33

Representativeness in Online Surveys through Stratified Samples

Jörg Blasius and Maurice Brandt

Introduction

The advantages and disadvantages of online surveys, as compared to other data collection methods, have often been extensively documented (for example, Dillman, 2000; Couper, 2001; Fricker and Schonlau, 2002; Couper et al., 2004); compared to face-to-face, telephone, and mail surveys, online surveys have the advantage of being cheaper, faster and independent in terms of time and space. The disadvantage is that they depend on the availability of Internet access. At present, online surveys in Germany are accounting for approximately 31% of all surveys; in 1998, it was only 1% (http://www.adm-ev.de/fileadmin/user_upload/PDFS/Jahresbericht_08.pdf, access as of March 10, 2010). As in other countries, Internet access in Germany with its 42.7 million people online (or 65.8% of the adult population, http://ard-zdf-onlinestudie.de/fileadmin/PM/PM2008_2.pdf, access as of March 10, 2010) is heavily distorted by age, education and sex (Van Eimeren and Frees, 2006: 404). Thereby, some groups are almost completely linked to the Internet, so that a survey of business clients (Deutskens et al., 2006), students (Kwak and Radler, 2002; Kaplowitz et al., 2004), and other selected target groups (e.g., users of online banking, eBay or Amazon) will presumably not raise any problems. But there are other groups who are almost entirely excluded from online

Source: *Bulletin de Méthodologie Sociologique*, 107 (2010): 5–21.

surveys; for example, elderly women with low educational attainment. With Internet accesses being so unequally distributed, it seems almost impossible to obtain representative results for the entire population – which, however, is often desired.

Despite the continuously growing number of Internet users, the lack of representativeness of the entire population will apparently remain an unsolved problem for another couple of years, as will the question of how to obtain representative results using online surveys. The current practice is to weight variables to make them representative with respect to socio-demographic characteristics such as age, sex and education (e.g., Faas and Schoen, 2006; Loosveldt and Sonck, 2008), or attitudes (e.g., Schonlau et al., 2004, 2007, 2009; Loosveldt and Sonck, 2008; Lee and Valliant, 2009). The operations of up-weighting and down-weighting – made to account for socio-demographic characteristics – and in cases of multiple weighting (e.g., sex × age (groups) × education (groups)), may cause the factors to differ by more than 100 (Vehovar et al., 1999), which may cause estimates that are highly inefficient and unstable. Bandilla et al. (2003: 241) already regard weighting factors greater than five as problematic and not very helpful. Furthermore, how can one justify assigning somebody's attitudes the hundredfold weight of another person's attitudes?

To avoid heavy weights and biased samples, online panel providers have recently started to offer so-called stratified samples. To evaluate this mechanism, we drew stratified samples from an online access panel, using population parameters of a three-way table of "sex × age groups × education groups" to estimate sample sizes. The online panel we used was a pool consisting of more than 20,000 individuals from the Western states of Germany who had expressed their readiness to participate in online surveys. As is probably true for all online panels, the panel is distorted with respect to the general population: The participation of well-educated young men is highly above average, whereas elderly women with low school attainment are scarcely represented. These kinds of over-representation of certain groups also result from other online panels that acquire respondents via conventional data collection methods such as doorstep surveys or telephone surveys (Hoogendorn and Daalmans, 2009). It follows that every simple random sample drawn from this panel will also be distorted. However, considering that we know age, sex and highest educational attainment of the panel members, and that this information is also available for the entire population, we are able to draw a sample that is subject to the restrictions of the three-way table of age (group) × sex × education (group); an elderly woman with low school attainment will have a high probability of being selected, while a young men with high educational attainment will hardly be likely to be selected.

As it will be shown later, we had to restrict our data to ages up to 49 years to be able to draw a sample without coverage error. Restricted by the

age of 49, we should receive an online sample that in terms of sex, age and education is comparable to a representative face-to-face sample drawn during the same time period. In this paper, we raise the question of whether or not an online sample that is stratified by the cells of the tree-way table of age × sex × education without coverage error is representative for the general population. To examine this question, we drew a stratified online sample and compared it with data from a representative face-to-face survey, the German Social Survey (GSS) 2002. As a basis for this comparison, we used a set of questions on attitudes towards society and socio-demographic characteristics.

Literature Review

The samples being used for large national and international face-to-face or telephone surveys, such as the German Socio-Economic Panel (GSOEP), the German Social Survey (GSS) or the International Social Survey Program (ISSP), are considered representative for the general population. By contrast, online samples are currently regarded as being representative of population subgroups only, even if the panel is based on a probability sample (Rookey et al., 2008; Hoogendorn and Daalmans, 2009). To overcome biased samples in Web surveys, a common procedure is to use weighting procedures based on stratification variables or propensity scores.

Weighting the Data

There are various possibilities for using the Internet for data collection. A technique applied in addition to self-selection methods with banners or pop-up windows, which do not even guarantee representativeness with regard to the universe of Internet users (e.g., Faas and Schoen, 2006), is the use of online access panels. These panels consist of a pool of persons or households which have been invited to participate in online surveys; for example, online – on the websites of institutes – or offline – at the end of a face-to-face or telephone interview. Such an online panel is used to draw random samples of a magnitude that allows researchers to obtain the desired number of cases. Taking into account that Internet connections are not distributed in a representative manner, the project will face severe distortions affecting both the entire online panel and the random sample to be drawn from it.

Weighting procedures, mostly with regard to socio-demographic variables, are used to adjust biased samples to the distribution of the general population (e.g., Bandilla et al., 2003; Faas and Schoen, 2006). Two weighting procedures should be distinguished: The first weights the single variables one by one; the weights provided by this method are relatively small

and sample distortions are only limited and small. For example, when up-weighting older people by an age group factor, this method does not account for the fact that elderly people in the online panel are represented with a percentage of university graduates much higher than should be expected, according to this age group's share of the entire population. The second method is multiple or cell weighting; that is, considering several variables that are assumed to be biased simultaneously. Thus, three educational groups, two sexes and five age groups already generate 30 cell weighting factors (or ten factors when successively weighting by single variables); some of them are presumably very high – elderly women with a low level of formal education – while others are very low, young men with high education.

Cell weighting factors in online surveys often seem to exceed values of 100. For example, Faas (2003: 130) reported weighting factors that range between 0.053 and 7.061: while one person is assigned a weight far below average, another one is given a weight far above average; comparing the weights of the members of the two respective cells shows they are related by factor of 133. To reduce such values to a maximum of 25, Vehovar, Lozar Manfreda and Batagelj (1999) proposed to aggregate the categories. Furthermore, with respect to the Internet population, Dever, Rafferty and Valliant (2008: 57) came to the conclusion that "the standard randomization justification of weighting the random sample to represent the target population does not apply to Internet survey estimates because the sample itself is not selected from the correct population."

An alternative to traditional weighting based on socio-demographic variables is propensity scoring, a method that can be traced back to propensity score matching (Rosenbaum and Rubin, 1984). The propensity scores can be estimated on the basis of attitudes, values, and also on socio-demographic variables (Duffy et al., 2005; Lee, 2006; Schonlau, van Soest and Kapteyn, 2007; Loosveldt and Sonck, 2008; Lee and Valliant, 2009; Schonlau et al., 2009). The application of propensity scoring requires at least two surveys with the same questions; for example, from an online and a face-to-face survey. Logistic regression facilitates the development of a statistical model for estimating the probability that each respondent, conditional on her or his characteristics (attitudes, values, socio-demographics), participated in the face-to-face survey rather than in the online one (for more details, see Duffy et al., 2005; Lee and Valliant, 2009; Schonlau et al., 2009).

Survey Mode Effects

Comparative studies of face-to-face, telephone and mail surveys revealed the existence of survey mode effects (e.g., De Leeuw, 1992). On the one hand, this means different response rates both in terms of individuals answering single questions and in terms of different population groups

being ready to respond at all. On the other hand, mode effects have an impact on a survey's contents; for example, in questions about sensitive issues where respondents may prefer the one or the other technique for feeling safe when giving true answers. Comparing a face-to-face and an online sample, Heerwegh and Loosveldt (2008) came to the conclusion that Web survey respondents produce higher "don't know" response rates, more item non-response, and differentiate less on rating scales than face-to-face survey respondents.

To examine mode effects between mail and Web-based surveys, Denscombe (2006) interviewed 338 15-year-old pupils on their health behavior, 69 of whom had to fill in the questionnaire in the school's computer laboratory and 269 in the classroom with paper and pencil. Pupils were assigned to one of the groups at random; there were no significant distinctions in terms of sex and ethnic origin. Descombe (2006: 252) did not find significant distinctions between the two samples, which could not be interpreted as accidental. The only exception in the 23 questions concerned the pupils' statement that they were smoking more than 20 cigarettes a week. While none of the pupils admitted this in the Web-based survey, one out of seven replying in the paper and pencil survey confessed to do so. Lozar Manfreda and Vehovar (2002) obtained a similar result in their survey of educational institutions in Slovenia. The authors contacted 200 institutions at random, asking them to reply to their questions by mail or via the Internet; they too did not find any conspicuous distinctions between the results. Finally, in a comparison of a Web survey with a traditional paper survey of employees, Gesell et al. (2007) reported that there are no mode effects which are important in terms of job satisfaction.

To examine mode effects, we would have to draw two samples from the online panel: one of them to be interviewed online; the other face-to-face or by mail. Since we were not interested in running such a survey, we cannot test for mode effects in our data. But from previous findings, we can assume that they do not affect the substantive solution.

Data and Method

This study examines whether or not it is possible to draw a representative sample from an online panel. We use German micro-census data, which are based on a one-percent probability sample of the entire population conducted by the Federal Statistical Office, to generate a population sample using the three-way table of age (group) × education (group) × sex as being representative in terms of these three indicators. The micro-census is partly realized by face-to-face interviews and partly by self administrated questionnaires. To avoid the still existing East-West differences in Germany, we restricted the study to the Western states. For age, we considered nine groups with a

range of five years (except for the first group) from 18 to 64; three groups for education, together with the two sex groups, and we have a total of 54 groups. The determinant for the number of education groups is that there are three main levels of school attainment in Germany: primary after nine years of school; secondary after ten years of school; and University entrance certificate after 13 years of school. The number of age groups was based on estimation. The aim was to obtain a net sample size of approximately 1,650 respondents for the online survey, which is the number of GSS 2002 participants of the mentioned age groups in the Western states of Germany. The attainment of cell occupancy rates of on average of 30 respondents allowed us to handle approximately 54 cells, which should be sufficient to avoid haphazard results.

The online panel, from which the sample was drawn, included more than 20,000 individuals who live in West Germany and whose age, sex and education are known by registration. To respect the standards of AAPOR (the German equivalent is "Arbeitsgemeinschaft Deutscher Marktforschungsinstitute" or, in short, ADM, see ADM, 2008) "the panel provider conducts a large Internet survey twice a year with a total of 40,000 respondents. The respondents are contacted as the *n*th visitor on very different Internet pages (with very different target audiences) via pop-up window or via layer. In case of an interest in participating in surveys, the respondents are asked to register. The panel provider checks all postal addresses for the future panel participants and makes sure that nobody is able to register twice" (translated from German, taken from the self-description of the panel provider, see: www.bvm.org/user/dokumente/kongress/vonHeesen.pdf, especially page 3; access as of March 10, 2010). As is probably true for all such panels, we were confronted with the fact that males, youngsters and better educated people were strongly over-represented (cf., Hoogendorn and Daalmans, 2009). Assuming that we are to analyze 1,650 panel members of up to 65 years of age and knowing from experiences with that panel that the response rate will be approximately 60%, we would have to invite approximately 2,800 panel members to participate in the survey.

In the first step, we compare the marginals of sex, education, and age groups from the online access panel with the micro-census data. Table 1 shows that the panel is biased – as probably all panels are – in a way that almost reflects the different user groups of the Internet: females, elderly and less educated people are under-represented. Under these conditions, the probability of being selected is quite different: panel members aged between 25 and 29 had a chance of 6.7% while 60 to 64 years old panel members had a chance of 90.5% (see Table 1). Considering only age for a stratified sample, a 60 to 64 year-old panel member would have a 13.5 times higher probability of being selected than somebody aged 25 to 29.

Because quota formation should be done simultaneously for all three variables, we built a three-dimensional contingency table of age (nine categories),

Table 1: Socio-demographic distribution in online-panel, micro-census 2002, and the probability of getting selected in a stratified sample

	Online-panel	*Micro-census*	*Probability to become selected*
Sex			
Male	58.2	49.5	11.9
Female	41.8	50.5	16.4
N	21,112		
Education			
Primary school	16.2	48.0	44.0
Secondary school	36.6	26.3	9.9
University entrance certificate	47.1	25.7	7.5
N	20,921		
Age groups			
18–24	15.0	10.7	10.0
25–29	18.3	8.8	6.7
30–34	16.6	12.1	10.2
35–39	17.6	13.8	10.9
40–44	12.1	12.7	14.6
45–49	8.7	11.3	18.2
50–54	5.8	10.5	25.3
55–59	4.1	8.9	30.5
60–64	1.7	1 1.1	90.5
N	20,413		

Table 2: Gross samples (total N = 2,828) as specified by the micro-census 2002

	Males			*Females*			
Education/age	*Low*[1]	*Medium*[2]	*High*[3]	*Low*[1]	*Medium*[2]	*High*[3]	*Total*
18–24	61	48	45	43	55	52	304
25–29	45	34	46	38	41	46	250
30–34	69	45	59	55	62	52	342
35–39	87	50	63	69	68	55	392
40–44	88	42	52	75	58	43	358
45–49	84	33	44	82	44	33	320
50–54	87	26	35	93	34	22	297
55–59	80	22	27	81	29	15	254
60–64	109	23	25	115	30	13	315
Total	710	323	396	651	421	331	2,832

[1] Primary school (9 years).
[2] Secondary school (10 years).
[3] University entrance certificate (13 years).

sex (two categories) and education (three categories), a total of 54 cells. We use the micro-census data to determine the shares for each of the 54 cells so that the gross sample we obtain is stratified in terms of age (group) × education (group) × sex. Table 2 shows the gross samples required for each cell; since only "entire" persons could be drawn from the panel, all values were rounded off upward.

As Table 2 shows, for the sample to be representative with respect to elderly women, we need many more women with a low level of education than women with medium or high educational attainment – whereas the situation

Table 3: Selection probability of an individual in a given cell for a sample from the universe (N = 20,230), percentages

	Males			*Females*		
Education/age	*Low*[1]	*Medium*[2]	*High*[3]	*Low*[1]	*Medium*[2]	*High*[3]
18–24	30.9	8.5	6.9	19.1	8.1	7.3
25–29	36.6	6.8	3.8	19.8	6.1	4.5
30–34	36.6	8.2	5.6	26.9	9.6	7.4
35–39	26.9	7.2	6.2	31.7	9.8	8.9
40–44	27.9	6.9	7.4	64.2	14.0	14.4
45–49	33.1	7.9	8.5	75.5	16.2	16.4
50–54	38.0	10.3	9.0	130.9	27.3	19.3
55–59	43.7	11.1	10.4	158.3	32.3	30.8
60–64	135.8	21.7	22.1	1,437.3	130.4	78.9
Total	37.5	8.3	6.7	54.5	11.6	8.8

[1] Primary school (9 years).
[2] Secondary school (10 years).
[3] University entrance certificate (13 years).

in the online panel is the contrary. In the group of 60 to 64 year-old women, the low to medium education ratio of the panel is 8 to 23 (not shown), but the ratio we need is 115 to 30 (Table 2). In other words, the probability that a woman with a medium educational level is drawn, rather than a woman with a low level of education, is nearly 3:1 in a random sample, but what we actually need is only one woman with medium education for every four women with low education. To clarify the relationships between the cases we have and the cases we need, we determine a person's probabilities to be drawn from each cell, which is obtained as the quotient of "persons to be drawn" and "all persons in the panel" (see Table 3).

Table 3 shows the probability for each person to become a member of the sample. For example, a 18 to 24 year-old man with high education has a probability of 6.9% to become a member of the sample, while a 45 to 49 year-old woman with low education has a chance of 75.5%. The persons who will be drawn in any case include all women aged 50 years or older with a low level of education and women between 60 and 64 years of age with a medium level of education, and men in the same age group with a low educational level. These five cells do not even include a sufficiently large number of persons required to draw a sample without coverage error. In the case of 60 to 64 year-old females with low education, even a panel that is ten times larger would not be sufficient to collect the desired sample. And if we would include even older people, for example 65 to 74 year-olds as done by Loosveldt and Sonck (2008), one would need an access panel that would include 300,000 members. Furthermore, those panel members who belong to the highly required groups in stratified samples cannot be included in every study. Since we are not able to draw a sample without coverage error in terms of age × education × sex for the general population, subsequent evaluations will be restricted to the age group of the 18 to 49 years-old.

Of the 1,966 individuals of up to 49 years-of-age, we invited participation (compare Table 2, rows 1 to 6), 1,396 (= 71.0%) in filling in the questionnaire; the online data were evaluated in July 2004. As our reference sample for comparison, we choose the German Social Survey (GSS) 2002, which was carried out as a CAPI survey between February and August 2002 (Blohm et al., 2003). It is based on a probability sample. All respondents were informed by postal mail that they were going to be interviewed in the next few days; the response rate was 47.3% (Blohm et al., 2003: 56). The GSS can be considered one of the main surveys for the social sciences in Germany.

To compare the online survey with the offline survey, we replicated – amongst others – nine questions concerning a person's attitude towards "society" from the GSS 2002 survey. We also considered demographic characteristics such as household size, marital status, religious affiliation, stratum allocation, gainful employment and income. We designed the visual layout of the questionnaire in compliance with rules by Dillman (2000) to minimize non-response as far as possible.

Results

Socio-demographic Variables

As a first analytical step, the socio-demographic data of the micro-census were compared with the data from the GSS 2002 and the stratified sample of the online panel. Table 4 displays the variables age, sex and education for

Table 4: Socio-demographic comparison between the micro-census, the GSS 2002 survey and the stratified sample of the online panel, percentages, 18 to 49 years

	Micro-census	*GSS*	*Online*
Sex			
Males	50.6	50.9	51.8
Females	49.4	49.1	48.2
N		1,115[1]	1,396[1]
School category			
Primary school	40.4	28.2	28.5
Secondary school	29.6	32.0	30.5
University entrance certificate	30.0	39.8	41.0
N		1,083[2]	1,344[2]
Age groups			
18 to 24 years	15.4	18.5	14.0
25 to 29 years	12.7	12.4	14.3
30 to 34 years	17.4	15.7	18.3
35 to 39 years	19.9	22.2	19.8
40 to 44 years	18.2	16.8	18.3
45 to 49 years	16.2	14.5	15.3
N		1,115[3]	1,396[3]

[1] $\chi^2 = 0.2$, where df = 1; n.s.
[2] $\chi^2 = 0.7$, where df = 2; n.s.
[3] $\chi^2 = 14.9$, where df = 5; $p < 0{,}05$; Cramer's V = 0.08.

the three data sources, all calculated with respect to an age of 18 to 49 years. It should be noted that the online data show a few discrepancies between panel information (given during panel registration) and survey information, which might be explained by time (in the case of education, a higher degree could have been obtained) and by the fact that another household member filled in the questionnaire. Happily, only few cases were affected and we will hereafter use only that survey information.

Comparing the GSS and the online survey, gender reveals only very small distinctions ($\chi^2 = 0.2$, with df = 1), discrepancies from the micro-census being marginal for both surveys. Regarding educational attainments, there is also no difference between the GSS and the online data ($\chi^2 = 0.7$, with df = 2), but – as anticipated – persons with primary school education are under-represented in both surveys, while persons with a university entrance certificate are over-represented. With respect to age, GSS and online surveys differ only marginally and not systematically. However, the differences are significant on the 5% level. Comparing both samples with the micro-census solution shows that the online sample performs a little better, except for the 25 to 29-year-olds; the values from the online survey are closer to the micro-census than those from the GSS. It can be concluded that the online sample covering individuals between 18 and 49 years of age has a distribution similar to that of the GSS. Comparing both data sets by the pooled variable of age × sex × education, the difference is not significant on the 5%-level ($\chi^2 = 48.9$ with df = 35, n.s., N = 2,427).

Individual Characteristics

In addition to sex, school attainment and age, respondents were requested to supply additional information on other demographic variables. The variables included the number of persons in the household, marital status, religious denomination and frequency of churchgoing. Table 5 shows that there is a highly significant difference in the number of persons in a household ($\chi^2 = 34.1$, with df = 4; $p < 0.001$; Cramer's V = 0.12), as one- and two-person households are over-represented in the online sample, whereas households with four and more persons are underrepresented. The average household size is approximately 3.0 persons in the GSS and 2.7 persons in the online sample. With respect to this variable, the GSS is clearly closer to the data from official statistics than those of the online panel.

There are also highly significant differences in marital status between the GSS and the stratified sample of the online panel. Thus, the sample of the online panel includes relatively few married people who live together with their spouses, but relatively many people who are divorced or have never been married. And again, the GSS performs clearly better than the online survey.

Table 5: Comparison of percentages and χ^2 test for structure variables, percentages

	Micro-census	*GSS*	*Online*
Persons in the household[1]			
Live alone	16.4	16.6	19.9
2 persons	21.9	22.2	29.0
3 persons	23.8	22.2	22.1
4 persons	26.0	25.5	19.7
5 persons and more	11.9	13.5	9.2
	1,115	1,384	
Marital status[2]			
Married and live together with spouse	52.1	51.3	43.1
Married and live in separation	2.3	2.2	4.3
Widowed	0.6	0.6	0.5
Divorced	5.8	6.6	9.3
Never married	39.2	39.2	42.8
	N	1,114	1,384
Religious denomination[3]			
Evangelical church		35.3	30.9
Free evangelical church		1.5	2.6
Roman-catholic church		38.6	31.7
Other Christian religious community		2.2	1.7
Other non-Christian religious community		4.7	1.6
No religious denomination		17.7	31.5
		1,108	1,383
Frequency of churchgoing[4]			
Never		28.3	35.7
Unfrequent		34.9	45.1
Several times a year		19.4	13.4
One to three times a month		9.0	3.0
Once a week		6.2	2.1
More than once a week		2.2	0.7
		1,113	1,383

[1] $\chi^2 = 34.1$, where df = 4; $p < 0.00l$; Cramer's V = 0.12.
[2] $\chi^2 = 24.5$, where df = 4; $p < 0.001$; Cramer's V = 0.10.
[3] $\chi^2 = 82.5$, where df = 5; $p < 0.001$; Cramer's V = 0.18.
[4] $\chi^2 = 117.6$, where df = 5; $p < 0.001$; Cramer's V = 0.22.

Religious denomination is also marked by highly significant differences between the online survey and the GSS (no micro-census data are available for this variable). Here, people without a religious denomination are clearly over-represented in the online sample. The differences in the frequency of churchgoing are even more pronounced: only 6% of the online respondents but 17% of the GSS respondents go to church at least once per month.

Although age, sex and education have – as we had hoped – a similar distribution in the two surveys, the respondents of the online panel differ rather strongly from those of the GSS regarding the number of persons in a household, marital status, religious denomination and frequency of churchgoing. These differences are a first indication that central socio-demographic variables are not sufficiently represented and that we cannot speak of representativeness in terms of the general population.

Table 6: ANOVA: Attitudes towards the society, GSS 2002 and online survey; mean values, standard deviation (in brackets), test statistic

	GSS	*Online*	*F*	*sig*	*η*
Diligence and performance	7.06 (2.04)	7.16 (2.04)	1.37	n.s.	
Responsibility	8.37 (1.66)	8.47 (1.78)	2.10	n.s.	
Conformity	8.05 (1.74)	8.07 (1.78)	0.16	n.s.	
Self-assured and critical	8.00 (1.77)	8.09 (1.68)	1.48	n.s.	
Laissez faire	3.87 (2.59)	4.18 (2.68)	8.08	$p < 0.01$	0.06
Wealth	7.89 (1.77)	8.52 (1.66)	83.94	$p < 0.001$	0.18
Political self-contribution	6.21 (2.23)	7.21 (2.07)	136.10	$p < 0.001$	0.23
Self-realization	7.50 (1.96)	7.93 (1.75)	33.45	$p < 0.001$	0.11
Religiosity	4.57 (2.59)	3.81 (2.39)	56.74	$p < 0.001$	0.15

Attitudes towards Society

To compare attitudes, both surveys used a "society module" of questions which included nine items in total. Among these items, respondents were asked: "How much would you like to live in a society that considers it important that people display diligence and performance?" (For question wording, see Table 6). The question responses used 10-point scales, ranging from 1 (= would not like it at all) to 10 (= would like it very much).

A comparison of the mean values of society-related variables between the two samples (ANOVA statistics, see Table 6) shows that general social values such as "diligence and performance", "responsibility" and "conformity" are considered equally important in both samples. However, there are clear distinctions concerning characteristics such as "laissez faire", "wealth", "self-realization" and, in particular, "political self-contribution", as the online respondents regarded these four properties as much more important than the GSS respondents did – the latter attributing more importance to the rather conservative value of "religiosity". With respect to the differentiation of the rating scales, we can only partly support the solution of Heerwegh and Loosveldt (2008): Using the standard deviation as an indicator for the variation in the data, there are five items of which the value for the face-to-face survey is higher than for the Web survey, while the opposite is true for three other items.

For a closer inspection of attitude patterns, we applied Categorical (or Nonlinear) Principal Component Analyses (in short, CatPCA; see Gifi, 1990; Michailidis and De Leeuw, 1998; De Leeuw, 2006) to scale the nine questions because this method does not require metric data as PCA does. In CatPCA, only the order of the successive categories is retained or at least not disrupted, the distances between the successive categories are calculated with an iterative procedure. The solutions of CatPCA can be interpreted as traditional PCA; together with eigenvalues and their explained variances, the method also provides factor loadings and factor scores. The scaling was done separately for both surveys and for the pooled data set. Thus we

Table 7: Categorical principal component analyses of attitudes towards the society, GSS 2002 and online survey, factor loadings

	GSS		*Online panel*		*GSS and online*	
	Factor 1	*Factor 2*	*Factor 1*	*Factor 2*	*Factor 1*	*Factor 2*
Diligence and performance	0.586	–0.410	0.594	–0.364	0.585	–0.384
Responsibility	0.737	–0.004	0.648	–0.125	0.692	–0.094
Conformity	0.686	–0.306	0.656	–0.399	0.663	–0.364
Self-assured and critical	0.699	0.283	0.711	0.188	0.702	0.207
Laissez faire	–0.110	0.674	–0.145	0.752	–0.124	0.707
Wealth	0.575	0.081	0.574	0.139	0.577	0.155
Political self-contribution	0.599	0.230	0.633	0.313	0.623	0.288
Self-realization	0.594	0.512	0.600	0.491	0.600	0.497
Religiosity	0.353	–0.452	0.194	–0.198	0.252	–0.374
Eigenvalue	3.024	1.322	2.857	1.306	2.913	1.324
Explained variance	33.6%	14.7%	31.7%	14.5%	32.4%	14.7%
Cronbach's Alpha	0.753	0.274	0.731	0.264	0.739	0.275

N = 1,102 (GSS 2002), N = 1,386 (Online panel), N = 2,488 (GSS and Online).

Table 8: ANOVA: Latent attitudes towards the society, GSS 2002 and online survey; mean values, standard deviation (in brackets), test statistic

	GSS	*Online*	*F*	*sig*	*η*
Dimension 1	–0.123 (1.02)	0.096 (0.97)	28.9	$p < 0.001$	0.11
Dimension 2	–0.184 (1.01)	0.147 (0.97)	69.0	$p < 0.001$	0.16

excluded all cases in which at least one item had a missing entry or a "don't know"; in the case of the GSS, we lost 13 cases, in the online survey we lost 10 cases. This solution does not support the finding by Heerwegh and Loosfeldt (2008) that online surveys produce more non-substantive responses than face-to-face surveys.

Table 7 shows that the variance, which can be explained by the first two factors, and the structure of the factors are almost identical in both the GSS and the online panel surveys. The first factor in both data sets is a "general factor", except "laissez faire" all elements are positively associated with it. The higher the corresponding factor scores of the respondents, the more they agree with values such as "responsibility", "conformity", and "self-assured and critical". For both surveys, the second factor reflects conservative values such as "religiosity" and "diligence and performance" on the negative part versus liberal values such as "laissez faire" and "self-realization" on the positive part. Although the mean values of five out of the nine items differ significantly (Table 6), the factor structures are almost the same (Table 7). This finding can be supported by the solution for the pooled data set: The values for the explained variances as well as for the factor loadings are in between both solutions (with the exception for "wealth").

Applying ANOVA with the survey mode as independent variable to both dimensions of the pooled solution shows that the online participants have significant higher values on the "general factor"; their agreements with the given values of the society are higher (Table 8). With respect to liberal vs.

conservative values (dimension 2), the online participants tend to be more liberal while the GSS participants tend to be more conservative. Furthermore, there are slight differences in terms of variation, which are in the same direction as found by Heerwegh and Loosveldt (2008).

Summary and Discussion

The survey design described here has made it possible to draw a sample without coverage error from an online panel of respondents up to 49 years of age with respect to the cells of the three-way table age × sex × education. Even if there could be as many respondents in the online panel as required for including groups such as "up to 65 year-old women with a low level of education" so that they reach the required sample size, they cannot be contacted in every survey that requires representativeness in terms of age × sex × education. In other words, stratified samples, as usually offered by many panel providers, can only be drawn for elderly and less educated panel members under certain restrictions. Loosveldt and Sonck (2008), and Duffy et al. (2005) only mentioned that they did "stratified sampling" with respect to age, sex, and education (Loosveldt and Sonck, 2008), or age, sex and region (Duffy et al., 2005); it is not documented what they did in situations where in some cells there were not enough cases.

Furthermore, even in the case of a genuine stratified sample by age × sex × education, representativeness will not be ensured since there are differences in other socio-demographic variables such as religious denomination and household composition. Regarding religious denomination, an above-average number of people do not belong to a religious community in the online panel – and, correspondingly, the respondents of the online panel go to church much less frequently than the GSS respondents do. Moreover, the respondents differ regarding attitude questions such as political self-contribution, economic safety or religiosity. This indicates that respondents who were interviewed offline differ from respondents who were interviewed online for indicators other than those covered by age, education and sex. Our results confirm those obtained by Couper et al. (2007): there are differences in the attitudes of persons who were asked online and persons who were asked offline, even when socio-economic variables are kept constant. However, further improvements might be possible when propensity scoring and calibration adjustment (Lee and Valliant, 2009) are additionally applied to stratified sampling. But applying propensity scoring includes the disadvantage of extending the questionnaire for the propensity variables, and it includes the decision for the variables that have to be used for this procedure.

If and to what extent online panels can be used in the future for the collection of representative data for a general population will not only depend on the choice of respondents, but also on the subjects on which these

respondents are to be interviewed. The members of the online panel we interviewed in the context of this study did not differ any stronger from the general population in terms of age, sex and education than the members of the GSS, but they did differ strongly in their household composition, their marital status and in their religious and political values.

References

ADM (2008) Jahresbericht 2007 (annual report 2007). (http://www.adm-ev.de/fileadmin/user_ upload/PDFS/Jahresbericht_08.pdf).

Bandilla W, Bosnjak M, and 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*, 2l, 235–243.

Blohm M, Harkness J, Klein S and Scholz E (2003) Konzeption und Durchführung der "Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften" (ALLBUS) 2002 (Concept and implementation of the "General Population Survey of Social Sciences" (GSS)). *ZUMA-Methodenbericht*, 2003/12.

Couper MP (2001) The Promises and Perils of Web Surveys. In: Westlake A. (ed.) *The Challenge of the Internet*. London: Associations for Survey Computing, 35–56.

Couper MP, Tourangeao R, and Kenyon K (2004) Picture This! Exploring Visual Effects in Web Surveys. *Public Opinion Quarterly* 68: 255–266.

Couper MP, Kapteyn A, Schonlau M and Winter J (2007) Noncoverage and Non-response in an Internet Survey. *Social Science Research* 36: 131–148.

De Leeuw E (1992) *Data Quality in Mail, Telephone and Face-to-face Surveys*. Amsterdam: TT-publ.

De Leeuw J (2006) Nonlinear Principal Component Analysis and Related Techniques. In: Greenacre M and Blasius J (eds) *Multiple Correspondence Analysis and Related Methods*. Boca Raton FL: Chapman & Hall, 107–133.

Denscombe M (2006) Web-based Questionnaires and the Mode Effect. An evaluation based on completion rates and data contents of near-identical questionnaires delivered in different modes. *Social Science Computer Review* 24: 246–254.

Deutskens E, de Jong A, de Ruyter K and Wetzels M (2006) Comparing the Generalizability of Online and Mail Surveys in Cross-national Service Quality Research. *Marketing Letters* 17: 119–136.

Dever JA, Rafferty A and Valliant R (2008) Internet Surveys: Can Statistical Adjustments Eliminate Coverage Bias? *Survey Research Methods* 2: 47–60.

Dillman DA (2000) *Mail and Internet Surveys. The tailored Design Method*. New York: Wiley.

Duffy B, Smith K, Terhanian G and Bremer J (2005) Comparing Data from Online and Face-to-Face Surveys. *International Journal of Market Research* 47: 615–639.

Faas T (2003) Umfragen im Umfeld der Bundestagswahl 2002, Offline und Online im Vergleich (Polls in the context of the elections to the federal parliament 2002. A comparison of offline and online data). *ZA-Informationen* 52: 120–135.

Faas T and Schoen H (2006) Putting a Questionnaire on the Web is Not Enough – A comparison of online and offline surveys conducted in the context of the German federal election 2002. *Journal of Official Statistics* 22: 177–190.

Fricker RD, Jr and Schonlau M (2002) Advantages and Disadvantages of Internet Research Surveys. Evidence from the Literature. *Field Methods* 14: 347–367.

Gesell SB, Drain M, Clark PA and Sullivan MP (2007) Test of a Web and Paper Employee Satisfaction Survey: Comparison of Respondents and Non-Respondents. *International Journal of Internet Sciences* 2: 45–58.

Gifi A (1990) *Nonlinear Multivariate Analysis*. Chichester: Wiley.

Heerwegh D and Loosveldt G (2008) Face-to-Face versus Web Surveying in a High-Internet-Coverage Population: Differences in Response Quality. *Public Opinion Quarterly* 72: 836–846.

Hoogendorn AW and Daalmans J (2009) Nonresponse in the Requirement of an Internet Panel Based on Probability Sampling. *Survey Research Methods* 3: 59–72.

Kaplowitz MD, Hadlock TD and Levine R (2004) A Comparison of Web and Mail Survey Response Rates. *Public Opinion Quarterly* 68: 94–101.

Kwak N and Radler B (2002) A Comparison between Mail and Web Surveys: Response Pattern, Respondent Profile, and Data Quality. *Journal of Official Statistics* 18: 257–273.

Lee S (2006) Propensity Score Adjustment as a Weighting Scheme for Volunteer Panel Web Surveys. *Journal of Official Statistics* 22: 329–349.

Lee S and Valliant R (2009) Estimation for Volunteer Panel Web Surveys Using Propensity Score Adjustment and Calibration Adjustment. *Sociological Methods & Research* 37: 319–343.

Loosveldt G and Sonck N (2008) An Evaluation of the Weighting Procedures for an Online Access Panel Survey. *Survey Research Methodology* 2: 93–105.

Lozar Manfreda K and Vehovar V (2002) Do Mail and Web Surveys Provide the Same Results? In: Ferligoj A and Mrvar A (eds) Development in Social Science Methodology 18. Ljubljana: FDV, 149–169.

Michailidis G and De Leeuw J (1998) The Gifi System for Descriptive Multivariate Analysis. *Statistical Science* 13: 307–336.

Rookey BD, Hanway S and Dillman DA (2008) Does a Probability-based Household Panel Benefit from Assignment to Postal Response as an Alternative to Internet-Only? *Public Opinion Quarterly* 72: 962–984.

Rosenbaum PR and Rubin D (1984) Reducing Bias in Oberservational Studies Using Subclassification on the Propensity Score. *Journal of the American Statistical Association* 79: 516–524.

Schonlau M, Zapert K, Simon L, et al. (2004) A Comparison between Responses from a Propensity-weighted Web Survey and an Identical RDD Survey. *Social Science Computer Review* 22: 128–138.

Schonlau M, van Soest A and Kapteyn A (2007) Are "Webographic" or Attitudinal Questions Useful for Adjusting Estimates from Web Surveys Using Propensity Scoring? *Survey Research Methods* 1: 155–163.

Schonlau M, van Soest A, Kapteyn A and Couper M (2009) Selection Bias in Web Surveys and the Use of Propensity Scores. *Sociological Methods & Research* 37: 291–318.

Van Eimeren B and Frees B (2006) ARD/ZDF Online Study 2006. Schnelle Zugänge, neue Anwendungen, neue Nutzer? (Fast access points, new applications, new users?). *Media-Perspektiven* 8: 402–415.

Vehovar V, Lozar Manfreda K and Batagelj Z (1999) Web Surveys: Can the Weighting Solve the Problem? *Proceedings of the Survey Research Method Section, ASA 1999*. Available a: http:// www.anstat.org/sections/srms/proceedings