Social Networks 40 (2015) 75–83

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

Methodological considerations in the use of name generators and interpreters

David E. Eagle[a,*], Rae Jean Proeschold-Bell[b]

[a] Department of Sociology, Duke University, PO Box 90088, Durham, NC 27708-0088, United States
[b] Duke Global Health Institute, Duke University, PO Box 90392, Durham, NC 27708-0392, United States

ARTICLE INFO

Keywords:
Name generators
Survey design
Longitudinal design
Panel conditioning
Interviewer effect
Clergy

ABSTRACT

With data from the Clergy Health Initiative Longitudinal Survey, we look for interviewer effects, differences between web and telephone delivery, and panel conditioning bias in an "important matters" name generator and interpreter, replicated from the U.S. General Social Survey. We find evidence of phone interviewers systematically influencing the number of confidants named, we observe that respondents assigned to the web survey reported a larger number of confidants, and we uncover strong support for panel conditioning. We discuss the possible mechanisms behind these observations and conclude with a brief discussion of the implications of our findings for similar studies.

1. Introduction

Survey researchers commonly use name generators and interpreters to generate a list of a respondent's closest confidants and their characteristics.[1] The U.S. General Social Survey (GSS) employs a popular approach, which asks respondents to report the names of all those people with whom they discussed important matters in the past six months. Following the name generator item, the GSS proceeds with a series of name interpreter questions, which collects information on the characteristics of the first five people named (Burt, 1984; Marsden, 1987). While the use of name generator items is a common method to collect information about respondent social networks, researchers have uncovered important methodological issues surrounding their use (adams and Moody, 2007; Campbell and Lee, 1991; Ferligoj and Hlebec, 1999; Hammer, 1984; Hlebec and Ferligoj, 2002; Kogovšek, 2006; Kogovšek and Ferligoj, 2005; Kogovšek et al., 2002; Kogovsek and Hlebec, 2009; Manfreda et al., 2004; Marsden, 1993, 2003; Matzat and Snijders, 2010; Van Tilburg, 1998; Zemljič and Hlebec, 2005).

For example, McPherson et al. (2008) discovered that, from 1985 to 2004, the discussion networks of Americans had shrunk significantly. This finding was met with skepticism by some (including the study's own authors) and was later revealed to be an artifact of the data collection process (Fischer, 2009; McPherson et al., 2006, 2008; Paik and Sanchagrin, 2013). Several of the interviewers, knowing that for every name given by respondents they would be forced to ask another long series of questions, simply skipped the section and reported the respondent as having no close confidants. Although not all studies have been subject to interviewer-induced error as egregious as this example, other research has shown that these types of questions are particularly prone to "interviewer effects," which refer to the tendency for answers to vary depending on the interviewer assigned to the case (Groves and Magilavy, 1986). These effects stem from the tone and manner in which interviewers ask questions and whether or how they prompt respondents for additional responses (Hox, 1994). Of the several studies that have looked for an interviewer effect on discussant network size, all of them found systematic variation associated with individual interviewers (Fischer, 1982; Marsden, 2003; Paik and Sanchagrin, 2013; Van Tilburg, 1998). The intraclass correlation coefficient (ICC) in these studies ranged from a low of about 0.10 in the 2010 GSS and the 2005 *National Social Life, Health and Aging Project* to more than 0.20 in the 2004 GSS, the 1998 GSS, the 1995 *Chicago Health and Social Life Survey*, and a 1992 study of older adults in the Netherlands (the ICC measures the proportion of variability due to interviewers). The most likely source of this variation is uneven prompting by interviewers (Bearman and Parigi, 2004). Seeking to avoid the added series of questions that comes with each additional name given, some interviewers fail to ask the respondent for any discussants they may have missed, while others follow study protocol and prompt for additional names.

Abbreviations: UM, United Methodist; GSS, United States General Social Survey; CHI, Clergy Health Initiative; ICC, intraclass correlation coefficient.

* Corresponding author. Tel.: +1 919 699 6321; fax: +1 919 660 5623.
E-mail addresses: david.eagle@duke.edu, david.eagle2@gmail.com (D.E. Eagle), rae.jean@duke.edu (R.J. Proeschold-Bell).

[1] Also note that the terms "discussant" and "confidant" are used interchangeably throughout.

http://dx.doi.org/10.1016/j.socnet.2014.07.005
0378-8733/

We also know that name generator items are sensitive to their placement within long surveys. When placed near the end of the survey, or after other name-generator or similar questions, people report having fewer close confidants (Paik and Sanchagrin, 2013). There is also evidence from an experimental study on the use of name generators in online surveys that the number of fields available to enter names on a web form affects the number of names generated. From this previous study, researchers discovered that respondents feel pressure to fill in as many of the available boxes on a web form, which leads to larger estimates of overall network size (Manfreda et al., 2004). They also found that small changes in question wording exert a major impact on the number of people named (Bidart and Lavenu, 2005).

Finally, research has demonstrated that so-called "panel conditioning" presents a significant problem in longitudinal surveys that interview respondents at multiple time points (Torche et al., 2012; Warren and Halpern-Manners, 2012). Panel conditioning refers to the bias that emerges when respondents use their previous experience with questions on prior waves of the survey to alter their response. Studies have uncovered several psychological mechanisms governing panel conditioning. First, in some cases respondents use their prior experience with the survey to give answers that they think will help the interviewer. In other situations, the questions answered by respondents spur the respondent to become more knowledgeable about the issues raised. Subsequent to the interview, they become more informed on the subject and change their answers in the next wave of the survey. Finally, respondents may work to reduce the amount of effort they need to expend on the survey. Therefore, panel conditioning is more common on more burdensome questions, when survey waves are spaced relatively close together, and with increasing numbers of survey waves (Kruse et al., 2009; Meurs et al., 1989; Pickery et al., 2001; Presser and Traugott, 1992; Van Der Zouwen and Van Tilburg, 2001). Research has also underscored the importance of separating panel conditioning bias from panel attrition bias, where a group of people with similar characteristics leaves between waves (Das et al., 2011; Kruse et al., 2009; Warren and Halpern-Manners, 2012).

Previous longitudinal research has failed to uncover the presence of panel conditioning on name generator questions. For instance, in one study of older adults, the authors discover that across two waves of a survey, the average network size decreased, the smallest networks became larger, and the largest networks became smaller (Van Der Zouwen and Van Tilburg, 2001). However, the authors conclude that little of this difference is due to panel conditioning, and is, instead, attributable to interviewer effects. Interviewers had access to the respondent's answers at wave 1, and prompted for the same number of respondents at wave 2. Other studies conclude that while the members of an individual's networks change over time, the aggregate properties of networks do not change a great deal (Lubbers et al., 2010; Morgan et al., 1996). There are predictable effects over time on network size from major life events – in particular, getting married, entering and leaving college, and moving (Bidart and Lavenu, 2005).

1.1. Research objectives

In the present study, we analyze data from a panel study of clergy conducted by the Duke Clergy Health Initiative. Below, we describe our focal research objectives.

1.1.1. Interviewer effects in telephone surveys

Because multiple interviewers gathered the telephone data, this research adds to existing knowledge about interviewer effects in the collection of social network characteristics. We measure the interviewer effect in this survey across the seven interviewers and compare it to results from other surveys. We also look for any patterns that might suggest the presence of systematic interviewer effects (Kogovšek, 2006; Kogovšek et al., 2002).

1.1.2. Implementation of name generators in web surveys

Through the random assignment of respondents to telephone interview and web survey conditions, this study allows for the comparison of responses to the name generator and interpreter questions across these two administration modes.

1.1.3. Panel conditioning in name generators

This study is one of the few to implement the GSS "important matters" name generator and interpreter items in a repeated-panel design. This allows us to investigate whether we observe patterns in these data that are consistent with what we would expect under panel conditioning (Torche et al., 2012; Warren and Halpern-Manners, 2012).

2. Data

The data come from the first three waves of the Clergy Health Initiative (CHI) Longitudinal Survey, a longitudinal study of the health of United Methodist (UM) clergy in North Carolina (NC). In 2008, the Duke CHI invited all currently serving UM clergy to participate in the hour-long survey. In the 2008 survey, investigators implemented an experimental comparison of the web survey to the telephone interview. Because web-based surveys offer considerable savings, they implemented this test to see if the web survey could be substituted for the phone interviews in subsequent waves. Investigators randomly assigned two-thirds of respondents to receive the survey via the web, and one-third to receive a telephone interview. To maximize the overall response rate, participants in the web condition could request a paper survey if they did not have reliable Internet access; participants in the telephone condition could also request to complete the survey via web or paper. The 2010 and 2012 waves were conducted only using online surveys (with an option to request a paper survey if Internet access was an issue) and included all of the previously invited participants – even those who had refused participation in the previous wave, retired, moved away, or left the profession. These waves also added any clergy newly meeting the original 2008 study criteria. The new clergy added to the survey were, on average, younger, less experienced in ministry, and slightly more racially diverse than the previously invited participants.

The 2008 survey contains 1726 cases collected by phone, mail, or web and has a 95% response rate. In total, 652 respondents completed phone interviews, 999 web surveys, and 75 mailed in their responses. Seven different interviewers conducted the telephone interviews. Investigators randomly assigned clergy respondents to the telephone condition. The interviewers' ages ranged from 54 to 65 years, and only 1 was male. The 2010 survey contains 1679 cases collected online and 70 by mail with a response rate of 87%. 1513 respondents participated in the survey in both 2008 and 2010, and 241 new cases were added in 2010. The 2012 survey contains 1724 cases collected online and 53 by mail, with a response rate of 81%. Of these, 1328 people participated in all survey waves, 272 people participated in the 2012 wave and *either* the 2010 or 2008 wave, and a total of 181 new cases were added. 96% of respondents in 2010 and 97% of respondents in 2012 used the web to complete the surveys, with the remainder completing paper surveys.

The CHI Longitudinal Survey replicates the name generator question from the GSS. Specifically, it asks, "From time to time, most people discuss important matters with other people. Looking back over the last 6 months, who are the people with whom you discussed matters important to you?" Respondents can report as many names as they like. If the respondent names less than five people, then they are prompted if there is anyone else.[2] On the web survey form, respondents can list up to 30 names, a value others have used as a hypothetical maximum (Manfreda et al., 2004). The respondents are then asked about the characteristics of the first five individuals named. Using categories similar to the GSS, they are asked how they are connected to the respondent; the frequency of contact with that individual; and whether the individual named is United Methodist, a pastor, or a member of the church the respondent serves. The name generator item and follow-up interpreter items occur in the first quarter of the survey, which took, on average, 1 h to complete. Placing the name generator item early in the survey is likely to reduce interviewer and interviewee fatigue and lead to more reliable estimates (Paik and Sanchagrin, 2013).

While this survey did not set out to investigate the methodological issues surrounding the use of name generators, it provides insight into several methodological considerations in the study of core discussant networks. Nevertheless, these data have limitations that affect the number of names generated. Clergy are older (2008 average = 52), more educated (in 2008, 77% had a graduate degree, which is normally required for ordination), and more likely to be married than the population at large (in 2008, 87% were married). These factors are associated with an increase in the average size of core discussant networks (Marsden, 1987). In addition, the nature of the clergy profession may lead to a larger number of confidants. Clergy engage in a great deal of interpersonal contact, often about religious matters. They conduct counseling sessions, participate in wedding and funeral planning, and offer spiritual advice. UM clergy also normally participate in Bible studies and peer-mentoring groups. During pre-testing of the name generator question, clergy expressed that any of these interactions could be construed as involving the discussion of important matters. Even though this is likely to increase variability in the results, to retain consistency with the GSS, the question was retained as is.

3. Methods

3.1. Interviewer effects

Given that interviewers were randomly assigned to respondents, respondent characteristics were relatively stable across interviewers (results available upon request), respondents are restricted to UM clergy in North Carolina, and all the interviewers, save one, conducted a reasonably large number of interviews, we expect relatively even distributions of network ties across interviewers. In order to measure interviewer effects in the 2008 telephone survey, we calculated the ICC using the 2008 telephone data, employing a multi-level regression (Groves and Magilavy, 1986; Hox, 1994). In the first level of the model, respondents are grouped by interviewers, which form the second model level. This procedure allows us to decompose the variance into that attributable to respondents, σ_{resp}, and interviewers, σ_{int}. From these estimates, we calculate the intraclass correlation coefficient, ρ_{int} as follows:

$$\rho_{\text{int}} = \frac{\sigma^2_{\text{int}}}{\sigma^2_{\text{resp}} + \sigma^2_{\text{int}}} \quad (1)$$

The intraclass correlation coefficient indicates the proportion of unexplained variance that can be attributed to interviewers and provides an estimate of the degree of correlation between responses given and the interviewers. As an overall measure of variability, the ICC will not necessarily pinpoint the potential interviewer effects. There is the possibility that the interviewer may have simply stopped prompting for names once 5 people were named, or, that they might prompt to get exactly 5 names, or, that they might have not prompted respondents to give more names if they reported fewer than five names. To reveal the by-interviewer patterns, we cross tabulate the number of cases with a network size less than five, equal to five and more than five. We also run a mixed-effects logistic regression with two different outcome variables – the first indicates a respondent with exactly 5 discussants, the second outcome is an indicator for respondents who have networks smaller than 5. The cases are clustered by interviewer and interviewer-specific predicted probabilities are calculated and compared. The regression equation is as follows:

$$Y_{i,j} \sim \log it(u_j) \quad \text{where } u_j \sim N(U, \sigma^2) \quad (2)$$

Here j is the index on interviewer (1...6), U is the average across interviewers and u_j is mean for interviewer j. To compare interviewers, plots of the median and credible values of u_j are presented.

3.2. Telephone versus web condition

In order to study the impact of the different delivery modes of this survey, we calculated the mean and standard deviation of the number of names generated by data collection mode. We also compared the size of clergy kin and non-kin discussion partner networks (kin are people related to the respondent by blood, marriage or adoption). Mean network characteristics by collection mode were compared using a t-test, and the standard deviations using an F-test.

3.3. Panel conditioning

If panel conditioning is a significant issue, it is reasonable to anticipate that repeat respondents, knowing that they will be asked interpreter questions on the first 5 people they name, will become more likely to report exactly 5 confidants. We also expect that repeat respondents may become more likely to report networks smaller than 5 in order to reduce the number of interpreter questions that they will be required to answer.

Measuring panel conditioning presents a challenge because it often occurs along with panel attrition bias (where people who leave the study vary systematically from the general population) and with real change. Panel conditioning bias is calculated as the observed total change less the panel attrition bias, less the real change. But, we do not observe the true change, nor do we observe the network size for attritionists at time 2. Therefore, we cannot calculate the value of panel conditioning bias without making additional assumptions (Das et al., 2011). One assumption that helps with identifiability is to assume that the true aggregate change in the average network size over time is equal to zero. Fischer's (2009) suggestion that, overall, the number of discussants has not changed for the US population, supports this assumption as a reasonable claim. The CHI Longitudinal Survey is a survey of the entire population of UM clergy within NC, and possesses relatively stable characteristics. These clergy are mostly married, share the same occupation, are of a similar age (as retirees are dropped and new clergy added at each wave) and, because of the structure of the UM church, relocate at similar intervals. To calculate panel conditioning bias, we also need an estimate of the network size for those who left the sample. However, we do not know the size of attritionist networks at time 2. We can compare the demographic characteristics of those who only participated in 2008 against those who participated in both 2008 and 2010, and likewise those who participated in 2010 against those who participated in both 2010

[2] This matches the 1985 GSS, but differs from the 2004 and 2010 GSS where respondents could name as many discussion partners as they liked, but only the first five names were recorded. If the respondents gave more than five names, the interviewer indicated the respondent's network size as "6 or more".

Table 1
Number of confidants named by interviewer.

Panel A		Network size	
Interviewer	N	Mean	SD
1	45	4.4	2.5
2	213	6.1	4
3	143	5.0	2.6
4	32	5.8	3.8
5	82	7.3	4.5
6	130	6.9	3.6
7	7	6.5	5.7
ICC (zero-order)	0.07		

Panel B	Proportion reporting numgiven			
Interviewer	<5	≤5	=5	>5
1	0.64	0.76	0.11	0.24
2	0.24	0.53	0.29	0.47
3	0.37	0.68	0.31	0.32
4	0.44	0.59	0.16	0.41
5	0.07	0.45	0.38	0.55
6	0.02	0.50	0.48	0.50
7	0.57	0.57	0.00	0.43

[a] Proportions do not sum to 1 because columns 2 and 3 are not mutually exclusive.

Table 2
Size of clergy discussion networks by survey type, 2008.

Network size	Percentage of respondents		
	Phone	Mail	Web
0	0.9	1.3	0.7
1	3.1	4.0	1.5
2	4.6	5.3	1.9
3	8.0	4.0	4.3
4	7.8	0.0	6.1
5	32.1	25.3	12.2
6	14.0	18.7	9.7
7	9.8	12.0	8.5
8	5.7	9.3	6.9
9	4.1	1.3	5.4
10	2.6	2.7	7.6
11–20	6.0	14.7	24.7
21–30	1.4	1.3	10.4
Mean[a]	6.1	7.1	10.4
Median	5.0	6.0	8.0
SD[b]	3.7	4.4	7.1
N	652	75	999
ICC	0.19		

Source: Clergy Health Initiative Survey, 2008 wave.
[a] All differences in mean are significant at the $p < 0.001$ level using a two-tailed t-test.
[b] All differences in the standard deviation are significant at the $p < 0.05$ level using an F-test.

and 2012. If the attritionists and non-attritionists do not differ significantly, then we can assume that panel attrition is not a major factor.

With these assumptions, we can then compare the network characteristics of people who switched response categories between waves. If repeat respondents remember that this survey asks them only for the characteristics of the first 5 people, then the probability of respondents switching to reporting 5 discussion partners in the subsequent survey wave should be larger than those switching into other categories (Das et al., 2011). This is a conservative approach, for we might expect people to switch to naming *fewer* than five respondents to reduce cognitive burden by avoiding follow up questions. However, if respondents switch to 5 confidants with a larger probability than other categories, this provides evidence of panel conditioning. In order to calculate the probability of switching from one value of the number of discussants named (*numgiven*) to another, we compute the following:

$$Pr_{Y_2|Y_1}(Y_2 = numgiven_{2010}|Y_1 = numgiven_{2008}) = \frac{Pr(\{Y_2 = numgiven_{2010}\} \cap \{Y_1 = numgiven_{2008}\})}{Pr(\{Y_1 = numgiven_{2008}\})} \quad \text{where } \{Y_2\}\backslash\{Y_1\} \tag{3}$$

As already mentioned, research suggests that panel conditioning is more likely to occur in those who participate in multiple waves of a study. In order to test if there is a cumulative effect of participating in all three waves of the survey, in a similar fashion to Eq. (3), we calculate:

$$Pr_{Y_3|Y_2,Y_1}(Y_3 = numgiven_{2012}|Y_2 = numgiven_{2010}, Y_1 = numgiven_{2008}) \text{ where } \{Y_3\}\backslash\{Y_2\} \text{ and } \{Y_2\}\backslash\{Y_1\} \tag{4}$$

4. Results

4.1. Interviewer effects

In Table 1, panel A, we show that for the seven interviewers, the mean size of the respondent's social network varies considerably. The zero-order ICC is 0.071, a modest effect, which indicates about 7% of the variance in reported network size is attributable to interviewers. In panel B of Table 1, we report, by interviewer, the proportion of interviews with a network size of less than 5, less than or equal to 5 equal to 5 and greater than 5. In particular, we were interested in seeing if there was interviewer variability in reports of reports of more than 5 discussants and exactly 5 discussants. Interviewers 2, 4, 5, 6, and 7 have percentages of more than 5 discussants that range from 41% to 55%. However, only 24% of interviewer 1's cases and only 32% of interviewer 3's cases have more than 5 discussants. Interviewers 5 and 6 have strikingly high percentages of exactly 5 discussants, namely 38% and 48%, which differs from interviewers 1 and 4, whose percentage of exactly 5 discussants was only 11% and 16%, respectively. To explore the statistical strength of these differences, Fig. 1 plots the predicted probability, with 68% and 90% credible intervals, of a respondent giving exactly five and less than five confidants by interviewer as predicted by the mixed-effects model. In terms of the probability of naming 5 respondents, interviewer 1 (significant in the 90% CI) and interviewer 4 (significant in the 68% CI) have much lower probabilities of reporting a network size of 5 than the other interviewers. Interviewer 6 has a much higher probability of reporting a network size of 5 (significant in the 68% CI). Similarly, interviewers 5 and 6 have a much lower probability of reporting fewer than 5 confidants (significant in the 90% CI) than the other interviewers, and interviewer 2 has a lower probability than interviewers 1, 3, and 4 (significant in the 90% CI). Interviewer 1 has the largest probability of reporting a network with fewer than 5 members (significant in the 68% CI).

4.2. Telephone versus web condition

Because respondents received random assignment into either the web or the telephone condition, these data allow us to assess the impact of web versus telephone delivery for this study. For completeness, we also report characteristics for those who filled out a mail survey, but we cannot make strong conclusions about this group because of the small number of cases ($n = 75$) and because they were not randomly assigned to this condition.

In Table 2, we report the characteristics of the core discussion networks reported to researchers in 2008 by telephone, mail, and

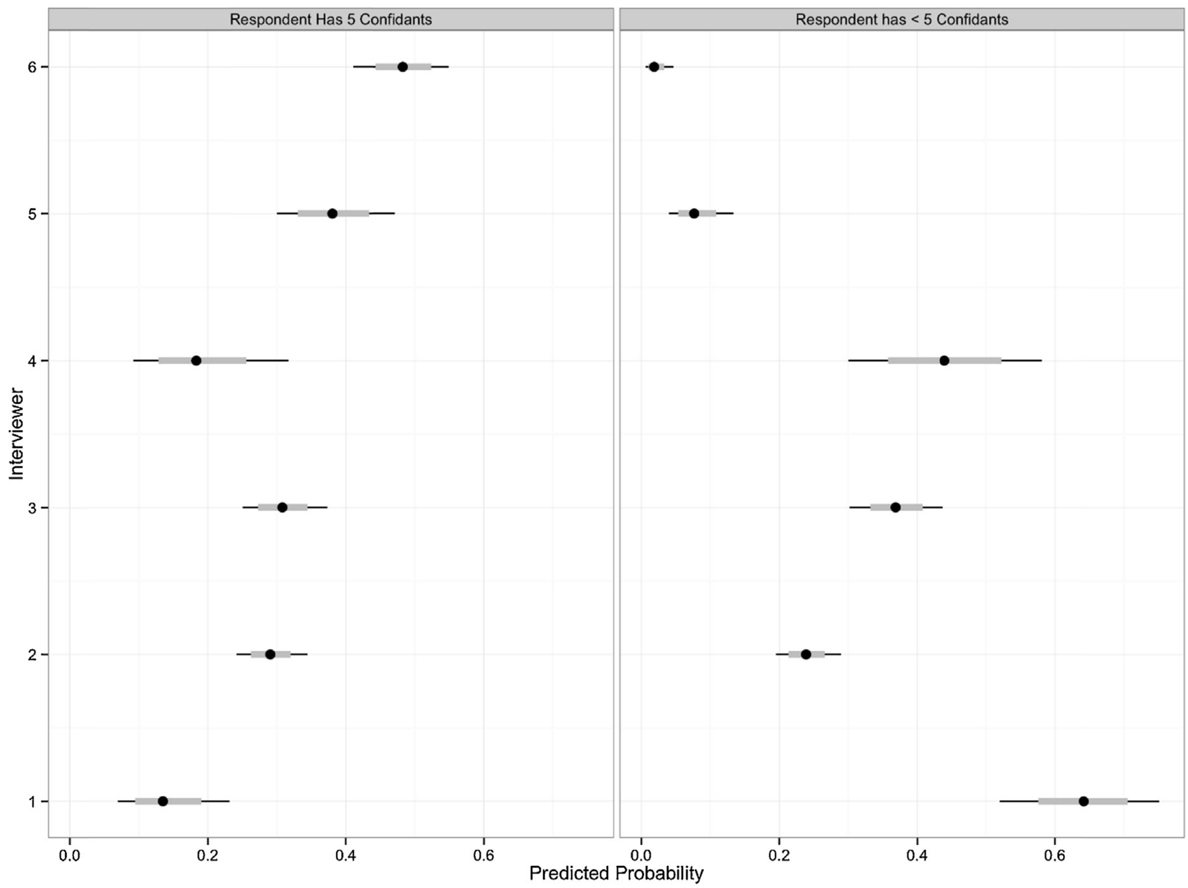

Fig. 1. Predicted probability of respondent reporting 5 and less than 5 confidants by interviewer. Heavy gray and thin black bands show 68% and 90% credible intervals, respectively.

web conditions. The mean network size for the telephone condition was 6.1 versus 7.1 for the mail survey and 10.4 for the web survey; the median was 5, 6, and 8, respectively. The standard deviation for the web survey was much larger than either the phone or the mail surveys (7.1 versus 3.7 for phone and 4.4 for mail). A significant proportion of this difference is attributable to web respondents who reported very large networks: 5.0% of web respondents listed 30 partners, whereas only 0.5% of respondents did so in the phone survey and none in the mail survey.

The CHI Longitudinal Survey contains name interpreter questions for the first 5 discussion partners named. One of these interpreters asks respondents to indicate how they are connected to each of the people they name. Turning to Table 3, we see that survey type does not exert a significant impact on the reported size of clergy kin/non-kin networks (kin are those directly related to the respondent through legal marriage, biological, or adoptive ties). The mean, median, and standard deviation of the size of kin and non-kin networks are not significantly different from one another based on survey condition. We also tested whether network density changed across survey type. Global network density was measured with a single question asking if all (coded 3), some (coded 2), a few (coded 1) or none (coded 0) of the first 5 named discussants knew each other. The average density was 2.09 for phone, 1.99 for mail and 2.06 for the web.

4.3. Panel conditioning

The CHI Longitudinal Survey administered the name generator questions across the three waves of the panel survey. In Table 4, we provide the mean, median, and standard deviation of the total network size by the waves of the panel survey in which the respondent participated (for 2008, we report statistics only for those who participated in the web condition to make the results directly comparable across waves). In 2008, retired pastors were not surveyed, but if they retired between waves, they were retained in the study. The study also followed pastors who left the occupation between

Table 3
Clergy kin and non-kin discussion networks by survey type, 2008.

Number of names given	Kin network			Non-kin network		
	Phone (%)	Mail (%)	Web (%)	Phone (%)	Mail (%)	Web (%)
Mean[a]	1.71	1.31	1.7	4.67	4.47	4.02
Median	1	1	1	4	4	4
SD[b]	1.2	0.87	1.26	3.34	3.05	2.66
N	650	71	974	652	75	999

Source: Clergy Health Initiative Survey, 2008 wave.

[a] None of the differences in the mean are significant at the $p < 0.05$ level, t-test, two-tailed.

[b] None of the differences in the standard deviation are significant at the $p < 0.05$ level, F-test.

Table 4
Population composition across waves, web survey data only.

	2008	2010	2012
Network size, mean	10.4	8.5	7.7
Median	8	7	6
SD	(7.07)	(6.21)	(5.68)
Years in ministry	16.8	17.9	17.5
SD	(11.83)	(12.3)	(12.13)
Age	51.8	52.9	53.1
SD	(7.07)	(6.21)	(5.68)
Married, %	87.4	88.0	87.8
Graduate degree, %	78.6	80.2	80.8
N	999	1620	1582

Source: Clergy Health Initiative Longitudinal Survey, Waves 2008, 2010, 2012, retirees and those who left ministry excluded.

Table 5
Characteristics of respondents who left the study prior to 2010 and 2012.

	Exited study between 2008 and 2010	Exited study between 2010 and 2012
Network size, mean	9.1	7.8
Median	7	7
SD	(6.77)	(5.54)
Years in ministry	16.8	17.1
SD	(9.59)	(9.92)
Age	49.6	51.7
SD	(6.77)	(5.54)
Married, %	76.9	88.0
Graduate degree, %	57.7	73.5
N	52	83

Source: Clergy Health Initiative Longitudinal Survey, Waves 2008, 2010, 2012, retirees and those who left ministry excluded.

waves. In order to keep the composition of the population comparable across all three waves, we dropped from the 2010 data all pastors who left the occupation (22) and all pastors who retired (122); and from 2012, all 19 pastors who left the denomination and all 176 pastors who retired.

These results reveal a decrease across each survey wave in the mean, median and standard deviation of the overall size of the respondent's network. Network size declined by an average of 1.9 discussants from 2008 to 2010 ($p < 0.001$) and by 0.8 discussants from 2010 to 2012 ($p < 0.001$). There was also a significant decline in the standard deviations from 7.07 to 6.21 ($p < 0.01$) and from 6.21 to 5.68 ($p = 0.005$). However, on other key measures known to influence network size – age, years in the occupation, marital status, and education – there were no significant changes.

In Table 5, we report the characteristics of those who exited the survey before the 2010 and 2012 waves. Attritionists who leave before the 2010 wave have smaller networks than the overall population (9.1 versus 10.4, $p = 0.19$), are less likely to be married (−10.5% as compared to the non-attritionists, $p = 0.03$) and less likely to have a graduate degree (−20.9%, $p < 0.001$). None of the differences are significant at the $p = 0.05$ levels for those who left between 2010 and 2012. Marriage and education are associated with larger networks, so we expect that attrition bias between 2008 and 2010 would contribute to an increase in network size.

In these data, a large proportion of respondents report a different number of confidants in wave 2 than in wave 1, in wave 3 than in wave 2, and in wave 3 than in waves 1 and 2. 89% of respondents switch the number of people they report in their confidant-network between 2008 and 2010 and 86% between 2010 and 2012. Only 10% report the same number of confidants across all three waves of the study. Using Eq. (3) (see Section 3), we plot the probability of a respondent at time 2 switching into a given response category from the previous survey wave at time 1. The plot is shown in Fig. 2. The largest change among repeat respondents is a shift into the exactly 5 named-discussant category. Between 2008 and 2010, 17% of the respondents shifted to a network of size 5; between 2010 and 2012, 16% of respondents made this shift. The next largest shift was to 6 named-discussants (11% between 2008 and 2010 and 9% between 2010 and 2012) and 4 named-discussants (8% between 2008 and 2010 and 9% between 2010 and 2012). Fig. 3, which utilizes Eq. (4) (see Section 3), plots the probability of switching into a given response category in 2012, conditional on giving a different response in 2008 and 2010. Once again, the same phenomenon is revealed. 12% of the respondents switch to reporting 5 discussants, given that they did not report 5 confidants in either of the prior waves. The next most likely category for respondents who switch is into 4 or 6 confidants (both 7%).

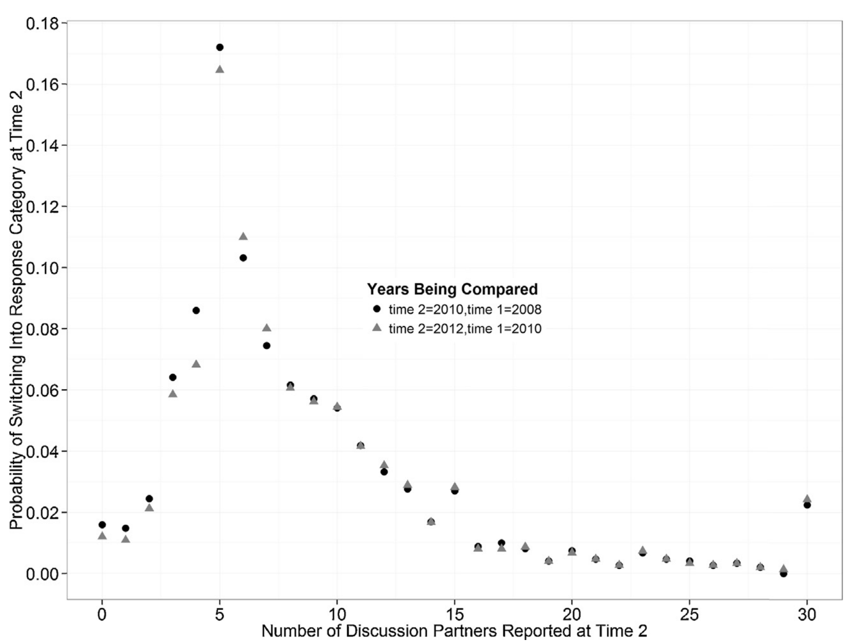

Fig. 2. Plot of the probability of switching into a given response category, repeat respondents only ($n_{switchers[2008-2010]} = 1343$; $n_{switchers[2010-2012]} = 1307$).

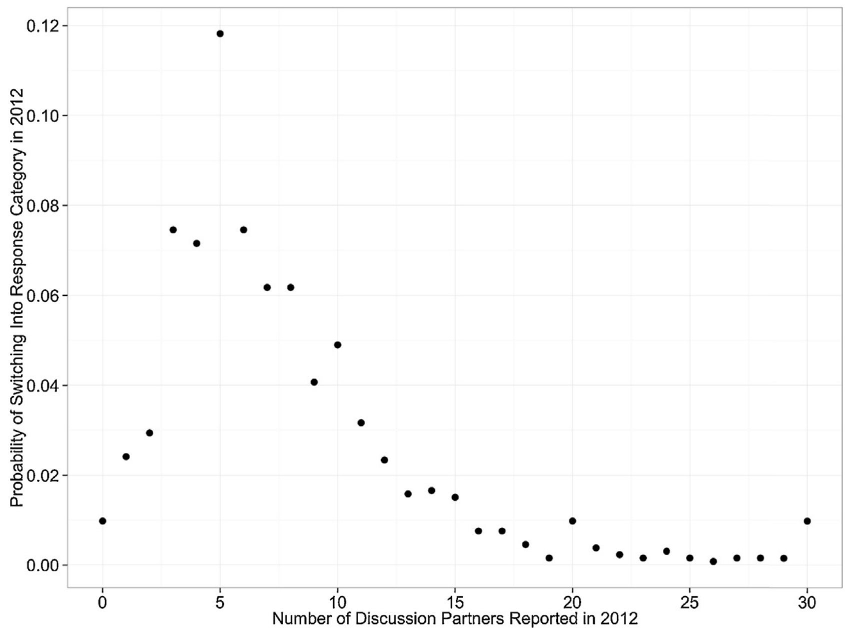

Fig. 3. Plot of the probability of switching into a response category not given in either the 2008 or 2010 waves ($n_{switchers[2012]}$ = 1030).
Source: Clergy Health Initiative Longitudinal Survey, 2008, 2010 and 2012 waves.

In terms of the composition of the social network, the average size of the kin and non-kin networks and the network density do not change significantly over time. There are also no significant differences in kin and non-kin network size and network density between participants in multiple survey waves versus new 2010 or 2012 participants (results available upon request).

5. Discussion and conclusions

This study replicates a number of results found in previous studies and also contributes new knowledge about using name generator and interpreter items to collect information on respondent social networks. Firstly, we replicated the finding that interviewers affect the number of discussants named. Secondly, we found that participants in the web versus the telephone condition were significantly more likely to report large numbers of confidants. Thirdly, in terms of the composition of the discussant network, the web and telephone conditions produced similar results. Finally, this study adds to knowledge about panel conditioning in name generators and interpreters. In our study, we found that the most popular number of confidants reported among those who switched response categories was 5, equal to the number of confidants who receive follow up questions. We further discuss our findings below.

5.1. Interviewer effects

The telephone interviews in the CHI Longitudinal Survey produced an ICC of 0.07, which is lower than other studies that contain a name generator for which an ICC is available (Paik and Sanchagrin, 2013). There are several possible pathways leading to this relatively low ICC. First of all, this survey is of clergy, who may be more compliant than other populations. Past research has shown that both the religiously affiliated and those with a more pro-social orientation are more likely to be positively oriented toward survey participation (Abraham et al., 2009; Presser and Stinson, 1998). The high response rates obtained in this survey are also evidence of a high degree of buy-in to the study's goals by respondents. Second, it has been shown that the extent and quality of interviewer training plays an important role in data quality (Billiet and Loosveldt, 1988). Interviewers in this study received extensive training and supervision from Westat, an experienced survey research company (de Leeuw, 1992, 2014).

However, as in past research, this study contains evidence of interviewer effects on the number of names generated. The number of interviews with exactly 5 names generated is difficult to imagine occurring completely at random. Indeed, 2 of the interviewers had almost half of their respondents reporting networks with exactly 5 participants and the other half with networks larger than 5. Two other interviewers recorded nearly a third of their respondents with exactly 5 discussion partners. It is hard to believe that 5 is a "magic" number of connections that an individual might have. There are several possible explanations for this pattern.

First of all, it was left to respondents to decide what constitutes "important matters." Interviewer effects could come about if respondents ask for clarification and the interviewers provide different definitions. Another possibility is that interviewers varied systematically in the extent of their probing. Previous research has demonstrated that this is a factor in studies with multiple interviewers (Groves and Magilavy, 1986). In this study, interviewers were instructed to probe for additional names if the respondent reported fewer than 5 respondents. It is possible that some interviewers failed to prompt for additional names if the respondent gave exactly 5 names. It is also possible that some interviewers, when faced with respondents who have less than 5 discussion partners, prompt for additional names until they have obtained 5. Finally, the fact that one interviewer generated a significantly smaller number of networks larger than 5 raises the possibility that this interviewer was trying to limit the number of names reported (Van Der Zouwen and Van Tilburg, 2001).

5.2. Telephone versus web condition

Respondents in the web condition were much more likely to report very large networks. There are several possible reasons for this effect. First, we know that burdensome questions produce poorer quality responses (Bell et al., 2007; Feld and Carter, 2002;

Krosnick, 1991). Name generators fall into the burdensome category (Paik and Sanchagrin, 2013). Not only do respondents need to dig into their social networks and generate names, but respondents also have to remember what names they have already given. This cognitive burden may lead respondents to stop giving names prematurely. In the telephone condition, respondents do not have ready access to the list of names already given, whereas in the web survey, they do. It is possible that the smaller-sized networks in the telephone condition may come from the relative difficultly in recalling what names have been given to the interviewer. In the web condition, being able to see the list could ease the cognitive burden of generating additional names, because respondents do not have to hold in their head everyone they have already named.

Another source of difference between the phone and web condition may relate to the visual clues provided by the web form. In an experimental study, Manfreda et al. (2004) show that respondents often pay more attention to the visual layout of the question, than to the question itself. In their web survey, which also contained 30 separate spaces in which to fill names, they found that the separate boxes provided a cognitive cue to generate a longer list. Although our results do not allow for systematic comparison of the mail survey, it is worth noting that the mail survey only produced a marginally larger network size than the telephone interview. This result is surprising, as it seems reasonable to assume that paper surveys function similarly to a web survey. The clue to the difference may lie in the design of the mail survey. It provided respondents with a large box where they could record multiple names, but the field was not broken into separate boxes in which to record a single name. The lack of a discrete number of boxes could have made any upper limit open to the respondent's interpretation.

While the name generator item appears to suffer the effects of survey type, the name interpreter items – in this case the size of kin and non-kin networks, and global network density – produced similar results whether collected in phone, mail or web conditions. This result is consistent with previous research that shows that lower-respondent-burden questions are subject to less variability (Krosnick, 1991). The name interpreter questions are less ambiguous than the name generator question – they simply require respondents to check boxes to indicate their relationship with the person already named or to provide an overall assessment of network density.

5.3. Panel conditioning

In the 2008 web survey, we observed a significantly higher average network size than in 2010. This is an unexpected result, given that the interview form, questionnaire, question placement, and sample characteristics remain the same across all three waves of the survey. A smaller decline of 0.8 people was observed from 2010 to 2012. These drops could be due to panel attrition bias, real change in network characteristics and/or panel conditioning bias. Our results did not provide strong support for the presence of attrition bias. While we detected some attrition bias between 2008 and 2010, because unmarried people and those with lower levels of education were slightly more likely to leave the sample, it is most probable that this would make the average network size larger, because marriage and education are associated with larger networks (Marsden, 1987).

We know that between waves, there is a great deal of switching among respondents in the number of confidants named. For switchers, by a large margin, the most popular category switched into is 5. This, along with the fact that our population evidences stability on a variety of dimensions related to social network characteristics, makes it less likely that real change in network characteristics is driving this result. Other research lends support to this interpretation. For instance, research using the GSS has shown that the average size of core discussion networks in America has remained stable (Fischer, 2009; Paik and Sanchagrin, 2013).

If panel conditioning is occurring, what are the possible mechanisms at play? It is possible that respondents with networks smaller than 5, in reaction to the survey question, decided they did not have enough confidants and, prior to the next wave, sought more discussion partners. However, it is difficult to see why people would change their confidant networks in response to a survey. The more plausible mechanism is that the survey design, independent of any real aggregate change in network characteristics, is driving this result. Repeat respondents remember that they are asked follow-up questions for only the first 5 names they provide and therefore report only 5 confidants in subsequent waves. A possible objection to this explanation is that rather than switching to 5 respondents, to reduce survey burden, people should switch to fewer than 5. The unique characteristics of the study population could provide a clue as to why we might expect people to switch to reporting 5 confidants rather than fewer than 5. As mentioned before, there is reason to expect that clergy will be more favorably disposed toward surveys in general. Also, this was a survey of UM pastors with the express purpose of designing programs to improve clergy well-being, which may have led to a greater willingness by respondents to provide optimally helpful responses. If respondents assumed that the name interpreter questions were of focal importance, and that any names, above and beyond 5, were extraneous information, this may have led them to report exactly 5 confidants.

5.4. Implications

This study underscores the importance for researchers to carefully consider the issues related to panel conditioning, interviewer effects, and survey design when implementing name generator and interpreter items to study the overall size of a respondent's social network.

We suspect that other studies will be subject to panel conditioning when implementing a name generator that is followed by name interpreter questions on a subset of those named. In our study, it appears likely that, even after two years, respondents remembered that they are not asked to provide the characteristics of more than 5 of their discussion partners, and become more likely to report only 5. With a less engaged group of respondents, this effect may not be as strong and it may result in a larger number of respondents switching to naming fewer than 5 confidants.

One alternative would be to ask respondents for the characteristics of every one of their discussion partners, rather than limiting the name interpreter questions to a subset of discussants named. This, however, significantly increases survey burden, which may, in turn, produce panel conditioning, where respondents name fewer confidants in subsequent waves. Part of the solution may lie in choosing a name generator question that elicits a more specific conversation domain and, hopefully, a smaller network. Part of the problem with the "important matters" name generator is that it is often interpreted expansively, which can lead to large networks (Bearman and Parigi, 2004). As our interview pre-tests indicated, this was especially acute in our study where the respondents are clergy, whose job it is to "discuss important matters."

Our study also affirms the importance of careful training and supervision of interviewers. Our results were suggestive of uneven prompting by interviewers, something documented in other studies (Paik and Sanchagrin, 2013; Van Der Zouwen and Van Tilburg, 2001; Van Tilburg, 1998). Interviewers should also be carefully trained and monitored, such that prompting is consistent across interviewers and respondents.

On the positive side, our study found little difference between web and telephone collection of the characteristics of a respondent's network. Moving to web-only collection resulted in significant savings. However, our respondents had higher than average educational attainment, were all in professional positions, and were overwhelmingly white. Our findings may not generalize to other groups, for instance among the elderly or low socioeconomic status immigrants. Less consistent and reliable results may also be produced when respondents are not as personally invested in the study's outcomes. Overall, findings from this study may assist researchers to conduct future name generator studies with greater accuracy.

Acknowledgements

This study was funded by a grant from the Rural Church Area of The Duke Endowment. We thank Achim Edelmann for his very helpful comments and suggestions and our research team at Westat, including Crystal MacAllum, Gail Thomas, and Ed Mann, for superb data collection efforts.

References

Abraham, K.G., Helms, S., Presser, S., 2009. How social processes distort measurement: the impact of survey nonresponse on estimates of volunteer work in the United States. Am. J. Sociol. 114, 1129–1165.
adams, J., Moody, J., 2007. To tell the truth: measuring concordance in multiply reported network data. Soc. Netw. 29, 44–58.
Bearman, P., Parigi, P., 2004. Cloning headless frogs and other important matters: conversation topics and network structure. Soc. Forces 83, 535–557.
Bell, D.C., Belli-McQueen, B., Haider, A., 2007. Partner naming and forgetting: recall of network members. Soc. Netw. 29, 279–299.
Bidart, C., Lavenu, D., 2005. Evolutions of personal networks and life events. Soc. Netw. 27, 359–376.
Billiet, J., Loosveldt, G., 1988. Improvement of the quality of responses to factual survey questions by interviewer training. Public Opin. Q. 52, 190–211.
Burt, R.S., 1984. Network items and the general social survey. Soc. Netw. 6, 293–339.
Campbell, K.E., Lee, B.A., 1991. Name generators in surveys of personal networks. Soc. Netw. 13, 203–221.
Das, M., Toepoel, V., van Soest, A., 2011. Nonparametric tests of panel conditioning and attrition bias in panel surveys. Sociol. Methods Res. 40, 32–56.
de Leeuw, E.D., 1992. Data Quality in Mail, Telephone and Face to Face Surveys. ERIC.
de Leeuw E.D. (2014). Mode effects in survey research: a comparison of mail, telephone, and face to face surveys. Dissertation. Free University of Amsterdam. ERIC Document Reproduction Service No. ED374136. Retrieved August 14, 2014, from ERIC database.
Feld, S.L., Carter, W.C., 2002. Detecting measurement bias in respondent reports of personal networks. Soc. Netw. 24, 365–383.
Ferligoj, A., Hlebec, V., 1999. Evaluation of social network measurement instruments. Soc. Netw. 21, 111–130.
Fischer, C.S., 2009. The 2004 GSS finding of shrunken social networks: an artifact? Am. Sociol. Rev. 74, 657–669.
Fischer, C.,1982. To dwell among friends. University of Chicago Press, Chicago, p. p.302.
Groves, R.M., Magilavy, L.J., 1986. Measuring and explaining interviewer effects in centralized telephone surveys. Public Opin. Q. 50, 251–266.
Hammer, M., 1984. Explorations into the meaning of social network interview data. Soc. Netw. 6, 341–371.
Hlebec, V., Ferligoj, A., 2002. Reliability of social network measurement instruments. Field Methods 14, 288–306.
Hox, J.J., 1994. Hierarchical regression models for interviewer and respondent effects. Sociol. Methods Res. 22, 300–318.
Kogovšek, T., 2006. Reliability and validity of measuring social support networks by web and telephone. Metodol. Zv. 3, 239–252.
Kogovšek, T., Ferligoj, A., 2005. Effects on reliability and validity of egocentered network measurements. Soc. Netw. 27, 205–229.
Kogovšek, T., Ferligoj, A., Coenders, G., Saris, W.E., 2002. Estimating the reliability and validity of personal support measures: full information ML estimation with planned incomplete data. Soc. Netw. 24, 1–20.
Kogovsek, T., Hlebec, V., 2009. Stability of typologies produced on the basis of repeated measurement with the role relationship and the name generator approach. Metodol. Zv. 6, 85–97.
Krosnick, J.A., 1991. Response strategies for coping with the cognitive demands of attitude measures in surveys. Appl. Cogn. Psychol. 5, 213–236.
Kruse, Y., Callegaro, M., Dennis, J.M., Subias, S., Lawrence, M., DiSogra, C., Tompson, T., 2009. Panel conditioning and attrition in the AP-Yahoo! news election panel study. In: Proceedings of the 64th conference of the American Association for Public Opinion Research (AAPOR), Hollywood, FL, pp. 5742–5756.
Lubbers, M.J., Molina, J.L., Lerner, J., Brandes, U., Ávila, J., McCarty, C., 2010. Longitudinal analysis of personal networks. The case of Argentinean migrants in Spain. Soc. Netw. 32, 91–104.
Manfreda, K.L., Vehovar, V., Hlebec, V., 2004. Collecting ego-centred network data via the web. Metodol. Zv. 1, 295–321.
Marsden, P.V., 1987. Core discussion networks of Americans. Am. Sociol. Rev. 52, 122–131.
Marsden, P.V., 1993. The reliability of network density and composition measures. Soc. Netw. 15, 399–421.
Marsden, P.V., 2003. Interviewer effects in measuring network size using a single name generator. Soc. Netw. 25, 1–16.
Matzat, U., Snijders, C., 2010. Does the online collection of ego-centered network data reduce data quality? An experimental comparison. Soc. Netw. 32, 105–111.
McPherson, M., Smith-Lovin, L., Brashears, M.E., 2006. Social isolation in America: changes in core discussion networks over two decades. Am. Sociol. Rev. 71, 353–375.
McPherson, M., Smith-Lovin, L., Brashears, M.E., 2008. Social isolation in America: changes in core discussion networks over two decades. Am. Sociol. Rev. 73, 1022.
Meurs, H., Wissen, L., Visser, J., 1989. Measurement biases in panel data. Transportation 16, 175–194.
Morgan, D.L., Neal, M.B., Carder, P., 1996. The stability of core and peripheral networks over time. Soc. Netw. 19, 9–25.
Paik, A., Sanchagrin, K., 2013. Social isolation in America: an artifact. Am. Sociol. Rev. 78, 339–360.
Pickery, J., Loosveldt, G., Carton, A., 2001. The effects of interviewer and respondent characteristics on response behavior in panel surveys: a multilevel approach. Sociol. Methods Res. 29, 509–523.
Presser, S., Stinson, L., 1998. Data collection mode and social desirability bias in self-reported religious attendance. Am. Sociol. Rev. 63, 137–145.
Presser, S., Traugott, M., 1992. Little white lies and social science models: correlated response errors in a panel study of voting. Public Opin. Q. 56, 77–86.
Torche, F., Warren, J.R., Halpern-Manners, a., Valenzuela, E., 2012. Panel conditioning in a longitudinal study of adolescents' substance use: evidence from an experiment. Soc. Forces 90, 891–918.
Van Der Zouwen, J., Van Tilburg, T., 2001. Reactivity in panel studies and its consequences for testing causal hypotheses. Sociol. Methods Res. 30, 35–56.
Van Tilburg, T., 1998. Interviewer effects in the measurement of personal network size: a nonexperimental study. Sociol. Methods Res. 26, 300–328.
Warren, J.R., Halpern-Manners, A., 2012. Panel conditioning in longitudinal social science surveys. Sociol. Methods Res. 41, 491–534.
Zemljič, B., Hlebec, V., 2005. Reliability of measures of centrality and prominence. Soc. Netw. 27, 73–88.