Combining Web and Face-to-Face Survey Modes in the GGS: Evidence from a Two-Country Pilot Study

Giorgio Piccitto (Bocconi University, giorgio.piccitto@unibocconi.it)
Aart C. Liefbroer (Netherlands Interdisciplinary Demographic Institute, University of Groningen & Vrije Universiteit Amsterdam; liefbroer@nidi.nl),

Extended abstract prepared for the European Population Conference, June 24-27, Padua, Italy

Abstract

The decreasing response rates and increasing costs of face-to-face surveys and the increasing internet penetration make it attractive to use web surveying as an addition to or replacement for face-to-face surveys. The Generations and Gender Survey, one of the most widely used resources of comparative research among demographers plans to move to web surveying as its major method in future waves of data collection. This paper uses data from a pilot study conducted in Germany and Croatia to examine the consequences of moving to web for estimates of central variables of interest to demographers and family scholars and for estimates of the strength of associations within theoretical models. Given that social desirability bias is assumed to be lower in web than in face-to-face mode, we expect less social desirable answers to questions that are sensitive to one's self-representation in web than in survey mode. We do not expect large differences between both modes in the strength of associations between variables in multivariate models. These expectations are tested on a range of variables and models of interest to demographers and family scholars.

Introduction

Online research has become the most popular mode of data collection in marketing research and is also becoming increasingly attractive as a tool for survey research. The response rates of face-to-face surveys (F2F from now on) have shown a clear, though not uniform, decrease over time (Beulens et al. 2018). At the same time, the costs of F2F interviews, in particular in countries where labour costs are high, have surged and act as a clear impediment against the use of F2F. Concomitantly, the internet penetration across the globe has increased sharply (Pandita 2017). This makes the use of the web (WEB from now on) to collect data increasing attractive, as people can answer questions whenever they feel ready to and with minimal interference. As result, WEB has become increasingly popular, either in combination with F2F (e.g. to boost response rates), leading to a kind of multi-mode survey design, or as a replacement of F2F. For instance, the new waves of the Generation and Gender Surveys (Gauthier et al. 2018), one of the most frequently used sources for comparative research in demography, plan to collect data by WEB in addition or as a substitution for F2F. However, the consequences of such a move for descriptive information on relevant variables and for the relationship between variables is unclear, as the literature has underlined the existence of differences between on-line survey responses and those collected with others modes (telephone, face-to-face, self-administered) (Currivan et al. 2004; Epstein et al. 2001). Such differences are defined as *mode effects*, a concept which refers to systematic dissimilarity among data collected with different survey modes. One important mechanism that could drive the mode effect is social desirability bias, the mechanism by which respondents over-report 'desirable'

behaviours and under-report 'undesirable' behaviours, especially on questions on sensitive topics (Tourengeau & Yan 2007). The social desirability bias rests on the idea that there are social norms governing a range of attitudes and behaviours, and that people may consciously or unconsciously misrepresent themselves to comply with these norms. In particular, this mechanism is likely to be at play in F2F interviews, with the physical presence of an interviewer, while it would be less biasing in WEB surveys due to their perceived level of anonymity and confidentiality (Aquilino 1994; Kreuter et al. 2008).

This paper aims to shed light on the mode effect of F2F versus WEB administration, by answering two research questions. First, we want to examine whether systematic patterns of differences in answering to sensitive topics exist between the two modes (*RQ1*). We expect that – as a result of social desirability bias – mode differences are larger for questions on family and work issues that are more subjective and central to one's self-representation (e.g. satisfaction with intimate relationships, couple disagreements, loneliness, quality of work relationships) than for questions that are more factual and less central to the representation of the self (e.g. factual information on employment, division of household labour). Second, we want to examine whether systematic differences exist in the association between variables in multivariate models (*RQ2*). This latter research question is particularly important as scholars often are more interested in testing theoretical models about the relationship between scientific concepts than about examining the average scores on concepts within a population. However, most of the empirical research on mode effects has focussed on comparisons of average scores on concepts rather than on comparisons of associations between concepts. We expect no major mode differences in associations. If anything, associations may be somewhat larger in WEB than in F2F mode.

Data and empirical strategy

The data we use derive from an experiment in the context of the Generations and Gender Project (GGP), which is a cross-national social science data infrastructure that collected micro- and macro-level data on demographically relevant topics in over 20 countries (Gauthier et al. 2018). In the past, the Generation and Gender Survey (GGS) – the focal survey instrument of the GGP – has been conducted in CAPI mode only. In 2018, an experimental pilot study was conducted in three countries (Germany, Croatia, Portugal) to study the possibilities of implementing a push-to-web design. Therefore, in these countries questions were answered in three different modes: a) F2F; b) WEB→F2F (start WEB, then those who did not respond were directed to F2F); c) WEB. In Portugal, the overall sampling strategy of the pilot turned out to be problematic, as an individual-level sampling frame was not available and internet penetration was low (Emery et al. 2019). Therefore, we restrict our analysis to data from Germany and Croatia. In Germany, response rates were somewhat lower for WEB than for F2F (23.7% versus 29.5%), whereas the reverse was true for Croatia (49.5% versus 27.7%) (Emery et al. 2019). In Germany, 1,816 interviews were conducted, of which 1,492 (82.2%) WEB, 192 (10.6%) F2F, and 132 (7.3%) WEB→F2F. In Croatia, 1,511 interviews were conducted, of which 1,296 (85.8%) WEB, 149 (9.9%) F2F, and 66 (4.4%) WEB→F2F.

For the analysis, we select a set of survey items and scales in the family and work domains that vary in the extent to which they are expected to be susceptible to social desirability bias. To answer our first research question on mode differences in the average scores, we run OLS and logistic

regression analyses in which item and scales scores were regressed on mode, with a number of additional control variables (age, gender, partner status, educational attainment, country). Additionally, we include an interaction between mode and country, to examine whether any mode differences showed comparative patterns across both countries. For our second research question, we run OLS and logistic regressions with substantively important variables, including interactions between theoretical important variables and mode to test whether the strength of substantively important relationships varies by mode.

Selected preliminary results

In this section, we present preliminary results for a set variables of interest to many demographers and family scholars. As an example of our analysis of the first research question, we present results on how average scores on a measure of loneliness vary by mode (*M1*). As an example of our analysis of the second research question, we present results on how the relationships between partner status, gender and age on the one hand and loneliness on the other vary by mode (*M2*). In the full paper, a much broader selection of variables and models will be tested. In the example, the focus is on the simple comparison of those who answered the questionnaire by WEB and those who did so by F2F. In the full paper, we will also include a distinction between those who were randomly selected into F2F and those who answered F2F after not responding in WEB.

Mode differences in levels of loneliness

The first research question examines whether mode effects on the average score of key concepts are observed. This is examined for loneliness. Loneliness refers to an experienced gap between the quality and quantity of relationships that one has and the quality and quantity of relationships that one aspires to have. It has been shown to be reliably and validly measured by a six-item instrument developed by De Jong Gierveld and Van Tilburg (2006). These items are included in the GGS and overall scores on the loneliness scale (running from 0 to 6) are calculated according to the guidelines provided by De Jong Gierveld and Van Tilburg (2006). In Model 1 (*M1*) we run an OLS regression with loneliness score as dependent variable and mode of survey administration (WEB or F2F) as main independent, including a range of socio-demographic controls. Then, we interact all the control variables with the variable mode of survey administration, in order to examine whether the relationship between loneliness and the included socio-demographic covariates varies by mode (*M2*). Table 1 shows the results of the two estimated model.

Table 1. OLS estimation results. Model 1 and Model 2.

Loneliness score				
	M1		M2	
	Coeff.	[95% Conf. Interval]	Coeff.	[95% Conf. Interval]
FTF mode	-0.75***	[-0.92 - -0.58]	-1.27***	[-2.22 - -0.32]
Age	0.02***	[0.01 - 0.03]	0.01***	[0.00 - 0.02]
Female	-0.14**	[-0.27 - -0.00]	-0.18**	[-0.32 - -0.03]
Without a partner	0.53***	[0.37 - 0.70]	0.60***	[0.42 - 0.78]
Education [ref.: Low]				
Middle	-0.00	[-0.28 - 0.28]	0.12	[-0.20 - 0.45]
Low	-0.24	[-0.52 - 0.05]	-0.22	[-0.55 - 0.11]
Croatia	0.37***	[0.23 - 0.50]	0.37***	[0.22 - 0.51]
FTF*Age			0.02**	[0.01 - 0.04]
FTF*Female			0.10	[-0.24 - 0.45]
FTF*Without a partner			-0.33	[-0.75 - 0.09]
*FTF*Education [ref.: Low]*				
FTF*Middle			-0.57*	[-1.20 - 0.06]
FTF*Low			0.03	[-0.63 - 0.69]
FTF*Croatia			-0.03	[-0.37 - 0.32]
Constant	1.61***	[1.20 - 2.02]	1.71***	[1.24 - 2.18]
Observations	2,392	2,392	2,392	2,392
R-squared	0.07	0.07	0.08	0.08
ci in parentheses; *** p<0.01, ** p<0.05, * p<0.1				

From Model 1 clearly emerges that the score on the loneliness scale is strongly influenced by whether or not respondents answered the loneliness items in WEB or F2F mode. Those who answered in F2F mode scored an average of .75 lower on the loneliness scale (that theoretically runs from 0 to 6). This is a very large difference, suggesting that respondents in F2F mode are much less likely to state that they experience loneliness than respondents in WEB mode. Regarding the covariates, loneliness scores increase with age, women are slightly less lonely than men, those without a partner are lonelier than those with a partner and respondents in Croatia are more lonely than German ones. No educational differences in loneliness are found. It is worthy to stress the stronger magnitude of the mode coefficient than the one for having a partner: this suggests the relevance of the mode effect in measuring loneliness, even higher than the presence of a partner (usually the most important determinant of feelings of loneliness).

Looking at Model 2, it comes up that the only statistically significant mode effect on the strength of relationships between socio-demographic characteristics and loneliness is for age. The effect for age is stronger in F2F mode than in WEB mode. In WEB mode, loneliness scores increase by .013 per additional year of age, whereas in F2F mode the increase is .013+.025=.038 per year. All other interactions are not statistically significant suggesting that the relationship between loneliness on the one hand and gender, partner status, level of education and country on the other does not differ on the bases of survey administration.

Discussion

In the full paper, the set of variables tested will be expanded to include other key ones in the family and work domains. Furthermore, more refined statistical techniques (such as the Multitrait

multimethod SEM approach). Based on the results, we will draw conclusions about the extent to which mode effects are visible in the average scores on key variables of interest and in key associations of interest.

Next, we will discuss ways of dealing with mode differences in comparative research. Given the potentially large effects of mode in substantively interesting variables and models, we suggest that – if a multi-mode approach is adopted - it is preferable to include both WEB and F2F in each country. At the very minimum, it is necessary to have one type of mode (e.g. WEB) included in the mode mix of each and every country. In that way, substantive country differences and mode effects can be separated in multivariate analyses and mean scores per country can be adjusted to account for differences in the mode mix between countries.

References

Aquilino, W. S. (1994). Interview Mode Effects in Surveys of Drug and Alcohol Use. *Public Opinion Quarterly, 58*(2), 210-240.

Beullens, K., Loosveldt, G., Vandenplas, C., & Stoop, I. (2018). Response Rates in the European Social Survey: Increasing, Decreasing, or a Matter of Fieldwork Efforts? *Survey Methods: Insights from the Field*. DOI:10.13094/SMIF-2018-00003

Currivan, D. B., Nyman, A. L., Turner, C. F., & Biener, L. (2004). Does Telephone Audio Computer-assisted Self-Interviewing Improve the Accuracy of Prevalence Estimates of Youth Smoking? Evidence from the UMass Tobacco Study. *Public Opinion Quarterly, 68*(4), 542-564.

De Jong Gierveld, J., & Van Tilburg, T.G. (2006). A 6-Item Scale for Overall, Emotional, and Social Loneliness: Confirmatory Tests on Survey Data. *Research on Aging, 28*(5), 582-598.

Emery, T. et al. (2019). *Report on the Results of the GGP Experiment on Modes*. The Hague: Generations and Gender Programme: Evaluate, Plan, Initiate.

Epstein, J. F., Barker, P. R., & Kroutil, L. A. (2001). Mode Effects in Self-Reported Mental Health Data. *Public Opinion Quarterly, 65*(4), 529-549.

Gauthier, A. H., Cabaço, S. L. F., & Emery, T. (2018) Generations and Gender Survey Study Profile. *Longitudinal and Life Course Studies, 9*(4), 456-465.

Kreuter, F., Presser, S, & Tourangeau, R. (2008). Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity. *Public Opinion Quarterly, 72*(5), 847-865

Pandita, R.(2017). Internet a Change Agent: An Overview of Internet Penetration and Growth across the World. *International Journal of Information Dissemination and Technology, 7*(2), 83-91.

Tourangeau, R., & Yan, T. (2007). Sensitive Questions in Surveys. *Psychological Bulletin, 133*, 859-883.