Modeling panel representativeness comparatively

Carina Cornesse, University of Mannheim and GESIS Leibniz Institute for the Social Sciences
Tobias Enderle, Leibniz Institute for the Social Sciences

In this paper, we use data from the German Internet Panel (GIP) and GESIS Panel to comparatively assess panel representativeness in the face-to-face recruitment stage of two probability-based online panels using the sample-based R-Indicator proposed by Schouten, Cobben, and Bethlehem (2009). Results show, that the R-Indicator for the GIP recruitment interview is significantly higher than the R-Indicator for the GESIS Panel. This indicates that the two panels' different sampling and fieldwork strategies have an effect on their survey representativeness. However, we are skeptical about the validity of these preliminary results because explanatory power and quality of the auxiliary variables underlying our R-Indicator calculations are questionable.

Probability-based online panels of the general population differ widely in their sampling and panel recruitment procedures, as well as the treatment of units of analysis without Internet access. Because these panel studies are so different, it is hard to assess their success in achieving representative survey data relative to one another (see e.g. Blom et al. 2015). In order to evaluate the success of different sampling and fieldwork procedures to generate representative survey data, being able to make cross-panel comparisons is, however, desirable. This is especially true for multi-thematic panel studies such as the GIP and GESIS Panel, where it is important to achieve a high degree of general representativeness in order to provide social scientists in different research fields with high quality survey data.

The goal of this research paper is to assess whether there are differences between the GIP and GESIS Panel with regard to the representativeness in their recruitment interviews that could be attributable to their different sampling and fieldwork procedures. If we find such differences, this would indicate that one panel is more successful than the other one in generating a representative panel response set.

Measuring representativeness using R-Indicators

In our analysis, we use the sample-based R-Indicator proposed by Schouten, Cobben, and Bethlehem (2009). It has the advantage of providing a single measure for survey representativeness that ranges between 0 and 1. Results should, therefore, be relatively easy to interpret and to compare across panels. The R-Indicator also allows for the inclusion of auxiliary variables, such as sampling frame information and paradata, which have been shown to have explanatory power in nonresponse bias analyses (Sinibaldi, Trappmann, and Kreuter 2014).

R-Indicators assess whether distributions of a set of auxiliary variables differ between the actual response set of a survey and its gross sample including respondents as well as nonrespondents. These indicators are based on logistic regression models for the propensity of a sample member to respond to a survey. The auxiliary variables are the independent variables in those response propensity models.

For the calculation of R-Indicators, the propensity models are used to investigate differences between respondents and nonrespondents on each of the included auxiliary variables. These propensities are then aggregated and re-scaled to range between 0 and 1 for better interpretation. If respondents differ a lot from nonrespondents on the set of auxiliary variables, then the R-Indicator is lower and if there is no such systematic difference between respondents and nonrespondents the R-Indicator is close to 1.

We choose sample-based R-Indicators in this research paper for two reasons: Firstly, we aim to investigate whether the GIP and GESIS Panel achieve representative response sets for their recruitment interviews on the analyzed auxiliary variables and R-Indicators should give us a good overview on this. Secondly, R-Indicators make it possible to assess cross-panel differences in the degree of representativeness on the auxiliary variables. Therefore, R-Indicators can show us the success of the GIP and GESIS Panel in achieving representative response sets relative to one another.

GIP and GESIS Panel

In this paper, we use data from the GIP and GESIS Panel in order to comparatively assess panel representativeness in the face-to-face recruitment stage of two probability-based online panels. The GIP has an area probability sample whereas the GESIS Panel has a register-based sample (see Blom, Gathmann, and Krieger 2015, and GESIS Panel 2015 for further information). The two panels have similar fieldwork approaches for their recruitment interviews, for instance with regard to the survey mode, fieldwork agency, contact strategy, and collection of paradata. Apart from the similarities, they differ in a number of aspects that might be relevant for achieving a representative response set, for example the incentives used (see Blom et al. 2015).

A major difference between the GIP and GESIS Panel, that could be relevant in their fieldwork processes, is the selection of the interviewed individuals within a household. The GESIS Panel draws a register-based sample of individuals and only these preselected individuals are interviewed. The GIP has an area sample where individuals are nested within households. The consequence is that in the GIP whoever opens the front door of a sampled address can be interviewed in the recruitment interview. During that interview, the respondent provides information on all other household

members. All household members between 16 and 75 years of age that consent to be recontacted are subsequently invited to the first online panel wave. This difference in respondent selection for the recruitment interview can cause differences between the two panels if one selection procedure is more successful than the other in generating a representative response set. For the recruitment interview, it might for instance be easier to contact anyone within a household who can provide information on all other household members than to contact a specific individual that might not be home a lot.

In this paper, we investigate whether at all the differences in sampling and fieldwork procedures in the GIP and GESIS Panel cause differences in the representativeness of their respondents with regard to their gross samples. If we find such differences, future research will be needed to explain which specific differences between the panels cause the (mis-)representation.

For the analysis reported in this paper, we have the nondirectional hypothesis that the measured degree of representativeness should vary between the two panels because they apply different respondent selection and fieldwork strategies that should be unequally successful in achieving a representative response set. To measure the success in generating representative survey data, we use auxiliary variables as independent variables in our R-Indicator calculations.

Auxiliary variables are all variables recorded for both respondents and nonrespondents before, during or after the two panels' recruitment interviews. In our analysis, they come from two different sources: One data source is the sampling frame and another is the call records. The sampling frame data used in our analysis are (1) geographic region and (2) degree of urbanity of the address of a potential respondent. Call records include (1) number of contact attempts by the interviewer, (2) type of building, (3) building condition, and (4) social class of the potential respondent estimated by the interviewer. Since the set of analyzed auxiliary variables needs to be identical for both panels in order to make cross-panel comparisons possible, we could not include any more variables in our analysis.

Representativeness in the panel recruitment interviews

Figure 1 shows the R-Indicator values and their respective 95% confidence intervals for the GESIS Panel and GIP. For the GESIS Panel, the R-Indicator is 0.83 with a confidence interval ranging from 0.81 to 0.84. The GIP has an R-Indicator of 0.87 with a confidence interval from 0.85 to 0.89. The propensity models underlying the R-Indicator calculations can be found in Table 1 in the appendix.

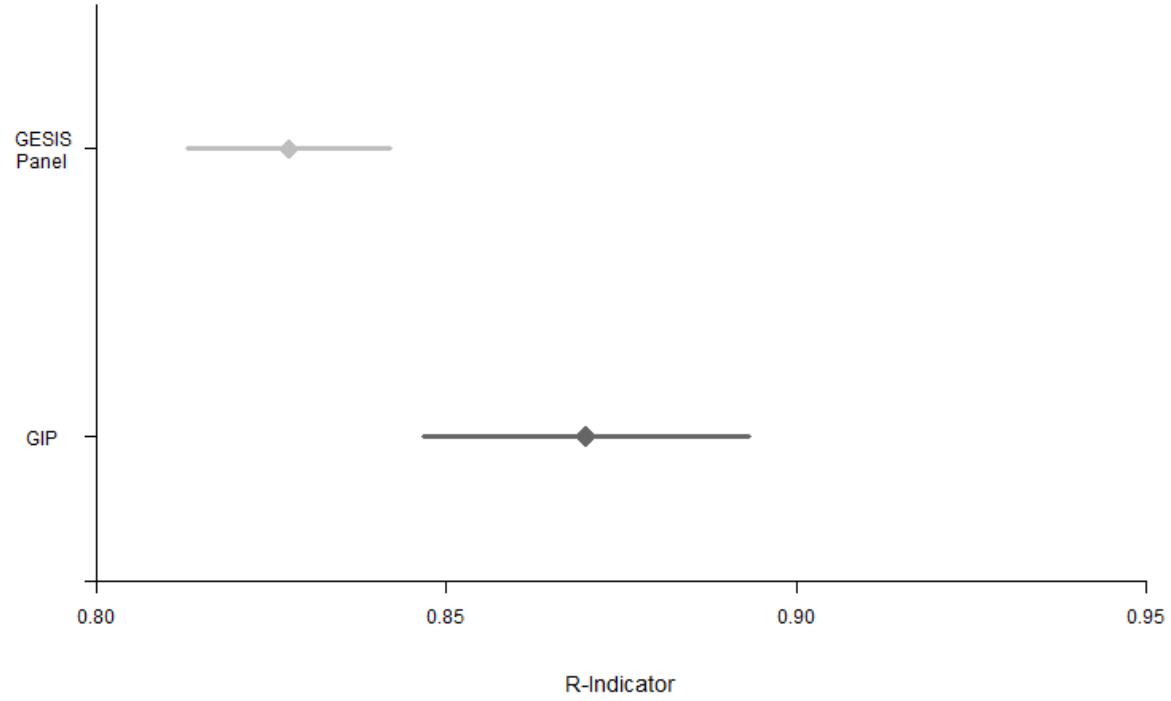

Figure 1: R-Indicators for GESIS Panel and GIP recruitment interviews (including 95% confidence intervals)

Since the confidence intervals of the R-Indicators for the GESIS Panel and GIP do not overlap, we conclude that the difference between the two panels is significant. The GIP generally seems to have a more representative response set with regard to the recruitment interview than the GESIS Panel. One possible explanation is that the GIP respondent selection procedure is more successful in contacting households where the inhabitants have unequal contact probabilities than the GESIS Panel approach is to contact a specific household member. But since there are more differences in the GIP and GESIS Panel recruitment, we cannot be sure that differences between the panels can be attributed to the different respondent selection procedures.

Discussion and future research

Preliminary results suggest that, based on their gross samples, the GIP recruitment interview response set is more representative than the GESIS Panel response set. This indicates that the GIP is more successful in generating representative survey data, at least in its panel recruitment stage. However, we question the validity of these results with regard to a number of aspects. Some of them are conceptual and concern the definition and operationalization of representativeness underlying the R-Indicators. Other concerns about the validity of our preliminary results are more practical with regard the limitations of the data we have for the analysis.

Conceptually, we have the problem that the R-Indicators do not take into account that already the gross samples could be biased. If, for instance, the coverage and sampling bias is higher for the GIP

than for the GESIS Panel, the GIP's higher R-Indicator values could just be an artifact caused by the lower representativeness of its gross sample with regard to the target population. Our future research on these cross-panel comparisons will have to take this possibility into account, for instance by comparing survey response sets and gross samples to external benchmarks. Such a benchmark can, for example, be a census that is supposed to contain correct variable distributions representative for the two panels' target population.

From a practical perspective we have identified two major problems for the validity of our preliminary results that need to be solved: Firstly, we only have a very limited set of auxiliary variables that we can include in the analysis. This is because we can only use data that are recorded identically for both panels and are available for respondents as well as nonrespondents. Additionally, the auxiliary variables used in our analysis might not have an influence on the variables that are of interest to the researchers using the survey data. Therefore, we are concerned that even if the GIP is more representative based on this set of auxiliary variables, this might not have an effect on any research question that is investigated by the researchers using the data. This is because if the variables of interest to the researchers are not related to the misrepresentation of the auxiliary variables, it does not have an influence on their analyses. A possibility for future research is to link identical sets of micro-geographic area data to the already available data of both panels in order to increase the amount of potentially relevant auxiliary variables.

Our second practical problem is that we find a high amount of systematically missing values in the interviewer observations in the call records (e.g., around 21% missing values on the social class evaluation for both panels). If the auxiliary variables themselves are biased, we cannot trust the validity of the results from the representativeness measure that is based on them. We have included a missing value category as an additional independent variable in the propensity model displayed in Table 2 in the appendix. The effect on the R-Indicators based on these propensity models can be seen in Figure 2 in the appendix. Both panels' R-Indicator values decrease relative to the model with listwise deletion of all cases with missing values in Figure 1. For the GESIS Panel, the drop in representativeness (from 0.83 to 0.65) is more severe than for the GIP (from 0.87 to 0.78). This suggests that in the GESIS Panel missingness on the social class evaluation is more systematic than for the GIP. In the future, we need to investigate why there are so many missing values on the interviewer observations. If we are able to model missingness appropriately data imputation procedures could be used to eliminate the missing values on the interviewer observations.

On the whole, since cross-panel representativeness comparisons are rare in the existing literature, there is still a lot of room for research.

Literature

Blom, A. G., Bosnjak, M., Cornilleau, A., Cousteaux, A.-S., Das, M., Douhou, S., & Krieger, U. (2015). A Comparison of Four Probability-Based Online and Mixed-Mode Panels in Europe. *Social Science Computer Review*.

Blom, A. G., Gathmann, C., & Krieger, U. (2015). Setting Up an Online Panel Representative of the General Population: The German Internet Panel. *Field Methods*, 27(4).

GESIS Panel (2015). Sample and Recruitment. http://www.gesis.org/en/services/data-collection/gesis-panel/sample-and-recruitment/

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

Sinibaldi, J., Trappmann, M., & Kreuter, F. (2014). Which Is the Better Investment for Nonresponse Adjustment: Purchasing Commercial Auxiliary Data or Collecting Interviewer Observations? *Public Opinion Quarterly*, 78(2), 440–473.

Appendix

Table 1: Participation in Recruitment Interview, GIP households 2012 and GESIS Panel individuals 2013, Logistic Regression (standard errors in parentheses, listwise deletion of missing values)

	GIP	GESIS Panel
Constant	-0.28	-0.48***
	(0.16)	(0.08)
Region_east *(ref.: west)*	0.18*	0.03
	(0.09)	(0.04)
Med_sized *(ref.: small town)*	-0.05	-0.03
	(0.08)	(0.04)
Large_City *(ref.: small town)*	-0.12	-0.32***
	(0.11)	(0.05)
Med_sized *(ref.: small town)*	-0.14	-0.21***
	(0.08)	(0.04)
Large_City *(ref.: small town)*	-0.13	-0.25***
	(0.10)	(0.05)
One_contact *(ref. more contact attempts)*	0.22**	0.46***
	(0.07)	(0.03)
Intercom *(ref. no intercom)*	-0.08	-0.01
	(0.07)	(0.03)
Single_Housing_Unit *(ref. more units)*	0.10	0.03
	(0.08)	(0.04)
Condition_ok *(ref. bad condition)*	0.30*	0.07
	(0.14)	(0.08)
Condition_good *(ref. bad condition)*	0.21	0.19*
	(0.17)	(0.08)
Middle_Class *(ref.: lower class)*	0.16*	0.24***
	(0.07)	(0.04)
Upper_Class *(ref.: lower class)*	0.50***	0.37***
	(0.14)	(0.06)
Pseudo R^2	0.01	0.04
R-Indicator	0.87	0.83
CI R-Indicator	(0.85|0.89)	(0.81|0.84)
N	4001	16890

Table 2: Participation in Recruitment Interview, GIP households 2012 and GESIS Panel individuals 2013, Logistic Regression (standard errors in parentheses, missing value category included)

	GIP	GESIS Panel
Constant	-0.22	-0.42***
	(0.16)	(0.08)
Region_east *(ref.: west)*	0.10	0.05
	(0.08)	(0.04)
Med_sized *(ref.: small town)*	-0.07	-0.04
	(0.08)	(0.04)
Large_City *(ref.: small town)*	-0.14	-0.33***
	(0.11)	(0.05)
Med_sized *(ref.: small town)*	-0.17*	-0.21***
	(0.08)	(0.04)
Large_City *(ref.: small town)*	-0.11	-0.27***
	(0.10)	(0.05)
One_contact *(ref. more contact attempts)*	0.17*	0.42***
	(0.07)	(0.03)
Intercom *(ref. no intercom)*	-0.09	-0.00
	(0.07)	(0.03)
Single_Housing_Unit *(ref. more units)*	0.11	0.03
	(0.08)	(0.04)
Condition_ok *(ref. bad condition)*	0.30*	0.02
	(0.14)	(0.08)
Condition_good *(ref. bad condition)*	0.18	0.14
	(0.16)	(0.08)
Middle_Class *(ref.: lower class)*	0.15*	0.25***
	(0.07)	(0.04)
Upper_Class *(ref.: lower class)*	0.50***	0.38***
	(0.14)	(0.06)
Missing_Class *(ref.: lower class)*	-1.35***	-3.29***
	(0.12)	(0.11)
Pseudo R^2	0.04	0.13
R-Indicator	0.78	0.65
CI R-Indicator	(0.75|0.80)	(0.64|0.66)
N	4542	20508

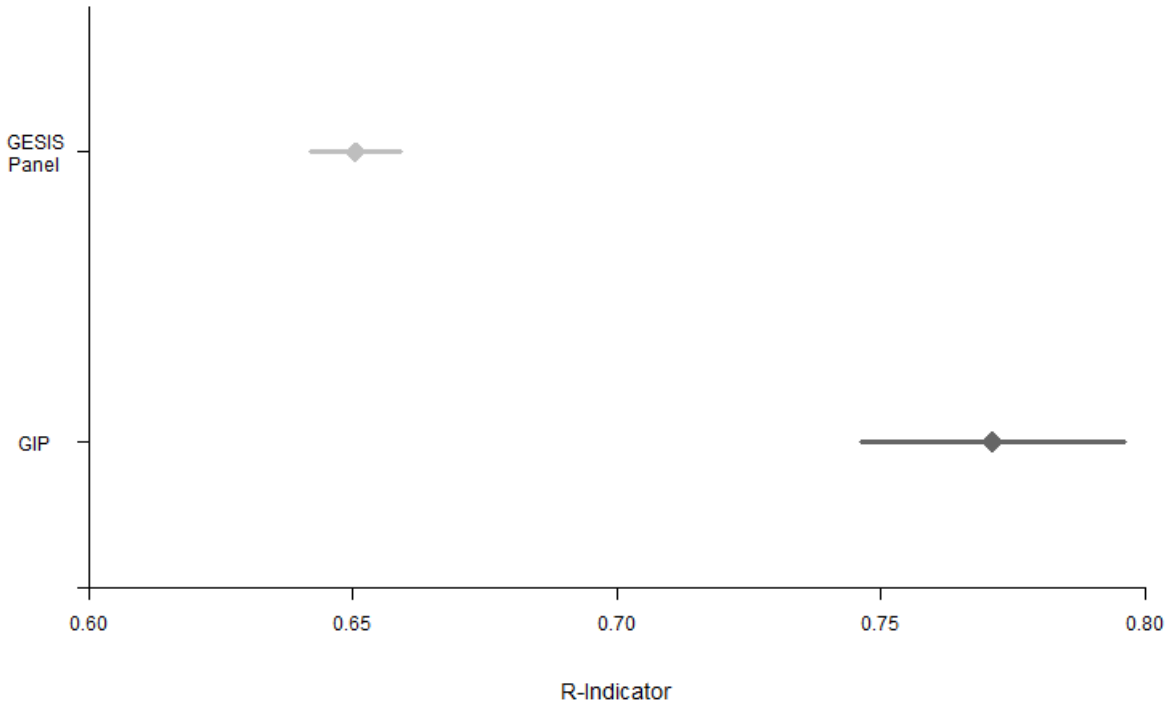

Figure 2: R-Indicators with missing values category additionally included as auxiliary variable for GIP and GESIS Panel recruitment interviews (including 95% confidence intervals)