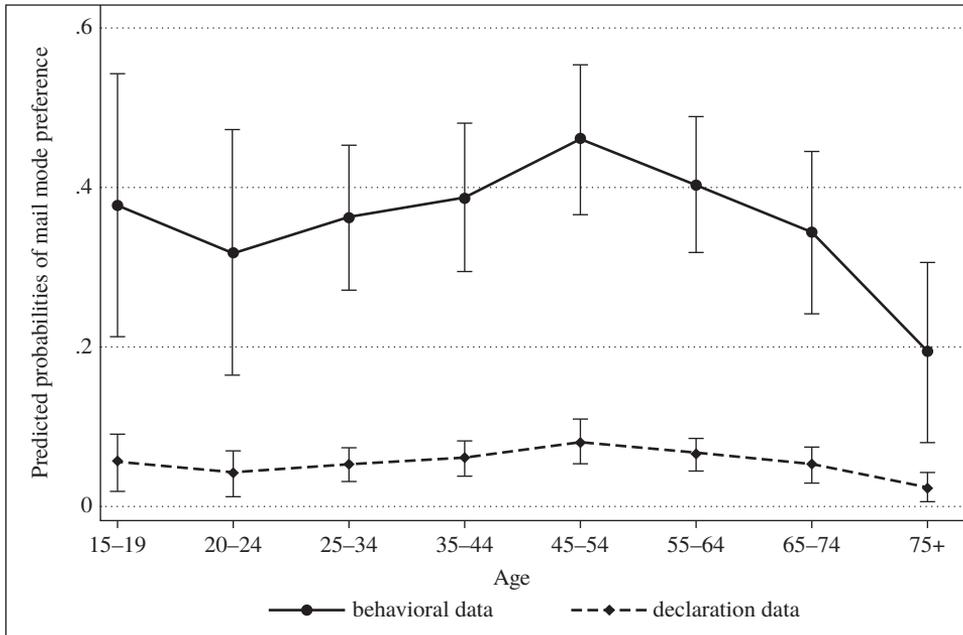


Discussion and Conclusion

The primary result of this study is that according to the model declarations do differ from actions (choices), especially concerning the mail mode. There are some relationships shapes maintained between results from both sources of data, and between the current model and previous model by Rybak (2018). Predicted probabilities of mail mode preference as collated with age categories still have the curvilinear shape, which can be seen in the odds ratio values or the predicted probabilities charts. The people aged 45–54 still have the highest mail mode propensities. Nevertheless, it cannot be omitted, that differences between declarations and choices are important. The odds of mail mode preference were nine times higher in the mixed-mode experiment than in the ESS8 preference question. This leads to the answer to the question asked in the introduction—can we really treat declarations as a good predictor of the preference? Probably not. They could be used in the absence of better auxiliary data, but they are imperfect.

Chart 2

Predicted probabilities—mail mode propensities for age categories.
 'Dataset' variable held at 'behavioral' and 'declaration' level



The study has many limitations, especially concerning the mixed-mode experiment part. The design of it was not strictly parallel, which can suggest that mail mode preference has been artificially inflated in comparison to face-to-face mode (because of the fact that interviewers contacted only nonrespondents in mail/Web modes). On the other hand, the declarations in favor of the mode with the interviewer present may be raised by the mode of the question itself. Another difficulty comes from the probable instability of mode preference. All of these problems were addressed in the previous sections of the article.

However, what can be seen in this study is that there are essential differences between mail and Web preferences. Mail preference was observed as much more frequent in real choices, as Web mode was much more frequent in declarations. This difference cannot be attributed to the defect in design, because Web and mail modes were offered simultaneously in the experiment. Change in Internet usage in Poland in the 1.5-year period between the studies also was not that big. One of the possible explanations of this fact could be based on treating the mode preference question as a sensitive question.¹ People asked by the interviewer can try to appear as more modern and tech-savvy by declaring Web mode preference. In the real situation, however, the usage of mail mode can be much more convenient to them, because of lack of digital literacy, the necessity of mode switch (when the sample is contacted by mail, because of the address frame) or other reasons.

The cost of face-to-face survey research in Poland is still relatively low compared to the cost in western countries. Because of this, mixed-mode surveys are still seen as one

¹ I would like to thank dr. Edith de Leeuw for suggesting me this interpretation.

of the possible options, while in the west they start to become a necessity. We can expect that the processes of the response rate reduction and fieldwork cost increase will accelerate in Poland in the future. Because of that, the preparation for the upcoming situation, also by refining the mixed-mode designs best fitted for the local environment (especially in the adaptive survey design framework) could be beneficial.

Acknowledgments

I would like to thank Paweł Sztabiński for the data on which this study is based and Piotr Jabkowski for supervision over my research.

References

- Ainsaar, M., Lilleoja, L., Lumiste, K., and Roots, A. 2013. *ESS Mixed Mode Experiment Results in Estonia (CAWI and CAPI Mode Sequential Design)*. Tartu: University of Tartu, Institute of Sociology and Social Policy.
- Al Baghal, T., and Kelley, J. 2016. The Stability of Mode Preferences: Implications for Tailoring in Longitudinal Surveys, *Methods, Data, Analyses 10* (102): 143–66.
- Dillman, D.A., Smyth, J.D., and Christian, L.M. 2014. *Internet, Phone, Mail, and Mixed Mode Surveys: The Tailored Design Method (4th Ed.)*. Hoboken: John Wiley & Sons Inc.
- Diment, K., and Garrett-Jones, S. 2007. How Demographic Characteristics Affect Mode Preference in a Postal/Web Mixed-Mode Survey of Australian Researchers, *Social Science Computer Review 25* (3): 410–17.
- European Social Survey. 2018a. *ESS-7 2014 Documentation Report*. Edition 3.2. Bergen, European Social Survey Data Archive, NSD—Norwegian Centre for Research Data for ESS ERIC.
- European Social Survey. 2018b. *ESS-8 2016 Documentation Report*. Edition 2.1. Bergen, European Social Survey Data Archive, NSD—Norwegian Centre for Research Data for ESS ERIC.
- Groves, R.M., and Couper, M.P. 1998. *Nonresponse in Household Interview Surveys*. New York: John Wiley & Sons.
- Groves, R.M., and Kahn, R.L.. 1979. *Surveys by Telephone: A National Comparison with Personal Interviews*. New York: Academic Press.
- Haan, M., Ongena, Y.P., and Aarts, K. 2014. Reaching Hard-to-Survey Populations: Mode Choice and Mode Preference, *Journal of Official Statistics 30* (2): 355–79.
- International Telecommunication Union. 2019. World Telecommunication/ICT Development Report and Database, Retrieved (<https://data.worldbank.org/indicator/IT.NET.USER.ZS>).
- Long, J.S., and Freese, J. 2014. *Regression Models for Nominal and Ordinal Outcomes*. College Station, Texas: Stata Press.
- Medway, R.L., and Fulton, J. 2012. When More Gets You Less: A Meta-Analysis of the Effect of Concurrent Web Options on Mail Survey Response Rates, *Public Opinion Quarterly 76* (4): 733–46.
- Millar, M.M., O’Neill, A.C., and Dillman, D.A. 2009. *Are Mode Preferences Real?* Washington: Social and Economic Sciences research Center.
- Olson, K., Smyth, J.D., and Wood, H.M. 2012. Does Giving People Their Preferred Survey Mode Actually Increase Survey Participation Rates? An Experimental Examination, *Public Opinion Quarterly 76* (4): 611–35.
- Rybak, A. 2018. The Selection of Survey Modes on the Basis of Respondents’ Preferences Analysis, *Ask: Research and Methods 27* (1): 23–39.
- Smyth, J.D., Olson, K., and Kasabian, A. 2014. The Effect of Answering in a Preferred Versus a Non-Preferred Survey Mode on Measurement, *Survey Research Methods 8* (3): 137–52.
- Smyth, J.D., Olson, K., and Millar, M.M. 2014. Identifying Predictors of Survey Mode Preference, *Social Science Research 48*: 135–44.
- Sztabiński, P.B. 2018. Does Mixed Mode Data Collection Improve the Achieved Sample? A Comparison of the ESS 7 and a Mixed Mode Experiment, in: F. Sztabiński, H. Domański, and P.B. Sztabiński (eds.), *New Uncertainties and Anxieties in Europe: Seven Waves of the European Social Survey*. Berlin, Bern, Bruxelles, New York, Oxford, Warszawa, Wien: Peter Lang.

- Sztabiński, P.B., and Sztabiński, F. 2016. Does Mixed Mode Data Collection Improve the Achieved Sample? A Comparison of the ESS PAPI Survey and a Mixed Mode Experiment, Retrieved 22 September 2017 (www.ifispan.pl/wp-content/uploads/2016/03/PBS_FS_Sztabinski_ESS7_dissconf.pdf).
- Tarnai, J., and Paxson, M.C. 2004. Survey Mode Preferences of Business Respondents, *59th Annual Conference, American Association for Public Opinion Research*.
- Tourangeau, R.J., Brick, M., Lohr, S., and Li, J. 2016. Adaptive and Responsive Survey Designs: A Review and Assessment, *Journal of the Royal Statistical Society. Series A: Statistics in Society*.
- Vannieuwenhuyze, J.T. a. 2013. On the Relative Advantage of Mixed-Mode versus Single-Mode Surveys, *Survey Research Methods* 8 (1): 31–42.
- Villard, A., and Fitzgerald, R. 2017. Using Mixed Modes in Survey Research: Evidence from Six Experiments in the ESS, in: M.J. Breen (ed.), *Values and identities in Europe: Evidence from the European Social Survey*. Abingdon: Routledge, pp. 273–310.
- Wardle, H., and Robinson, C. 2007. Choosing Web Surveys: Mode Choices among Youth Cohort Study Respondents, in: M. Trotman, T. Burrell, L. Gerrard, K. Anderton, G. Basi, M. Couper, K. Morris, D. Birks, A. Johnson, R. Baker, M. Rigg, S. Taylor, and A. Westlake (eds.), *The Challenges of a Changing World. Proceedings of the Fifth ASC International Conference*. Southampton: Association for Survey Computing, pp. 193–202.

Biographical Note: Adam Rybak is a Ph.D. student at the Faculty of Sociology, Adam Mickiewicz University in Poznań, Poland. His current research focuses on survey methodology, with using meta analysis of comparative survey projects, in particular.

ORCID iD: <https://orcid.org/0000-0003-3981-1081>

E-mail: adam.rybak@amu.edu.pl