Variety of Modes – Variety of Results? Comparability between Different Sources of Data Collection

Sven Vollnhals[1]

1   Introduction

Falling participation rates still threaten the quality of surveys (e.g. Heer 1999; Atrostic et al. 2001; Leeuw/Heer 2002; Stoop 2004; Curtin/Presser/Singer 2005; Brick/Williams 2012). Allthough low response rates do not automatically end up in a bias (e.g. Keeter et al. 2000; Groves/Heeringa 2006) it holds that when „[...] the rate of nonresponse rises, [...] the potential for bias increases [...]“ (Massey/Tourangeau 2012: 227). Therefore, to analyse the reasons for and consequences of nonrespone – and noncoverage – in surveys remains a central challenge for survey methodology to be able to cope with – or even to fix – the resulting influences on data quality (e.g. Groves 2011; Brick 2011; Brick 2013; Baker et al. 2013).

But does this necessarily imply that datasets from different sources of data collection on a unifying topic are less comparable regarding their results on central key variables because they are affected in *different* ways by nonresponse and noncoverage problems?

In the following paper, I will analyse this problem in the context of the federal election in Germany which took place on the 22th of September 2013.

2   Data Source

The „German Longitudinal Election Study“ (GLES) is a multi-component study to monitor the electorate in the forefront of the federal elections 2009, 2013 and 2017 (upcoming). Components 1, 2 and 3 for the election in 2013 will be used.[2]

The first one was conducted via personal interviews of 2003 (1744 entitled to vote) persons from a „[...] disproportional stratified multistage random sampling based on the ADM-Design.“ (Rattinger et al. 2013b: 2). The dataset provides a design weight to correct for the overrepresentation of the population from East-Germany and to simultaneously transform the household sample into a personal sample. The fieldtime ranges from 07/29/2013 up to 09/21/2013 (Rattinger et al. 2013b: 2 f.).

---

[1]Johannes Gutenberg-University Mainz. Contact: sven.vollnhals@politik.uni-mainz.de.

[2]For a detailed description of the design and aim see `http://gles.eu/wordpress/english/`. Components 1 and 2 target the population of Germany of 16 years or older living in a private household and that are eligible to vote. Component 3 restricts this definition to those with internet access.

Outline to be presented at the Nonresponse-Workshop in Leuven (4th of September 2015).

page 1 of 7

The second component consists of telephone interviews from a „Rolling Cross-Section" (RCS) design sampled over 76 days (n=7882, 7585 entitled to vote) from a „Multistage random sampling procedure based on the ADM-Design [...]" (Rattinger et al. 2013c: 2). The fieldtime is from 07/08/2013 up to 09/21/2013 (Rattinger et al. 2013c).

The third component is a sample drawn from an access-panel of (mainly) online recruited members hosted by Respondi.[3] The participation is rewarded with monetary incentives. The selection of the 4226 survey participants[4] is actively controlled through quota-sampling by age, gender and educational attainment (Rattinger et al. 2013a: 7 ff.). The fieldtime is 06/20/2013 up to 07/07/2013 (Rattinger et al. 2013a: 2). Two complementary samples from the same access-panel with the same questionnaire are merged with the following field times: 08/01/2013 up to 08/11/2013 and 09/02/2013 up to 09/12/2013.[5] All in all 6479 participants from the access-panel are entitled to vote and are therefore part of the targeted population.

3   Sociodemografics

Table 1 shows the differences according to sociodemografic standard variables in comparison to the official statistics from the census (CENSUS) conducted in 2011 in Germany.[6] Two main trends occur: The personal (PERS) and telephone (TEL) samples overrepresent the elderly and underestimate the percentage of the young. The access-panel (ONL) instead clearly underrepresents the population of those being 75 years old or older, who are nearly nonexistent. A typical problem arising in non-probability based online samples due to noncoverage – which occurs although a quota system was used.

Regarding the highest formal degree achieved, especially the telephone interview clearly underestimates the low educational level and overestimates those with a high level of educational attainment by nearly 20 percentage points. The mean size of the household, the percentage of women and the percentage of the population from East-Germany show no noticeable deviations from the official statistics in all components.

---

[3]Over 75% of the around 100 000 members are recruited through online advertisement, Facebook and search engines.

[4]1030 of the 5256 sampled persons participated already in 2009, they were excluded due to potential conditioning effects.

[5]For a description of the design of these additional samples see `http://gles.eu/wordpress/english/design/wkp/`. The third mentioned sample was conducted after the election and will not be used.

[6]The reference data from the census has been extracted from `https://ergebnisse.zensus2011.de/`. For a description of the methodology of the „Zensus 2011" see `http://tinyurl.com/py7j74z`.

Table 1: Sociodemografics Compared to Official Statistics

Survey	Age 16-29	Age 30-39	Age 40-49	Age 50-64	Age 65-74	Age 75-99	Age All
PERS	10 % (-)	11 %	14 % (-)	30 % (+)	18 % (+)	17 % (+)	| 100 %
TEL	9 % (-)	10 %	17 %	34 % (+)	18 % (+)	12 %	| 100 %
ONL	16 %	19 % (+)	20 %	32 % (+)	11 %	2 % (-)	| 100 %
CENSUS	20 %	14 %	19 %	24 %	13 %	10 %	|100 %

Survey	EastGer	Women	LowEdu	MidEdu	HiEdu	SizeHH
PERS	22 %	47 %	35 % (-)	37 % (+)	28 %	$\bar{x} = 2.27$
TEL	23 %	51 %	19 % (-)	32 %	49 % (+)	$\bar{x} = 2.34$
ONL	22 %	51 %	35 % (-)	32 %	33 %	$\bar{x} = 2.14$
CENSUS	20 %	51 %	42 %	29 %	29 %	$\bar{x} = 2.20$

Note: The personal interview (PERS) was weighted by the described design-weight. Red marked when percentage is five points or more below official statistics, brown marked when above. LowEdu includes no and the lowest school leaving certificate („Hauptschule"). MidEdu refers to the intermediary secondary qualification („Realschule") and HiEdu to the higher educational certificate required for university entrance („(Fach-)Abitur") in Germany.

4 Electoral Comparability

Table 2 shows statistics of general survey-related questions.

Table 2: Comparison of Survey-Related Questions

Study	Work Federal Government		Left-Right Self-Perception		Party Identification	Political Interest
	$\bar{x}$	s_x	$\bar{x}$	s_x	%	$\tilde{x}$
PERS	6.37	2.53	5.49	1.76 (-)	80 (+)	3
TEL	6.47	2.79	5.34	2.22	72	3
ONL	4.91 (-)	2.83	5.55	2.31	71	3

Note: The personal interview (PERS) was weighted by the described design-weight. Variables scaled as follows: 1-"Totally dissatiesfied" to 11-"Totally satisfied" (Work Federal Government), 1-"Left" to 11-"Right" (Left-Right Self-Perception), 1-"Very interested" to 5-"Not interested at all" (Political Interest).

The access-panel clearly shows a lower satisfaction with the work of the federal government, which consisted in the forefront of the election of a coalition of the Conservatives („CDU/CSU") and the Liberals („FDP").[7] Regarding the self-perception, the personal interview seems to have a tendency to the centre ground as the low standard deviation suggest – which may result from an interviewer effect sanctioning a too radical positioning. Party identification is slightly increased in the personal interview, political interest instead does not vary between the components according to the median.

---

[7]The following question was asked: „On the whole are you satisfied or dissatisfied with the way the CDU/CSU and FPD federal government is running the country?"

How to correct for the sociodemografic and survey-related discrepancies between the datasets?

In contrast to rake the dataset to the marginal distributions of e.g. the census data by „Iterative Proportional Fitting" (IPF), „Propensity Score Matching" (PSM) allows to account for the multivariate distribution of all auxiliary information – and to (better) incorporate for survey-related questions. As the PSM needs a reference dataset, a choice has to be made. The telephone interview heavily overestimates the proportion with a higher education, the online interview dramatically underestimates the group of 75+ years and the satisfaction with the government. Therefore the personal interviews will function as a reference.

Correction will take place for sociodemografics (age and education – „Socio") solely and combined with the survey-related evaluation of the satisfaction with the work of the government („SocioSurvey").[8] The design-weight is included in the matching model, to incorporate the survey design as e.g. DuGoff/Schuler/Stuart (2014): 289 suggest.[9]

Figure 1 shows the percentage of votes based on all three studies in comparison to the results of the German federal election (BTW) on the 22th of September 2013, varied by different weighting. Table 3 shows the corresponding absolute number of percentage points that the two pairwise regarded distributions differ.

Table 3: Bias According to Percentage of Votes

| Weighting	PERS-TEL	PERS-ONL	BOTH (PERS)	TEL-BTW	ONL-BTW	BOTH (BTW)
Design	8.62	34.39	43.01	17.51	36.66	54.17
Socio	8.91	29.06	37.97	14.99	37.07	52.06
SocioSurvey	8.66	13.25	21.91	20.41	18.87	39.28

Note: Table shows the overall percentage points that the telephone and online sample differ from the telephone sample and the official results of the election separately and combined.

The access-panel heavily underestimates the conservative party and overestimates the Pirates and the Leftist. The Greens are overrepresented in all components, especially in the telephone interview which additionally underestimates the Leftist (1). Correcting for age and educational attainment reduces the difference between the personal and the

---

[8] For all matching solutions „Genetic Matching" (Diamond/Sekhon 2013) has been used as an algorithm, which achieved the best results in terms of balancing. All matching solutions achieved a nearly 100 % reduction in the standardized mean differences of the calculated „Propensity Scores" and auxiliary information, see Ho et al. (2011): 23 for the computation.

[9] To rake the personal interview at first to the census data according to age, education, regional belonging and gender and include this weighting vector removed all differences between the telephone data and the census on the aggregated level but clearly reduced the comparability of the voting behaviour afterwards between the datasets and is therefore not reported.

Figure 1: Sociodemografic and Survey-Related Voting Behaviour
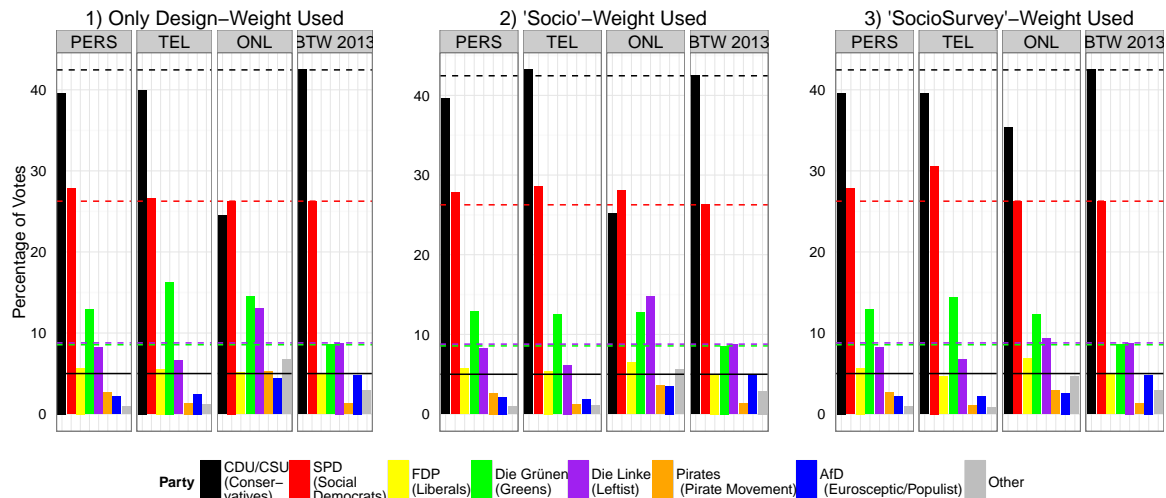
Note: Black reference line refers to the 5 % treshold for a party to move into parliament in Germany. The coloured dashed reference lines refer to the result of the general election as a reference.

online sample (2) but to get the online and personal data on a clearly more comparable level needs the additional correction for the satisfication with the performance of the government (3). Although weighting on sociodemografics reduces the difference between the personal and telephone samples regarding the Greens (2), this difference increases again when the survey-related factor is included in the calculation of the weights.

Even in a multi-component study, it seems hard to find a general weighting solution, which is highly dependend on the type of bias that occured. But correcting for nonresponse and noncoverage in the context of an/this electoral study is necessary when evaluating other items that are influenced by the political orientation. The additional incorporation of survey-related questions improves this comparability if it fixes the underrepresentation of a relevant group (here: the partisans of the conservative party).

A bunch of practical questions still remains:

1. When does the incorporation of survey-related questions become tautological?

2. How to measure comparability between different datasets on common variables? Are R-Indicators (e.g. Schouten et al. 2012) a possibility?

3. How to distinguish between the mode effect and the nonresponse bias?

4. What defines a „good" reference dataset? The probability-based sampling?

5. How to deal with the incorporation of complex survey designs into PSM?

6. Are contextual information (e.g. of the electoral districts) valuable information?

Literature

Atrostic, B. K. et al. (2001): Nonresponse in U.S. Government Household Surveys: Consistent Measures, Recent Trends, and New Insights, in: *Journal of Official Statistics*, 17/2: 209–226.

Baker, R. et al. (2013): Summary Report of the AAPOR Task Force on Non-probability Sampling, in: *Journal of Survey Statistics and Methodology*, 1/2: 90–143.

Brick, J. Michael (2011): The Future of Survey Sampling, in: *Public Opinion Quarterly*, 75/5: 872–888.

Brick, J. Michael (2013): Unit Nonresponse and Weighting Adjustments: A Critical Review, in: *Journal of Official Statistics*, 29/3: 329–353.

Brick, J. Michael/D. Williams (2012): Explaining Rising Nonresponse Rates in Cross-Sectional Surveys, in: *The ANNALS of the American Academy of Political and Social Science*, 645/1: 36–59.

Curtin, R./S. Presser/E. Singer (2005): Changes in Telephone Survey Nonresponse over the Past Quarter Century, in: *Public Opinion Quarterly*, 69/1: 87–98.

Diamond, Alexis/Jasjeet S. Sekhon (2013): Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies, in: *Review of Economics and Statistics*, 95/3: 932–945.

DuGoff, Eva H./Megan Schuler/Elizabeth A. Stuart (2014): Generalizing Observational Study Results: Applying Propensity Score Methods to Complex Surveys, in: *Health Services Research*, 49/1: 284–303.

Groves, Robert M. (2011): Three Eras of Survey Researchs, in: *Public Opinion Quarterly*, 75/5: 861–871.

Groves, Robert M./Steven G. Heeringa (2006): Responsive design for household surveys: tools for actively controlling survey errors and costs, in: *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169/3: 439–457.

Heer, Wim de (1999): International Response Trends: Results of an International Survey, in: *Journal of Official Statistics*, 15/2: 129–142.

Ho, Daniel et al. (2011): MatchIt: Nonparametric Preprocessing for Parametric Causal Inference, in: *Journal of Statistical Software*, 42/8: 1–28.

Keeter, Scott et al. (2000): Consequences of Reducing Nonresponse in a National Telephone Survey, in: *Public Opinion Quarterly*, 64/2: 125–148.

Leeuw, Edith de/Wim de Heer (2002): Trends in household survey nonresponse: A longitudinal and international comparison, in: Robert M. Groves ed.: *Survey nonresponse.* Wiley series in probability and statistics. New York: Wiley: 41–54.

Massey, Douglas S./Roger Tourangeau (2012): Where Do We Go from Here? Nonresponse and Social Measurement, in: *The ANNALS of the American Academy of Political and Social Science*, 645/1: 222–236.

Rattinger, Hans et al. (2013a): *Campaign Panel (GLES 2013): ZA 5704 Datafile Version 3.0.0.* GESIS Data Archive. Cologne.

Rattinger, Hans et al. (2013b): *Pre-election Cross-section (GLES 2013): ZA5700 Data file Version 2.0.0.* GESIS Data Archive. Cologne.

Rattinger, Hans et al. (2013c): *Rolling Cross-Section Campaign Survey with Post-election Panel Wave (GLES 2013): ZA5703 Data file.* GESIS Data Archive. Cologne.

Schouten, Barry et al. (2012): Evaluating, Comparing, Monitoring, and Improving Representativeness of Survey Response Through R-Indicators and Partial R-Indicators, in: *International Statistical Review*, 80/3: 382–399.

Stoop, Ineke A. (2004): Surveying Nonrespondents, in: *Field Methods*, 16/1: 23–54.