Article

Field Methods
25(4) 339-360

DOI: 10.1177/1525822X12472951
fm.sagepub.com
SAGE

Sample Composition Discrepancies in Different Stages of a Probability-based Online Panel

Michael Bosnjak[1], Iris Haas[2], Mirta Galesic[3], Lars Kaczmirek[2], Wolfgang Bandilla[2], and Mick P. Couper[4]

Abstract

We report sample composition discrepancies related to demographic and personality variables occurring in different stages of development of a probability-based online panel. The first stage—selecting eligible participants—produces differences between Internet users and nonusers in age, education, and gender distribution as well as in the personality traits of openness to experience, conscientiousness, and extraversion. The second and third stages of panel development—asking about willingness to

[1] School of Economics and Management, Free University of Bozen-Bolzano, Bolzano, Autonomous Province of South Tyrol, Italy
[2] GESIS—Leibniz Institute for the Social Sciences, Survey Design and Methodology, Mannheim, Germany
[3] Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Berlin, Germany
[4] Institute for Social Research, University of Michigan, Ann Arbor, MI, USA

Corresponding Author:
Michael Bosnjak, Free University of Bozen-Bolzano, School of Economics and Management, Piazza Universita 1, I- 39100 Bolzano, Autonomous Province of South Tyrol, Italy.
Email: michael.bosnjak@unibz.it

participate in the panel and actual participation in online surveys—result in fewer and smaller discrepancies. The results suggest that among the three potential sources of sample composition bias considered, the largest impact comes from coverage differences with regard to Internet access.

Keywords
probability-based online panels, nonresponse, attrition, sample composition

Introduction and Research Questions

Online access panels are frequently being used to collect survey data in various fields (Couper 2007; Göritz 2010; Scherpenzeel and Bethlehem 2011; Vis and Marchand 2011), including public opinion research (e.g., Smith 2003), and can be classified into two broad categories: nonprobability (volunteer) panels and prerecruited probability-based panels (Baker et al. 2010; Couper 2000; Couper and Bosnjak 2010).

In nonprobability (volunteer) panels, respondents recruit themselves and sign up to participate in online surveys regularly. These respondents might have become aware of the panel by any kind of referral or advertisement, such as banner ads, pointers placed on web sites, and word-of-mouth communication. Questions have been raised about the inferential value of such panels, most recently by Baker et al. (2010). The sampling frame and specific mechanisms of self-selection into the panel are unknown. As a consequence, design-based inferential statistics such as standard errors and confidence intervals make little sense (see Baker et al. 2010). Examples of such volunteer opt-in panels include those summarized on metasites such as yellowsurveys.com and money4surveys.com.

In contrast, in prerecruited probability-based online panels, there is a known nonzero probability of selection from a given sampling frame. In most cases, panel members are recruited through random-digit-dialing (RDD) telephone sampling or address-based sampling. Since there is knowledge about the sampling frame and the recruitment processes, coverage and nonresponse error can be estimated and then used to weight and adjust the resulting survey data (see e.g., Lee 2006), the success of which depends on the amount and quality of information on the frame. Examples of such panels are those operated by RAND Corporation (American Life Panel) and Knowledge Networks in the United States (Huggins and Eyerman 2001; Smith 2003) or the panels operated by CentERdata at Tilburg University in the Netherlands (e.g., the CentERpanel with RDD-type recruitment as described by Hoogendoorn and Daalmans 2009, and the Longitudinal Internet Studies for the Social Sciences (LISS) panel with a recruitment process based on mandatory resident registration data, as described by Knoef and de Vos 2009). To reduce coverage error, most of these panels provide respondents with a free computer and Internet access if necessary (for the LISS panel, the procedure is described in detail by Scherpenzeel and Das 2011). Because of their methodological advantages, we focus on prerecruited probability-based online panels.

Despite the relative advantage of probability-based online panels in comparison to their nonprobability-based counterparts, sample composition bias might still be present. Following Chang and Krosnick (2009), the term sample composition bias denotes the deviation of (online panel-based) sample characteristics in relation to a probability national sample such as that used in the General Social Survey in the United States. In Germany, the equivalent survey is ALLBUS, a large national survey on attitudes, behavioral patterns, and social structure. Data from these hour-long computer-assisted personal interviews (CAPI) are used to make important policy decisions. However, this expensive study can be conducted only once every 2 years. In this article, we describe the development of an online panel recruited from ALLBUS participants and investigate whether it can be used for short, relatively inexpensive studies on important topics emerging between the two waves of ALLBUS. Specifically, we ask two questions about the existence and nature of the sample composition bias of the online sample compared to the original ALLBUS sample.

First, in building and operating a probability-based online panel, which stages are most affected by which potential sources of sample composition bias? Following Couper et al. (2007), we distinguish three serially related stages potentially contributing to sample composition biases when building and operating a probability-based online panel: (1) access to, or use of the Internet among sample members; (2) willingness to participate in the panel; and (3) actual participation in online surveys conducted. While the first stage pertains to coverage error, the latter two address nonresponse issues. Therefore, various conceptually distinct sources of bias at different stages of building and operating an online panel may contribute to the resulting overall sample composition bias. One goal of this article is to conceptually disentangle and quantify potential differences during three selected stages of panel development, namely coverage-related differences (Internet users vs. nonusers), differences between those willing versus those not willing to participate in online panel surveys, and nonresponse-related differences (actual online panel survey participants vs. nonparticipants).

Second, for which variables do we find sample composition discrepancies? Past research has primarily focused on demographics and some specific content domains such as voting behavior (Chang and Krosnick 2009), environmental attitudes (Bandilla et al. 2003), attitudes toward and understanding of science and technology (Fricker et al. 2005), and other specific public opinion issues (Smith 2003). However, the spectrum of affected variables might be larger, encompassing more general psychological parameters such as personality traits, which can then affect specific survey topics.

Personality traits have been shown to be systematically related to political attitudes and voting behavior (e.g., Caprara and Zimbardo 2004; Saucier 2000), personal values (e.g., Parks and Guay 2009; Roccas et al. 2002), subjective well-being (e.g., DeNeve and Cooper 1998; Diener and Lucas 1999), physical health and longevity (e.g., Caspi et al. 2005), mental health (e.g., Trull and Durrett 2005), job performance (e.g., Barrick et al. 2002), consumer behavior (e.g., Bosnjak et al. 2007; Mowen 2000), and survey participation (e.g., Dollinger and Leong 1993; Lönnqvist et al. 2007; Marcus and Schütz 2005). Therefore, using personality traits as proxy variables for a broader range of substantive variables in sociology, psychology, political science, and public health appears reasonable and contributes to the understanding of how a broad set of substantive areas could be affected by sample composition discrepancies in probability-based online panel surveys.

Literature Review

A few studies of sample composition bias in probability-based online surveys exist. For instance, based on data from a two-wave survey conducted before and after the U.S. 2000 presidential election, Chang and Krosnick (2009) found systematic sample composition biases between the demographic profile of the probability-based online panel used (Knowledge Networks Panel [KNP]) and the Current Population Survey [CPS]). Relative to the CPS, the KNP sample underrepresented individuals with low education (high school or less), those under age 25 or over age 65, African American respondents, and the lowest-income individuals.

Similar data come from a German study (Bandilla et al. 2003). Conducted in 2000, this study compared—among content-related aspects such as environmental attitudes—demographic sample composition differences between the respondents of the representative German general population survey ALLBUS and those of a probability-based panel operated by FORSA, a public opinion research institute in Berlin. Overall, better-educated respondents (defined as being eligible to enter a university program) were considerably overrepresented (web: 70%, population: 23%); younger (ages 18–29; web: 40.4%, population: 16.4%) and middle age groups (ages 30–44; web: 44.8%, population: 31.1%) were overrepresented; and females were underrepresented (web: 33.9%, population: 51.8%).

Differences between the U.S. (Chang and Krosnick 2009) and German (Bandilla et al. 2003) findings might be attributable to various factors, such as the fact that both coverage biases and nonresponse errors may have been confounded in the Bandilla et al. (2003) study, and due to the time lag in Internet adoption in Germany compared to the United States in 2000. In 2000, the Internet penetration rate in Germany was 13% lower than in the United States (Germany: 29%; United States: 42%). By 2008, Internet adoption rates for the two countries were almost the same (National Telecommunications and Information Administration [NTIA] 2011; van Eimeren and Frees 2011).

Fricker et al. (2005:384, Table 3) found that compared to the U.S. adult population in 2003, the following characteristics were overrepresented in a probability-based panel web survey: female (web survey: 59.4%, population estimate: 51.1%), white (web survey: 89.2%, population: 80.7%), more highly educated (e.g., college graduates in the web survey: 48.2%, population: 24.7%), and middle-age segments (34–54 years of age; web survey: 47.4%, population: 39.6%). After adjusting for coverage bias using other demographic variables, these differences were reduced but not eliminated (female web survey: 59.4%, female online: 52.0%; graduated from college web survey: 48.2%, graduated from college online: 36.5; middle-age segments web survey: 47.4%, middle-age segments online: 46.3%).

Coverage error is reduced (but not necessarily eliminated) in two probability-based CentERdata panels in the Netherlands (because those who did not have access to the Internet were provided with the required equipment, see, e.g., Hoogendoorn and Daalmans [2009] for the CentERpanel and Knoef and de Vos [2009] for the LISS panel). However, sample composition bias due to nonresponse appears to be still present. In the CentERpanel, Hoogendoorn and Daalmans (2009) found that in all stages of panel recruitment such as (1) responding to an initial computer-assisted telephone interviewing recruitment survey; (2) expressing the intention to participate; and (3) actually becoming a member of the panel, demographic variables (e.g., age, gender, ethnicity, and income) have a small but significant influence on each of the selection steps. In comparison to representative national statistics, Knoef and de Vos (2009) found for the LISS panel that the elderly are underrepresented, especially elderly women (LISS panel: 4.0% women 65 or older, population: 7.8%). In all other age segments, females tend to be slightly overrepresented, paralleling the results reported by Fricker et al. (2005).

The present study augments the previous literature by investigating separately three possible sources of sample composition biases: access to the Internet, willingness to participate in the panel, and willingness to participate in surveys conducted on the panel. This information can help survey researchers identify stages of panel development that are most critical to the overall bias and consequently require the largest input of time and resources. In addition, in the present article, we focus not only on demographic biases as did most of the previous studies but specifically on the personality traits that may underline a host of otherwise unrelated attitudes and behaviors. The findings can guide researchers in interpreting results obtained in online samples.

We first describe the methods used in recruiting a probability-based online panel and in administering the first two surveys within the panel. We also explain how we measured the key variables, namely demographic variables and personality traits. Next, we summarize the results on the coverage- and nonresponse-related differences in sample composition at different steps in recruiting and operating the probability-based online panel. Finally, we explore sample composition biases on both demographic *and* personality variables.

Method

Sample and Procedure

Online panelists were recruited with the aid of the German General Social Survey (ALLBUS) conducted between spring and summer 2008. ALLBUS is a biennial survey that has been in place since 1980 on the attitudes, behavior, and social structure of residents in Germany. A representative cross section of the population is questioned using, on average, 1-hour face-to-face interviews (CAPI).

The sample was drawn in a two-stage design of all German-speaking adults (age $\geq$ 18) living in private households. In the first stage, communities were selected with a probability proportional to the number of adult residents; in the second stage, individuals were randomly selected from the community registers. Sampling, data collection, and editing were conducted by TNS Infratest, the fieldwork agency for ALLBUS. Participation in the survey was voluntary and participants were not financially rewarded.

In 2008, the final ALLBUS sample encompassed 3,469 persons and had—using AAPOR's (2011) response rate type 1 (RR1) definition—a response rate of 40%, which is in line with former studies based on register samples. There are no large deviations between ALLBUS and the 2007 German Microcensus, so weighted and unweighted distributions are very similar. We use the unweighted ALLBUS data for our analyses. Full details of the sample including comparisons to the microcensus are presented in the technical report on ALLBUS 2008 (Wasmer et al. 2010).

At the end of the ALLBUS interview, all participants were asked whether they have Internet access at home or not. After that, all participants with Internet access were asked whether they were willing to participate in the online panel or not. Those ALLBUS 2008 participants who reported having Internet access at home and were willing to participate in "scientific online surveys regularly conducted by GESIS Leibniz Institute for the Social Sciences" were included in the online panel. Of the 3,469 respondents, 1,865 reported having Internet access at home (54% of ALLBUS 2008 respondents). Of these, 812 Internet users (44% of the 1,865 Internet users) expressed willingness to participate in the online panel. Among those willing to join the panel, 223 persons were screened out for several reasons (invalid email addresses [$n = 27$] and a random subgroup used for a different online survey that is not reported here [$n = 196$]), resulting in 589 initial panel members.

The initial panel members were invited to two online surveys, fielded from November to December 2008 (wave 1) and from March to April 2009 (wave 2). Figure 1 depicts the flow of participants of the study.

The first survey was about "health and life satisfaction." Of the 589 initial panel members, 260 persons participated in this first wave (RR2 = 44%). Only 483 of the 589 panelists (82% of the initial panel members) were invited to the second wave, which was about "Internet usage and privacy concerns." The reasons for this panel attrition were that, following the first wave, 40 persons were no longer reachable and 66 opted out of the panel. Of the 483 invited to the second wave, 250 responded to the survey (RR2 = 52%).

Following these recruitment steps, four types of panel participation can be distinguished: (1) 153 panelists took part in both waves (26% of the initial panel members); (2) 107 responded only to the first survey (18% of the initial panel members; $M = 42.96$ years, $SD = 14.39$ years; 51% men;

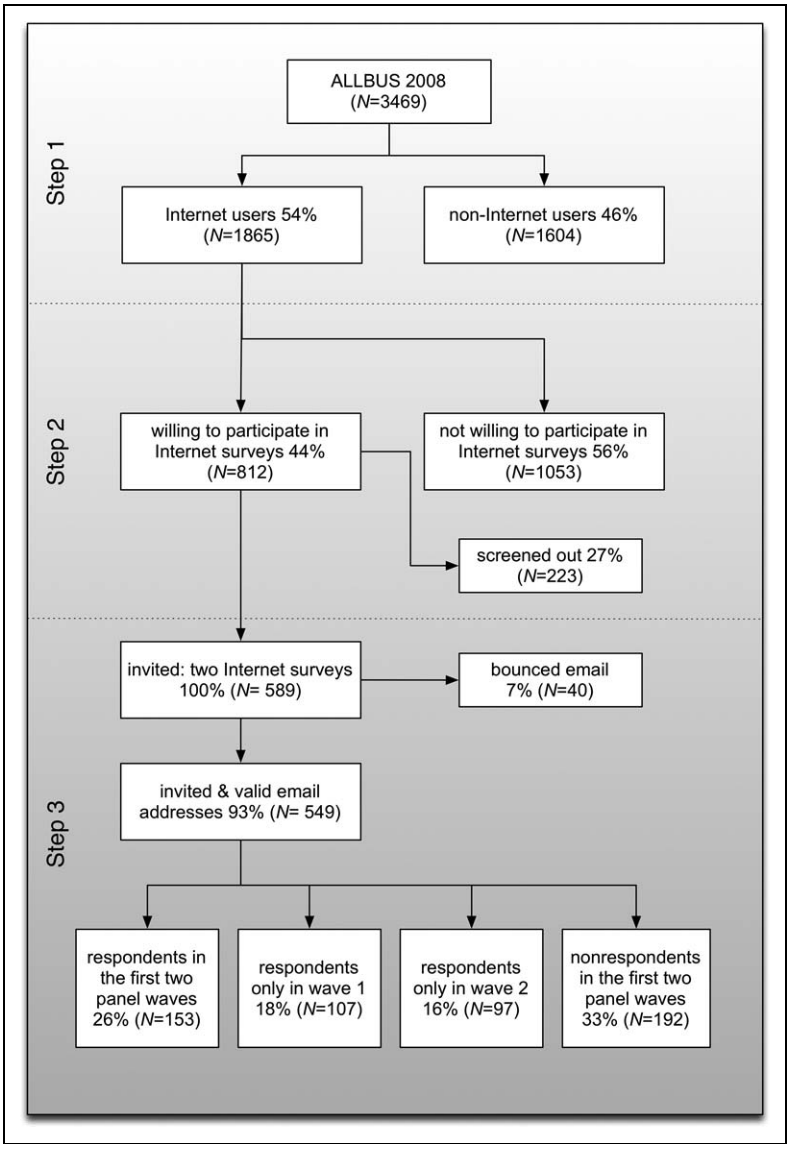

Figure 1. Recruitment procedure.

educational level: 13% low, 31% medium, 55% high); (3) 97 responded only to the second survey (16% of the initial panel members; $M = 41.86$ years, $SD = 14.25$ years; 53% men; educational level: 17% low, 31% medium, 52% high); and (4) 192 responded to neither survey (33% of the initial panel members).

To isolate the effects of demographic characteristics and personality on (complete) participation versus nonparticipation, those subjects who participated only partially (i.e., groups 2 and 3 as described above) are excluded from the analysis. Therefore, we used extreme groups (participants to both waves vs. nonrespondents to both waves), assuming that if we do not find substantive differences for these extreme groups in terms of sample composition, we should not worry about possible attrition steps in between. Sociodemographic characteristics of the two groups considered in the analysis are reported in Tables 1–3.

Measures

The most accepted descriptive taxonomy for human personality traits is based on a five-dimensional model, also known as the Big Five (Goldberg 1990) or the five-factor model of personality (McCrae et al. 1996). In most contemporary approaches (e.g., De Raad 2000; John and Srivastava 1999; Ostendorf and Angleitner 2004), the five personality trait dimensions are termed openness to experience (O), conscientiousness (C), extraversion (E), agreeableness (A), and neuroticism (N).

The openness factor captures individual differences in active seeking and appreciation of experiences for their own sake. People scoring high on the O-factor are dispositionally open to their inner feelings and emotions and value new experiences, ideas, impressions, and aesthetics.

Conscientiousness focuses on degree of organization, persistence, control, and motivation in goal-directed behaviors. People scoring high on the C-factor stress fulfilling moral obligations; they are self-disciplined, deliberate, well organized, punctual, and believe in their own self-efficacy.

The central issues of the extraversion dimension lies in the quantity and intensity of energy directed outward into the social world. High scores on the E-factor are, for instance, associated with assertiveness, social activity, and the tendency to seek environmental stimulation.

Agreeableness assesses the kinds of interactions an individual prefers, ranging from compassion to tough mindedness. People scoring high on the A-factor are compliant, trusting, altruistic, modest, and kind hearted.

Table 1. Coverage-related Differences between Internet Users and Nonusers for Personality and Selected Demographic Variables.

		ALLBUS ($N = 3{,}469$)						
		Internet Users ($N = 1{,}865$)			Nonusers ($N = 1{,}604$)			Group Contrasts (Cohen's *d*)
		M	*SD*	95% CI	*M*	*SD*	95% CI	
Personality traits	Openness	3.61	0.83	[3.57, 3.65]	3.30	0.93	[3.25, 3.34]	**0.35**
	Conscientiousness	4.03	0.71	[4.00, 4.06]	4.24	0.70	[4.20, 4.27]	−**0.30**
	Extraversion	3.38	0.88	[3.33, 3.42]	3.18	0.95	[3.14, 3.23]	**0.22**
	Agreeableness	3.18	0.75	[3.15, 3.22]	3.31	0.80	[3.27, 3.35]	−0.17
	Emotional stability	3.47	0.84	[3.43, 3.51]	3.49	0.93	[3.44, 3.53]	−0.02
Demographics	Age	42.3	14.32	[41.7, 43.0]	60.6	16.32	[59.8, 61.4]	−**1.19**
			%			%		Hasselblad and Hedges (1995) *d**
	Is male		54.7			43.1		**0.26**
	High education		41.6			12.7		**0.90**

Notes. *d* values larger than .2 are givenin boldface, denoting at least small effects according to Cohen (1992). Values above .4 are classified as practically meaningful according to Ferguson (2009).
High education means that these participants have at least a general higher education entrance qualification (so-called Abitur).

Neuroticism describes differences between emotional sensitivity on one hand, and emotional stability on the other. Neurotic individuals are more prone to psychological distress such as anxiety, impulsivity, nervousness, and vulnerability.

The Big Five personality traits were an integral part of ALLBUS and measured using the 10-item personality scale (Big Five Inventory–10 [BFI-10]; Rammstedt and John 2007). This short-scale version[1] of the well-established BFI was developed to provide a personality inventory for research settings with extreme time constraints (Rammstedt 2007). In general social surveys like ALLBUS, an extremely brief measure is needed. Traditional BFIs with 20 items and more are much too lengthy. All the questions in ALLBUS surveys are intended for replication in different years, so there is a strong limitation on the inclusion of new items. Among the standard questions for replication in 2008, the ALLBUS questionnaire included the short BFI-10 version, which assesses the five dimensions using only two items for each dimension.

Results

In accordance with Couper et al. (2007), we distinguish three serially related recruitment steps in the early stages developing a probability-based online panel: (1) access to, or use, of the Internet among the sample members (a coverage-type bias); (2) willingness to participate in the panel (a source of nonresponse error); and (3) actual participation to the online surveys conducted (another source of nonresponse error). In line with our study goals, in each recruitment step, we describe sample composition differences in terms of the demographic characteristics age, gender, and education as well as the personality traits of openness, conscientiousness, extraversion, agreeableness, and emotional stability.

As effect size measures, we use Cohen's (1992) *d* for mean personality differences and Hasselblad and Hedges's (1995) *d** for differences in percentages on demographic variables. Conceptually, *d* values are computed by forming a contrast between two central tendency measures (e.g., mean differences for *d* according to Cohen 1992) and by dividing this contrast by a measure of dispersion (e.g., pooled standard deviation of two samples compared for the Cohen's *d* metric, for instance). Therefore, *d*-type effect size measures allow for estimating the magnitude or strength of sample parameter differences regardless of sample sizes. To classify the magnitude of these two types of *d* values, we refer to Cohen (1992), who denoted *d* values of .2 as small, those around .5 as medium, and those of .8 or higher as large effects. Ferguson (2009) recommended treating effects larger than .4 as those representing "practically meaningful" differences for social science data. Differences above this threshold are expected to be consistently replicable in practice.

First Selection Stage: Coverage-related Differences

Starting from those who responded to the ALLBUS 2008 survey (N = 3,469), Table 1 summarizes the results for the differences in demographic characteristics and personality traits between Internet users, who are potentially eligible for the panel, and those who reported not having Internet access at home.

As Table 1 shows, we find small to medium differences for three of the five personality traits: Internet users tend to have higher scores on openness ($d = .35$), to be less conscientious ($d = -.30$), and to be more extraverted ($d = .22$). However, practically meaningful differences according to Ferguson (2009) are found for demographics, especially for age. Nonusers of the Internet are considerably older ($M = 60.6$ compared to 42.3 for Internet users, $d = -1.19$), have a higher level of formal education ($d^* = .90$), and are more often male ($d^* = .26$). Overall, the findings suggest that coverage-related sample composition differences tend to be large for demographics and smaller for personality traits.

Second Selection Stage: Differences Related to Willingness to Participate in the Panel

During the ALLBUS 2008 survey, participants were asked if they would be willing to become members of an online panel regularly conducting scientific studies. Table 2 summarizes sample composition differences for those who agreed to participate compared to those who declined to do so. Possible differences on this level may contribute to nonresponse error.

As Table 2 shows, we found only a few differences, all of them small in size. Those willing to participate appeared to be less conscientious ($d = -.28$), slightly younger ($d = -.21$), and had a higher level of formal education ($d = .30$). Once we make Internet access a condition, the personality trait differences between those willing to join the panel and those not willing become smaller. This is consistent with the findings of Couper et al. (2007).

Table 2. Differences between Those Willing and Those not Willing to Participate in Online Surveys for Personality and Selected Demographic Variables.

		Willing to Participate in Online Panel Surveys ($N = 812$)			Not Willing to Participate in Online Panel Surveys ($N = 1{,}053$)			Group Contrasts (Cohen's *d*)
		M	*SD*	95% CI	*M*	*SD*	95% CI	
Personality traits	Openness	3.65	0.83	[3.60, 3.71]	3.58	0.83	[3.53, 3.63]	0.08
	Conscientiousness	3.92	0.72	[3.87, 3.97]	4.12	0.70	[4.08, 4.16]	−**0.28**
	Extraversion	3.39	0.87	[3.33, 3.45]	3.36	0.89	[3.31, 3.42]	0.03
	Agreeableness	3.14	0.75	[3.09, 3.19]	3.22	0.75	[3.17, 3.26]	−0.10
	Emotional stability	3.46	0.82	[3.40, 3.52]	3.47	0.86	[3.42, 3.53]	−0.01
Demographics	Age	40.6	14.56	[39.6, 41.6]	43.6	14.00	[42.8, 44.5]	−**0.21**
			%			%		Hasselblad and Hedges (1995) *d**
	Is male		57.1			52.9		0.09
	High education		50.1			36.8		**0.30**

Note: d values larger than .2 are given in boldface, denoting at least small effects according to Cohen (1992). Values above .4 are classified as practically meaningful according to Ferguson (2009).
High education means that these participants have at least a general higher education entrance qualification (so-called Abitur).

Table 3. Respondents versus Nonrespondents in the First Two Panel Survey Waves.

		Respondents in the First Two Panel Survey Waves (N = 153)			Nonrespondents in the First Two Panel Survey Waves (N = 192)			Group Contrasts (Cohen's d)
		M	*SD*	95% CI	*M*	*SD*	95% CI	
Personality traits	Openness	3.67	0.81	[3.54, 3.80]	3.64	0.81	[3.52, 3.75]	0.04
	Conscientiousness	3.89	0.70	[3.76, 3.98]	3.93	0.75	[3.82, 4.03]	−0.06
	Extraversion	3.17	0.83	[3.03, 3.30]	3.38	0.90	[3.25, 3.52]	−**0.24**
	Agreeableness	3.19	0.71	[3.08, 3.30]	3.13	0.76	[3.02, 3.24]	0.08
	Emotional stability	3.42	0.85	[3.28, 3.55]	3.51	0.74	[3.40, 3.61]	−0.11
Demographics	Age	44.3	14.01	[42.1, 46.6]	40.3	14.83	[38.14, 42.37]	**0.28**
			%			%		Hasselblad and Hedges (1995) d*
	Is male		56.9			55.7		0.03
	High education		56.3			40.7		**0.35**

Note: d values larger than .2 are given in boldface, denoting at least small effects according to Cohen (1992). Values above .4 are classified as practically meaningful according to Ferguson (2009).
High education means that these participants have at least a general higher education entrance qualification (so-called Abitur).

Third Selection Stage: Respondents versus Nonrespondents in the First Two Panel Survey Waves

Finally, among those who signed up for the panel, only a subset participated in both of the first two panel survey waves. Sample composition differences for respondents versus nonrespondents to both waves are summarized in Table 3.

As seen in Table 3, better educated ($d^* = .35$) and older ($d = .28$) panelists tend to be overrepresented among the respondents compared to nonrespondents. Moreover, respondents tend to be less extraverted ($d = -.24$). No other differences were statistically significant.

Sample Composition Bias: ALLBUS Participants versus Respondents in the First Two Panel Survey Waves

Table 4 compares the group of respondents to the ALLBUS 2008 survey ($N = 3{,}469$) with those who responded to both online panel waves ($N = 153$), enabling quantification of sample composition bias for selected sample characteristics.

Table 4 shows substantial sample composition bias for age (respondents to the online panel survey waves are younger compared to the ALLBUS benchmark, 44.3 vs. 50.8 years, $d = .41$) and education (online panel participants are better educated, $d = .64$). Moreover, substantial differences were found for two personality traits: Respondents to the two online panel waves had higher scores on openness ($d = .21$) and lower scores on conscientiousness ($d = .33$) compared to the ALLBUS benchmark.

Summary and Conclusions

The increasing use of probability-based online panels in survey research makes it important to identify and measure any sample composition biases they may be associated with. While previous research has found evidence for such biases on demographic variables, the aim of this study was to identify *where* possible biases are introduced during development of a probability-based online access panel. Moreover, our study extends previous research by looking at sample composition differences related to both demographic *and* personality variables occurring at different stages of development of a probability-based online panel in Germany.

When looking at sample composition bias, defined here in accordance to Chang and Krosnick (2009) as the deviation of the (online panel-based)

Table 4. ALLBUS Participants versus Respondents in the First Two Panel Survey Waves.

		ALLBUS Participants (N = 3,469)			Respondents in the First Two Panel Survey Waves (N = 153)			Group Contrasts (Cohen's d)
		M	SD	95% CI	M	SD	95% CI	
Personality traits	Openness	3.49	0.89	[3.44, 3.50]	3.67	0.81	[3.54, 3.80]	−**0.21**
	Conscientiousness	4.12	0.71	[4.10, 4.15]	3.89	0.70	[3.76, 3.98]	**0.33**
	Extraversion	3.29	0.92	[3.26, 3.32]	3.17	0.83	[3.03, 3.30]	0.14
	Agreeableness	3.24	0.78	[3.21, 3.27]	3.19	0.71	[3.08, 3.30]	0.07
	Emotional stability	3.48	0.88	[3.45, 3.51]	3.42	0.85	[3.28, 3.55]	0.07
Demographics	Age	50.8	17.79	[50.17, 51.36]	44.3	14.01	[42.1, 46.6]	**0.41**
		%			%			Hasselblad and Hedges (1995) d*
	Is male	49.4 (1,712 of 3,469)			56.9 (87 of 153)			−0.17
	High education	28.7 (977 of 3406)			56.3 (85 of 151)			−**0.64**

Note: d values larger than .2 are given in boldface, denoting at least small effects according to Cohen (1992). Values above .4 are classified as practically meaningful according to Ferguson (2009).
High education means that these participants have at least a general higher education entrance qualification (so-called Abitur).

sample characteristics in relation to a probability national sample survey, we find substantial differences for age and education. Moreover, the two personality traits openness and conscientiousness do differ.

When looking at the different stages of development of a probability-based online panel, it became evident that the first stage—selecting eligible participants—produces the most differences and the largest effect sizes. People who are Internet users tend to be much younger, better educated, and more often male than nonusers. Besides these expected demographic differences reflecting differences in socioeconomic status and access to technology, we also find differences in personality traits. Internet users are higher in openness to experience and extraversion and lower in conscientiousness. These personality differences might be related to the younger age of this group. A recent study on a national German sample has shown that younger people tend to have higher openness to experience and extraversion and lower conscientiousness than older people (Donnellan and Lucas 2008). However, controlling for age, education, and gender in multivariate models, the relationships between the personality factors and Internet use persist.

The second stage of panel development—willingness to participate in the panel—is associated with few significant differences and smaller effects. Younger and better-educated Internet users are more likely to agree to join an online panel, although the sizes of these effects are much smaller than in the first stage. These people are possibly more technologically skilled than older Internet users and thus more comfortable with completing online surveys. In terms of personality traits, we find differences in conscientiousness: People who are willing to participate in a panel are less conscientious than those who are not. In a seeming contradiction, other studies have found that respondents to mail and online surveys tend to score higher on conscientiousness than nonrespondents (Kanuk and Berenson 1975; Rogelberg et al. 2003). However, in this stage of panel development, participants were only agreeing in principle to participate in further studies, not yet actually participating. Therefore, participants who are more conscientious might have been more concerned about their ability to follow up on their commitment and hence more likely to decline participation in the panel.

In the third stage of panel development, in which members were invited to actually participate in surveys, we again find only a few sample composition biases. Panel members who participated in both surveys tend to be a bit older than those who did not participate in either survey, which actually reduced the overall age bias compared to people who were not eligible for participation in the panel. On the other hand, the education bias increased further in this third stage: Panel members who participated in both surveys were somewhat better educated than those who did not participate. In terms of personality traits, the only small bias we found was related to extraversion: Participants in both surveys were somewhat less extraverted, thus reducing the initial bias toward higher extraversion that occurred in the first stage.

In sum, these results speak in favor of putting extra effort into reducing sample composition biases in the first stage of development of probability-based online panels. One solution could be to equip nonusers of the Internet with technology that enables them to complete online studies (Baker et al. 2010). This could reduce the major biases in age and education (Hoogendoorn and Daalmans 2009). Additional but smaller reduction in demographic biases could be achieved by increasing the motivation of Internet users to become members of the panel. Biases in personality traits that persist to the last stage of panel development (participating in surveys) are related to openness to experience and conscientiousness. They should be taken into account when investigating survey topics that could be affected by dispositions such as political attitudes and consumer behavior.

Limitations

Our findings about differences in personality traits of the participants in different stages of panel recruitment should be used with caution. Our study was exploratory in nature and did not start from strong theoretical predictions about personality differences we could expect. Furthermore, because of the tight limits on the amount of time that we could spend examining personality traits in the ALLBUS survey, we had to rely on a short form of the BFI that used only two items for each of the five dimensions.

Further studies could use more comprehensive personality inventories to examine the replicability of our results. Future research could also investigate the possibility of reducing the personality differences between participants and nonparticipants by constructing weights based on demographic variables. Numerous weighting algorithms exist (Bethlehem 2002; Gelman and Carlin 2002), and it is likely that different procedures would have different influence on biases in personality traits. If our findings on the differences in personality traits hold in further studies, careful investigation could determine what demographic variables and what weighting schemes would be most successful in reducing personality biases.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Funding for this research came from the German Science Foundation (Deutsche Forschungsgemeinschaft), research grant BA 3627/1-1.

Note

1. Limitations and weakness of the short BFI version in comparison to traditional Big Five inventories with more items are discussed by Rammstedt and John (2007) and Rammstedt et al. (2010).

References

Baker, R., S. J. Blumberg, J. M. Brick, M. P. Couper, M. Courtright, J. M. Dennis, D. A. Dillman, et al. 2010. Research synthesis: AAPOR report on online panels. *Public Opinion Quarterly* 74:711–81.

Bandilla, W., M. Bosnjak, and P. Altdorfer. 2003. Survey administration effects? A comparison of web-based and traditional written self-administered surveys using the ISSP environment module. *Social Science Computer Review* 21:235–43.

Barrick, M. R., G. L. Stewart, and M. Piotrowski. 2002. Personality on job performance: Test of the mediating effects of motivation among sales representatives. *Journal of Applied Psychology* 87:43–51.

Bethlehem, J. G. 2002. Weighting nonresponse adjustments based on auxiliary information. In *Survey nonresponse*, eds. R. M. Groves, D. A. Dillman, J. L. Eltinge, and R. J. A. Little, 275–88. New York: Wiley.

Bosnjak, M., M. Galesic, and T. Tuten. 2007. Personality determinants of online shopping: Explaining online purchase intentions using a hierarchical approach. *Journal of Business Research* 60:597–605.

Caprara, G. V., and P. G. Zimbardo. 2004. Personalizing politics: A congruency model of political preference. *American Psychologist* 59:581–94.

Caspi, A., B. W. Roberts, and R. L. Shiner. 2005. Personality development: Stability and change. *Annual Review of Psychology* 56:453–84.

Chang, L.-C., and J. A. Krosnick. 2009. National surveys via RDD telephone versus the Internet: Comparing sample representativeness and response quality. *Public Opinion Quarterly* 73:641–78.

Cohen, J. 1992. A power primer. *Psychological Bulletin* 112:155–59.

Couper, M. P. 2000. Web surveys: A review of issues and approaches. *Public Opinion Quarterly* 64:464–94.

Couper, M. P. 2007. Issues of representation in eHealth research (with a focus on web surveys). *American Journal of Preventive Medicine* 32: S83–89.

Couper, M. P., and M. Bosnjak. 2010. Internet surveys. In *Handbook of survey research*, eds. P. V. Marsden and J. D. Wright, 527–50. Bingley, UK: Emerald Group Publishing.

Couper, M. P., A. Kapteyn, M. Schonlau, and J. Winter. 2007. Noncoverage and nonresponse in an Internet survey. *Social Science Research* 36:131–48.

DeNeve, K. M., and H. Cooper. 1998. The happy personality: A meta-analysis of 137 personality traits and subjective well-being. *Psychological Bulletin* 124:197–229.

De Raad, B. 2000. *The Big Five personality factors*. Toronto, Canada: Hogrefe & Huber Publishers.

Diener, E., and R. E. Lucas. 1999. Personality and subjective well-being. In *Well-being: The foundations of hedonic psychology*, eds. D. Kahneman, E. Diener, and N. Schwarz, 213–29. New York: Russell-Sage.

Dollinger, S. J., and F. T. L. Leong. 1993. Volunteer bias and the five-factor model. *The Journal of Psychology* 127:29–36.

Donnellan, M. B., and R. E. Lucas, 2008. Age differences in the Big Five across the life span: Evidence from two national samples. *Psychology and Aging* 23:558–66.

Ferguson, C. J. 2009. An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice* 40:532–38.

Fricker, S., M. Galesic, R. Tourangeau, and T. Yan. 2005. An experimental comparison of web and telephone surveys. *Public Opinion Quarterly* 69:370–92.

Gelman, A., and J. B. Carlin. 2002. Poststratification and weighting adjustments. In *Survey nonrersponse*, eds. R. M. Groves, D. A. Dillman, J. L. Eltinge, and R. J. A. Little, 289–302. New York: Wiley.

Goldberg, L. R. 1990. An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Personality* 59:1216–29.

Göritz, A. 2010. Using lotteries, loyalty points, and other incentives to increase participant response and completion. In *Advanced methods for conducting online behavioral research*, eds. S. Gosling and A. Johnson, 219–33. Washington, DC: American Psychological Association.

Hasselblad, V., and L. V. Hedges. 1995. Meta-analysis of screening and diagnostic test. *Psychological Bulletin* 117:167–78.

Hoogendoorn, A. W., and J. Daalmans. 2009. Nonresponse in the recruitment of an Internet panel based on probability sampling. *Survey Research Methods* 3:59–72.

Huggins, V. J., and J. Eyerman. 2001. Probability based Internet surveys: A synopsis of early methods and survey research results. Paper presented at the Research Conference of the Federal Committee on Statistical Methodology, Arlington, VA, November 14–16, 2001. http://www.fcsm.gov/01papers/Huggins.pdf (accessed October 30, 2012).

John, O. P., and S. Srivastava. 1999. The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In *Handbook of personality: Theory and* *research* (2nd ed.), eds. L. A. Pervin and O. P. John, 102–38. New York: Guilford Press.

Kanuk, L., and C. Berenson. 1975. Mail surveys and response rates: A literature review. *Journal of Marketing Research* 12:440–53.

Knoef, M., and K. de Vos. 2009. *The representativeness of the LISS panel*. Tilburg, the Netherlands: CentERdata.

Lee, S. 2006. An evaluation of nonresponse and coverage errors in a prerecruited probability web panel survey. *Social Science Computer Review* 24:460–75.

Lönnqvist, J.-E., S. Paunonen, M. Verkasalo, S. Leikas, A. Tuulio-Henriksson, and J. Lönnqvist. 2007. Personality characteristics of research volunteers. *European Journal of Personality* 21:1017–30.

McCrae, R. R., A. B. Zonderman, P. T. Costa, M. H. Bond, and S. V. Paunonen. 1996. Evaluating replicability of factors in the revised NEO Personality Inventory: Confirmatory factor analysis versus procrustes rotation. *Journal of Personality and Social Personality* 70:552–66.

Marcus, B., and A. Schütz. 2005. Who are the people reluctant to participate in research? Personality correlates of four different types of nonresponse as inferred from self- and observer ratings. *Journal of Personality* 73:959–84.

Mowen, J. C. 2000. *The 3M model of motivation and personality: Theory and empirical applications to consumer behavior*. Boston: Kluwer Academic.

NTIA (National Telecommunications and Information Administration). 2011. *Digital nation: Expanding Internet usage*. Washington, DC: U.S. Department of Commerce.

Ostendorf, F., and A. Angleitner. 2004. *Neo-Persönlichkeitsinventar (revidierte Form, NEO-PI-R) nach Costa und McCrae* (The NEO personality inventory, revised form, NEO-PI-R according to Costa and McCrea). Göttingen, Germany: Hogrefe.

Parks, L., and R. Guay. 2009. Personality, values, and motivation. *Personality and Individual Differences* 47:675–84.

Rammstedt, B. 2007. The 10-item Big Five inventory—Norms values and investigation of sociodemographic effects based on a German population representative sample. *European Journal of Psychological Assessment* 23:193–201.

Rammstedt, B., L. R. Goldberg, and I. Borg. 2010. The measurement equivalence of Big-Five factor markers for persons with different levels of education. *Journal of Research in Personality* 44:53–61.

Rammstedt, B., and O. P. John. 2007. Measuring personality in one minute or less: A 10-item short version of the Big Five inventory in English and German. *Journal of Research in Personality* 41:203–12.

Roccas, S., L. Sagiv, S. H. Schwartz, and A. Knafo. 2002. The Big Five personality factors and personal values. *Personality and Social Psychology Bulletin* 28: 789–801.

Rogelberg, S. G., J. M. Conway, M. E. Sederburg, C. Spitzmueller, S. Aziz, and W. E. Knight. 2003. Profiling active and passive nonrespondents to an organizational survey. *Journal of Applied Psychology* 88:1104–14.

Saucier, G. 2000. Isms and the structure of social attitudes. *Journal of Personality and Social Psychology* 74:366–85.

Scherpenzeel, A. C., and J. G. Bethlehem. 2011. How representative are online panels? Problems of coverage and selection and possible solutions. In *Social and behavioral research and the Internet: Advances in applied methods and research strategies*, eds. M. Das, P. Ester, and L. Kaczmirek, 105–32. New York: Routledge.

Scherpenzeel, A. C., and M. Das. 2011. "True" longitudinal and probability-based Internet panels: Evidence from the Netherlands. In *Social research and the Internet*, eds. M. Das, P. Ester, and L. Kaczmirek, 77–104. New York: Taylor and Francis.

Smith, T. W. 2003. An experimental comparison of knowledge networks and the GSS. *International Journal of Public Opinion Research* 15:167–79.

The American Association for Public Opinion Research (AAPOR). 2011. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 7th edition. AAPOR.

Trull, T. J., and C. Durrett. 2005. Categorical and dimensional models of personality disorder. *Annual Review of Clinical Psychology* 1:355–80.

van Eimeren, B., and B. Frees. 2011. Drei von vier Deutschen im Netz – ein Ende des digitalen Grabens in Sicht? Ergebnisse der ARD/ZDF-Onlinestudie 2011 (Three out of four Germans are online – Is the end of the digital devide in sight? Results of the ARD/ZDF online study 2011). *Media Perspektiven* 7/8:334–349.

Vis, C. M., and M. Marchand. 2011. Challenges in reaching hard-to-reach groups in Internet panel research. In *Social and behavioral research and the Internet: Advances in applied methods and research strategies*, eds. M. Das, P. Ester, and L. Kaczmirek, 271–90. New York: Routledge.

Wasmer, M., E. Scholz, and M. Blohm. 2010. Konzeption und Durchführung der Allgemeinen Bevölkerungsumfrage für die Sozialwissenschaften ALLBUS 2008 (Conceptualization and implemetation of the German General Social Survey ALLBUS 2008). *GESIS Technical Report* 2010/04. http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2010/TechnicalReport_10-4.pdf (accessed October 30, 2012).