The association between survey characteristics and representativeness: A meta-analysis of two common approaches

Carina Cornesse, SFB 884 “Political Economy of Reforms”, University of Mannheim and GESIS – Leibniz Institute for the Social Sciences
Michael Bosnjak, GESIS – Leibniz Institute for the Social Sciences and School of Social Sciences, University of Mannheim

Abstract

In light of decreasing response rates, new technological as well as survey statistical developments, and the rise of commercial nonprobability online panels, the social scientific research community faces new challenges when deciding how to collect survey data and when to trust the results. In an ongoing project on modeling the association between survey design features and representativeness, we identify, synthesize, and analyze the existing literature on survey representativeness using meta-analytic techniques. So far, our findings indicate that mixed-mode surveys, surveys with high response rates, and probability surveys have higher degrees of representativeness than single-mode surveys, surveys with low response rates, and nonprobability surveys. Furthermore, web surveys and personal interviews tend to be less representative than surveys that use any other common survey mode.

Introduction

In this paper, we conduct meta-analyses on two highly established measures of survey representativeness (R-Indicators and benchmark comparisons). In the meta-analyses, we compare a large number of surveys and rank them according to their reported degrees of representativeness. We then regress representativeness on a number of survey design characteristics that are assumed to influence survey representativeness. Using these advanced meta-analytic techniques, we seek to provide evidence about which survey characteristics make for a highly representative survey. Furthermore, we assess the robustness of our results by comparing findings across measures and look into how the statistical properties of the measures influence the results.

From the existing literature on representativeness in general and the individual error sources that the concept includes (coverage error, sampling error, and nonresponse error) we derive a number of hypotheses that we test using meta-analytic techniques:

H1: Face-to-face surveys are more representative than surveys in any other mode.

H2: Web surveys are less representative than surveys in any other mode.

H3: Mixed-mode surveys are more representative than single-mode surveys.

H4: The higher the response rate, the higher is the degree of representativeness.

H5: Probabilistic surveys are more representative than nonprobability surveys.

H6: Large numbers of variables in the representativeness models lead to low measured degrees of representativeness.

H7: The higher the degree of representativeness on one measure, the higher is the degree of representativeness on another measure.

Data

We conduct two separate meta-analyses in this paper using different effect sizes. Effect sizes are treated as dependent variables in the meta-analysis. The first meta-analysis uses R-Indicators as an effect size. The second meta-analysis uses aggregated percentage point differences between a survey and a benchmark as an effect size. In our meta-regression models, we use the identical set of moderators on both effect sizes except for select variables that are only available for either R-Indicators or benchmark comparisons. Additionally, we use the same types of meta-regression models in our analyses on both effect sizes. For all of our analyses, we use the *metafor* package in R (see Viechtbauer 2010).

The effect sizes

R-Indicators

With regard to R-Indicators as a measure of survey representativeness, we include in our meta-analyses all articles (journal articles, book chapters, published working papers) that contain general sample-based R-Indicators for a survey. R-Indicators are measures of survey representativeness that are based on logistic regression models for the propensity to respond to a survey. The independent variables in the regression models are auxiliary variables that need to be available for both respondents and nonrespondents to the survey. The individual response propensities from the regression models are then aggregated. In the last step of computing R-Indicators, the aggregated results from the propensity models are rescaled to range between zero (not representative) and one (very representative) using the formula

$$R(\rho) = 1 - 2S(\rho) \qquad (3)$$

where $S(\rho)$ is the standard deviation of the average response propensity (see Schouten, Cobben, and Bethlehem 2009 for a detailed description of R-Indicators).

Benchmark comparisons

Benchmark comparisons are comparisons of proportions of a variable in a survey and a benchmark. We compute absolute percentage point differences between the survey and the benchmark for each variable of a study with

$$D_{benchmark,c} = \left| \left(\frac{n_{R_C}}{n_R}\right) - \left(\frac{N_C}{N}\right) \right| \qquad (5)$$

where n_R is the number of respondents, N is the benchmark estimate, and C stands for the category of a variable.

The raw estimates of the survey and benchmark for our computations are usually reported in tables so that $D_{benchmark,c}$ can easily be computed. To aggregate the results, we compute the median $\widetilde{D}_{benchmark}$ of these individual percentage point differences across all categories and variables in the study. $\widetilde{D}_{benchmark}$ is the effect size we use in our meta-analysis.

Moderators

Moderators are the independent variables in our meta-regression models and operationalize the expectations described above. In this paper, we use the following moderators:

1) Survey mode
2) Survey response rate
3) Probability versus nonprobability samples
4) Number of variables used to measure representativeness

Method

Meta-analyses synthesize the results of individual studies and analyze this data using statistical procedures (see Cooper 2010). They can be used to integrate and evaluate a larger number of empirical studies and can thereby contribute to the development of scientific discussions and research fields as a whole. The advantages of meta-analyses include that they can answer questions that no single study can answer individually. By this, they can settle controversies in a field. Moderators play a crucial part in most meta-analyses as independent variables that can be used to test hypotheses and model associations. Commonly, they are coded to operationalize a hypothesis of the association with the effect size, which is the dependent variable in the analysis.

Results

We divide this section into two parts. In the first part, we display the R-Indicators results and in the second part the benchmark comparisons results.

R-Indicators

In this paper, we include 22 papers on R-Indicators with 111 individual effect sizes. Generally, we find a mean R-Indicator of 0.83 with a confidence interval of 0.82 to 0.85. The individual R-Indicators range from 0.63 in a working paper by Ariel and Schouten in 2008 to 0.98 in a working paper by deNooij in 2008. This is a relatively narrow value range considering that R-Indicators can theoretically range between 0 and 1. Nevertheless, the Q-test for heterogeneity is statistically highly significant, which means that there is variability across the R-Indicators worth exploring.

Table 1: Mixed effects meta-regression models of R-Indicators with single moderators (and std. errors)

	Model 1	**Model 2**	**Model 3**	**Model 4**
Intercept	0.85***	0.82***	0.77***	0.84***
	(0.01)	*(0.01)*	*(0.00)*	*(0.02)*
Personal interview	-0.04**			
(vs. other modes)	*(0.01)*			
Mixed mode		0.04***		
(vs. other modes)		*(0.01)*		
Response rate			0.0009*	
			(0.00)	
Number of auxiliary variables				-0.00
				(0.01)
k	97	97	111	111
R^2	7.58%	12.76%	3.42%	0.00%
I^2	98.75%	98.64%	99.16%	99.19%

*** p<0.001 **p<0.01 *p<0.05

Table 1 shows the results of our mixed-effects meta-regression analyses on R-Indicators. It only includes moderators with sufficient numbers of cases. That is why there are models using personal interview and mixed mode as moderators, but no models using the web survey mode. Furthermore, there are models using response rate and number of auxiliary variables as moderators, but no models with probability vs. nonprobability surveys. In the first model (Model 1) we regress the R-Indicators on personal interviews. The number of cases included in the analysis is 97, because there are some missing values in the data and modes that did not fit into any of the mode categories. The R^2 for the model is 7.58% and the I^2 is 98.75%. The intercept of Model 1 is 0.85, highly significant, and has a standard error of 0.01. The coefficient of the personal interview moderator is -0.04, significant, and has a standard error of 0.01. In the second model (Model 2) we regress R-Indicators on a binary mixed mode variable. The number of effect sizes is again 97. The R^2 is 12.76% and the I^2 is 98.64%. The intercept of Model 3 is 0.82, highly significant, and has a standard error of 0.01. The mixed mode coefficient is 0.04, highly significant, and has a standard error of 0.01. The third model (Model 3) regresses R-Indicators on a continuous response rate variable. The number of cases in the analysis is 111, which means that all the studies in our data set are included in the model. The R^2 is 3.42 and the I^2 is 99.16%. The model intercept is 0.77, highly significant, and has a standard error of 0.00. The response rate coefficient is 0.0009, weakly significant, and has a standard error of 0.00. The last model in Table 1 is Model 4. It shows results of a regression of R-Indicators on the number of auxiliary variables that were used in the logistic propensity model underlying the R-Indicator calculation. The number of cases is again 111. The R^2 is 0.00 and the I^2 is 99.19%. The model coefficient is 0.84, highly significant, and has a standard error of 0.02. The coefficient of the number of auxiliary variables is -0.00, insignificant, and has a standard error of 0.01.

Benchmark comparisons

In this subsection, we show the results for the analyses on the benchmark comparisons as an effect size. We include 34 studies reported in 19 papers in the analysis. The mean deviation of a set of characteristics between a survey and a benchmark is 3.97 percentage points with a confidence interval from 2.69 to 5.25 percentage points. The largest median deviation of a survey compared to a benchmark across a set of variables included in our paper is 17.90 percentage points in a study by Bandilla, Bosnjak, and Altdorfer (2003).There are six studies with a median misrepresentation of less than 1% and all but three studies have a median misrepresentation of less than 10%. Furthermore, all but three effect sizes have standard errors of less than 0.15.

Table 2: Mixed effects meta-regression models of percentage point differences with single moderators (and std. errors)

	Model 5	**Model 6**	**Model 7**	**Model 8**
Intercept	1.81*	*3.85****	3.23***	6.15***
	(0.81)	*(0.72)*	*(0.86)*	*(1.33)*
Web survey (vs. other modes)	4.08***			
	(1.11)			
Response rate		-0.05*		
		(0.02)		
Nonprobability sample			1.67***	
(vs. probability-based sample)			*(1.30)*	
Number of variables compared				-0.12
				(0.06)
k	34	34	34	34
R^2	27.47%	0.00%	1.99%	6.91%
I^2	99.97%	99.98%	99.98%	99.98%

*** p<0.001 **p<0.01 *p<0.05

Table 2 shows the results for the mixed-effects meta-regressions on median percentage point deviations of a survey to a benchmark as an effect size. For some moderators, the number of cases per category is insufficient. Therefore, we include a binary variable on web surveys, a continuous variable on the response rate, a binary variable on probability versus nonprobability surveys, and a continuous variable on the number of variables that is compared between a survey and a benchmark per study. There are no missing values on any of the moderators. Therefore, the number of cases is 34 in all models. The first model (Model 5) regresses the percentage point deviations on web surveys versus other modes. The R^2 of the model is 27.47% and the I^2 is 99.97%. The model intercept is 1.81, weakly significant, and has a standard error of 0.81. The web survey intercept is 4.08, highly significant and has a standard error of 1.11. In the second model (Model 6) we regress median percentage point deviations on response rates. The R^2 is 10.81% and the I^2 is 99.98%. The model intercept is 6.55, highly significant, and has a standard error of 1.31. The response rate coefficient is -0.05, weakly significant, and has a standard error of 0.02. For the third model (Model 7) we regress percentage point deviations on nonprobability samples. The R^2 is 1.99% and the I^2 is 99.98%. The model intercept is 3.23, highly significant, and has a standard error of 0.86. The moderator coefficient is 1.67, highly significant, and has a standard error of 1.30. In the last model (Model 8) we regress percentage point deviations on the number of variables that is compared between a survey and a benchmark in each study. The R^2 is 6.91% and the I^2 is again 99.98%. The model intercept is 6.15, highly significant, and has a standard error of 1.33. The moderator coefficient is -0.12, insignificant, and has a standard error of 0.06.

Conclusion

In this section of our paper we revisit each of our expectations from above and interpret our empirical findings. Our first hypothesis states that face-to-face surveys are more representative than surveys in other modes. We find evidence against this in our R-Indicator sample, where the mixed-effects meta-regression shows a significant negative association between personal interviews and the degree of representativeness (see Model 1 in Table 1). We do not find the same result in the meta-analysis on benchmark comparisons, because the number of cases of personal interviews in our data set is insufficient for a mixed-effects meta-regression. The second expectation that we formulate is that web surveys are less representative than surveys with other modes of data collection. In our R-Indicator data set, the number of web survey studies is insufficient to test the hypothesis. Model 5 in Table 2, however, supports the hypothesis for benchmark comparisons. Our third expectation states that mixed-mode surveys are more representative than single-mode surveys. Model 2 in Table 1 supports this hypothesis for the R-Indicators. There is a highly significant effect of mixed-mode on the degree of representativeness as measured using R-Indicators. Due to insufficient numbers of observations, we cannot test the hypothesis for benchmark comparisons. Our fourth expectation is that there is a positive linear relationship between the response rate and the degrees of representativeness on R-Indicators and benchmark comparisons. We find weakly significant effects in favor of the expectation both in the R-Indicator (Model 3 in Table 1) and benchmark comparison (Model 6 in Table 2) analysis. Expectation 5 is that probabilistic surveys are more representative than nonprobability surveys. We cannot test this hypothesis for the R-Indicator studies because there are no R-Indicators for nonprobability surveys. Model 7 in Table 2 shows that nonprobability surveys, indeed, have significantly larger median percentage point deviations between a survey and a benchmark. Therefore, the data support our hypothesis. Our last expectation is that results of the meta-analyses should be consistent across measures for representativeness. We find that this is mostly true for all the moderators that we could include in both analyses. The effects for response rates and the number of variables included in the computation of the effect sizes are robust across effect sizes. Additionally, results for the mode hypotheses are not contradictory across R-Indicators and benchmark comparisons. However, due to insufficient numbers of cases per category, some of the hypotheses could only be tested for one of the measures, so that we cannot assess the robustness of the findings across measures. Therefore, the question remains whether associations between mode and sample type with the degree of representativeness are consistent across R-Indicators and benchmark comparisons.

Literature

Ariel, A., & Schouten, B. (2008). Representativity of the Time Use Survey. *Discussion Paper (08013)*, Statistics Netherlands.

Bandilla, W., Bosnjak, M., & Altdorfer, P. (2003). Survey Administration Effects?: A Comparison of Web-Based and Traditional Written Self-Administered Surveys Using the ISSP Environment Module. *Social Science Computer Review*, 21(2), 235–243.

Cooper, H. M. (2010). *Research synthesis and meta-analysis: a step-by-step approach (4th ed)*. Los Angeles: Sage.

De Nooij, G. (2008). *Representativity of Short Term Statistics.* Statistics Netherlands.

Schouten, B., Cobben, F., & Bethlehem, J. G. (2009). Indicators for the representativeness of survey response. *Survey Methodology*, 35(1), 101–113.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *J Stat Softw*, 36(3), 1–48.