

# Response Bias over Time: Interviewer Learning and Missing Data in Egocentric Network Surveys

Thomas W. Valente<sup>1</sup>, Leanne Dougherty<sup>2</sup>,  
and Emily Stammer<sup>2</sup>

## Abstract

This study investigates potential bias that may arise when surveys include question items for which multiple units are elicited. Examples of such items include questions about experiences with multiple health centers, comparison of different products, or the solicitation of egocentric network data. The larger the number of items asked about each named individual or location, the greater potential interviewer and respondent burden accrues to the naming of more names. Interviewers may be inclined to limit the number of names elicited to reduce the amount of time required to complete the interviews. We tested whether such bias occurred from data collected in northwest Ghana by contrasting group learning with individual learning. The results provided mixed evidence for both group and individual learning and stress the need to take actions such as increased training, change in incentives, and/or monitoring responses to guard against such results.

<sup>1</sup> Department of Preventive Medicine, Keck School of Medicine, Los Angeles, CA, USA

<sup>2</sup> JSI Research & Training Institute, Inc., Rosslyn, VA, USA

## Corresponding Author:

Thomas W. Valente, Department of Preventive Medicine, Keck School of Medicine, 2001 N. Soto Ave., Los Angeles, CA 90032, USA.

Email: tvalente@usc.edu

Survey methodologists have for years cautioned against overly long surveys that result in poor or missing data (Converse and Presser 1986; Valente 2002). Proscriptions for survey length have been put forward, and the wise researcher is cautioned to include in surveys only those questions that are essential to answering the research questions at hand. This challenge can become exacerbated when collecting data on topics for which multiple units are elicited. For example, loop options occur when a researcher wishes to know about a study participant's experience with visits to multiple health centers, comparison of different products, or various family members. In some maternal child health surveys, for example, respondents are asked to name all their children and provide information on each one.

Most studies of interviewer bias focus on the sociodemographic characteristics of the interviewer and whether those characteristics match the participants (e.g., Flores-Macias and Lawson 2008). Some studies have examined whether interviewers bias survey administration (Kosyakova et al. 2014). Other studies focus on whether the measurement of demographic characteristics or sensitive subjects such as substance use or sexual activity vary by interviewer characteristics (Davis et al. 2010). When it comes to survey length, most researchers intuitively argue that shorter surveys are better because they reduce respondent burden and increase response rate. With the exception of mailed surveys, however, the potential respondent does not know the survey length beforehand (Bogen 1996). And an earlier literature review found no association between survey length and response rate (Bogen 1996).

More recently, however, there is evidence that survey length depresses respondent participation, but this is by no means a uniform finding (Rolstad et al. 2011). Instead, the evidence seems to indicate the survey content matters as much as survey length (Rolstad et al. 2011). One prior study showed that interviewers might manipulate the size of the network they measure if incentivized by the number of interviews completed compared to being paid hourly (Josten and Trappman 2016).

The problem of interviewer bias can be particularly acute when researchers attempt to measure social networks via egocentric name generator questions (Campbell and Lee 1991; Eagle and Proeschold-Bell 2015; Marsden 2003; McCarty et al. 1997; Valente 2010; Valente and Vlahov 2001). The typical approach is to ask the respondent or ego to name up to three, five, or more communication partners and to elicit information about each one (Burt 1984; Marsden 1987). For example, in two iterations of the General Social Survey, respondents were invited to provide the names or initials of up to five people they discuss important matters with (Burt 1984). Research

has focused on whether the ensuing information should be gathered by rows or columns (i.e., whether to ask the alter questions for each alter before proceeding to the next one or ask the attribute questions of each alter; the former approach seems to work best; Kogovšek and Ferligoj 2005). Collecting egocentric network data can be demanding because the data collection process involves eliciting a set of contact names using a name generator and then asking a series of questions about each person named.

An additional problem can occur, however, if the interviewers realize that the survey can be shortened if fewer names are elicited (Eagle and Proeschold-Bell 2015). If interviewers are incentivized on a completed interview basis rather than on a complete data basis, they may be inclined to elicit fewer network names. In other words, over time, interviewers may record smaller networks as they learn that eliciting smaller networks provides the same reward but with less work or time. An interview with no named contacts can take considerably less time and effort than one with three or four named contacts, depending on how many questions are asked of each alter.

Eagle and Proeschold-Bell (2015) reviewed prior studies on interviewer bias and note that variation in the number of names provided can depend on how well the interviewers follow the protocol, the tone they use when probing, and the placement of the name generator in the survey. Eagle and Proeschold-Bell (2015) demonstrated interviewer effects in their data collected from clergy in North Carolina. Although interviewer bias might often be treated as random and not affecting scientific results, Paik and Sancha-grin (2013) show that interviewer effects were responsible for estimates that core discussion networks of Americans were getting smaller.

In the present study, the name generator was restricted to three. Due to branching, the number of questions asked about each alter varied widely from person to person. The number of questions depended on responses to questions about alter sex, marital status, education, whether they had children, the age of the child, and whether their discussions were about pregnancy, breastfeeding, or both. The minimum number of questions (16) were asked of alters who were single men with no education or children, who discussed only breastfeeding with the respondent. The maximum number of questions (48) were asked of alters who were educated women with children under the age of 5 with whom the respondent discussed pregnancy and breastfeeding. Questions about alter background were not sensitive, although a number of questions about pregnancy and breastfeeding conversations required the respondent to think about who gave/received more information and who was more knowledgeable, which

may have taken some time to answer if the alter was a peer instead of a health-care worker.

The interviewers were compensated for completed interviews, regardless of how many alters were named. This created the potential problem of interviewers learning that it was in their best interest to record few named alters rather than prompting the participant to name up to three. This study seeks to understand if interviewers individually or collectively modified their data collection procedures over time to reduce network size and as a result, the length of time it took to administer a questionnaire.

## **Individual, Peer, and Group Learning Models**

Biased network size data could arise from at least three different sources when administering an egocentric network survey. First, individual learning could have occurred as interviewers conducted the interviews and learned that the time was much shorter if they elicited fewer network partners. Second, they could have learned this information from other interviewers. In other words, interviewers who learned this fact could have told other interviewers to restrict the number of names recorded. Third, the study participants, the women in the community, could have learned from each other. The study protocol was designed to collect a census of respondents from nine study areas. Consequently, the women interviewed early could have communicated to other women in the network, the hazards of eliciting too many network partners.

Evidence for individual interviewer learning would exist if individual interviewers decreased the recording of network size over time. Thus, a negative correlation between day of interview and network size would support an individual learning model. Evidence for a group learning or diffusion model, one in which interviewers told each other how to shorten the survey duration, would exist if network size decreased within the community, regardless of which particular individual did the interviewing. In this scenario, some interviewers may have learned to shorten the survey and then told others.

Finally, evidence for learning or diffusion among respondents would exist if we could map the networks of respondents and determine whether there were ties between later and earlier respondents and corresponding smaller and larger networks respectively. Unfortunately, if later respondents provided no or little network data, this hypothesis would be impossible to test and indeed this is the case with these data. The proximity of communities within different Community-based Health and Planning

Service (CHPS) zones varied by CHPS. In general, however, communities interact with other communities within and between CHPS zones on a regular basis. Market days, which occur weekly, often bring vendors from neighboring communities and CHPS. Funerals are another event that occur regularly and bring members from different communities together. The goal of this article is to determine whether there is evidence for individual or group learning. We hypothesize that we will find evidence for both individual and group learning.

## Methods

The parent study was designed to evaluate the effects of an intervention on the use of skilled birth attendance, antenatal and postnatal care, and breastfeeding behavior using a pre–post quasi-experimental survey design. This study used the baseline data collected from November 25, 2013, to February 1, 2014, with 99% of the interviews completed by January 10, 2014 (46 days). The study is being conducted in three districts in the Upper West Region of Ghana. Regional Ghana Health Service officials recommended three communities within each district for program implementation, and a census of women between the ages of 15 and 49 who had given birth to at least one child in 5 years preceding the survey was conducted. For more information on the parent study, see Community Benefits Health Baseline Report, July 2014 (<http://www.jsi.com/JSIInternet/Resources/publication/display.cfm?txtGeoArea=INTL&id=18359&thisSection=Resources>).

## Data Collection

Data were collected in three CHPS areas within each district (nine in total). Interviewers were recruited locally, and the surveys were administered using mobile phones by 28 data collectors who were hired and managed by the local research partner Kintampo Health Research Center. The project is being implemented by staff employed at John Snow International. The data collectors, all of whom had graduated from secondary school, attended a five-day training workshop in November 2013. The workshop covered the basic operation of the mobile phone (a Samsung Galaxy Ace Plus), operation of the survey software (SurveyCTO, v1.161), and overview of the content of the household survey. The data collectors also participated in role-play activities and extensive piloting of the survey among one another and in pairs in the community. During the training, errors in skip patterns and translations were corrected.

Data collectors were assigned to specific communities and were asked to reside there during data collection to familiarize themselves with the health services of the community. Three supervisors oversaw three groups of data collectors. The supervisors' role included advising on any questions about the survey, arranging the data collection schedule, and transmitting collected data from the phones to the online server using the available local mobile network. Supervisors also collected the mobile phones and charged them each day as electricity was not available in many of the communities.

Overall, 1,606 women were interviewed. Of these, 967 (60.2%) provided zero names, 359 (22.3%) provided one name, 200 (12.5%) two names, and 80 (5.0%) three names of people they talked to about ante- or postnatal care. Thus, 60.2% of respondents responded with no name when asked who they had "chatted with about breastfeeding or receiving care before or after pregnancy." If respondents named fewer than three, they were prompted with "Can you think of anyone else? How about sitting in on a conversation, even if you yourself didn't say anything?"

The 28 interviewers conducted between 26 and 82 interviews. All interviewers began on the first or second day of data collection except one started on day 3 and one on day 5. On average, interviewers were in the field for 45 days with a minimum of 22 and a maximum of 68. The average number of interviews conducted per interviewer was 57.4 ( $SD = 16.8$ ). Eight of the interviewers only interviewed in one CHPS area, whereas the remaining 20 conducted interviews in two or more areas. Two CHPS areas had only two interviewers, whereas one had as many as 16.

### **Control Variables**

In the analyses reported below, we included a set of sociodemographic variables: education, age, religion, number of children, being a member of a women's group, and listening to the radio at least once per week that might also be associated with network size (Table 1). Education ranged from no formal schooling (71.4%) to some or all primary (18.4%), some or all middle (8.8%), and some or all secondary (1.4%). Age had four categories: less than 20 years (4.6%), 20–29 years (43.3%), 30–39 years (40.1%), and 40–49 years (12.0%). We used dummy variables for religion: Protestant (7.2%), Muslim (9.9%), traditional (7.5%), and no religion (12.9%). Number of children was measured as one (19.1%), two–three (29.2%), four–six (38.4%), and seven or more (13.3%). Some 63.4% reported belonging to a women's group, and 49.9% reported listening to

**Table 1.** Univariate Statistics for Length of Interview, Network Size, Day of Interview, and Sociodemographic Control Variables.

	Mean (Standard Deviation) or %
Length of interview (in minutes)	44.7 (42.2)
Network size	0.62 (0.89)
Named 0	60.2%
Named 1	22.3%
Named 2	12.5%
Named 3	5.0%
Day of interview	23 (14.0)
Religion	
Catholic	62.5%
Protestant	7.20%
Muslim	9.80%
Traditional	7.50%
Nonreligious	12.9%
Age category	
Less than 20 years	4.60%
20–29 years	43.30%
30–39 years	40.10%
40–49 years	12.00%
Education	
None	71.40%
Completed primary	18.40%
Completed middle and above	10.20%
Parity	
1 Child	19.10%
2–3 Children	29.20%
4–6 Children	38.40%
7+ Children	13.30%
Belong to women's group	65.40%
Listen to radio 1+ times/week	49.90%

the radio once a week or more. Day of interview is the day from the start of interviewing when the interview took place (e.g., 5 = November 30).

### Analysis Plan

To test the hypothesis of whether later interviewers recorded smaller networks than earlier ones, we regressed network size on day of interview. Because the data were collected in nine CHPS areas, we used random effects regression specifying CHPS area as the random variable in the

overall model only. We included the control variables in all models, but these did not substantially affect the estimates between network size and day of interview, so we report only the bivariate coefficients (adjusted tables available on request).

To compare the group and individual learning models, we regressed network size on day of interview separately by district and CHPS area and then by individual interviewer. Significant negative correlations between network size and day of interview by district or CHPS area suggest group learning whereas significant negative ones by individual suggest individual learning. Because our research question assumes that recording more names during the interview takes more time, we check this assumption by analyzing the relationship between network size and length of interview.

## Results

Table 2 shows that overall there was no correlation between network size and day of interview ( $r = .04$ ), which was unchanged when controlled for age, religious preference, education, number of children, belonging to a women's group, and radio listenership. When a random effects model was used, however, with CHPS area as the random variable, the correlation was negative and significant. This suggests differences by CHPS area. By district, the correlation between size and time was negative for two districts (one,  $r = -.21$ ,  $p < .01$  and two,  $r = -.13$ ,  $p < .01$ ); but not the other. The overall Spearman rank order correlation between length and day of interview was  $.057$  ( $p < .05$ ), indicating that later interviews were longer.

By CHPS areas, three areas had significant ( $p < .05$ ) negative correlations (C, E, and F) and two had positive ones (B and G), suggesting that in some communities learning functioned to decrease network size whereas in others it increased network size. Figure 1 graphs the estimated bivariate regression lines between network size and day of interview for the nine CHPS areas. (In the graphs, we omitted interviews conducted during the last 19 days of interviewing as these were eight "mopping up" interviews.)

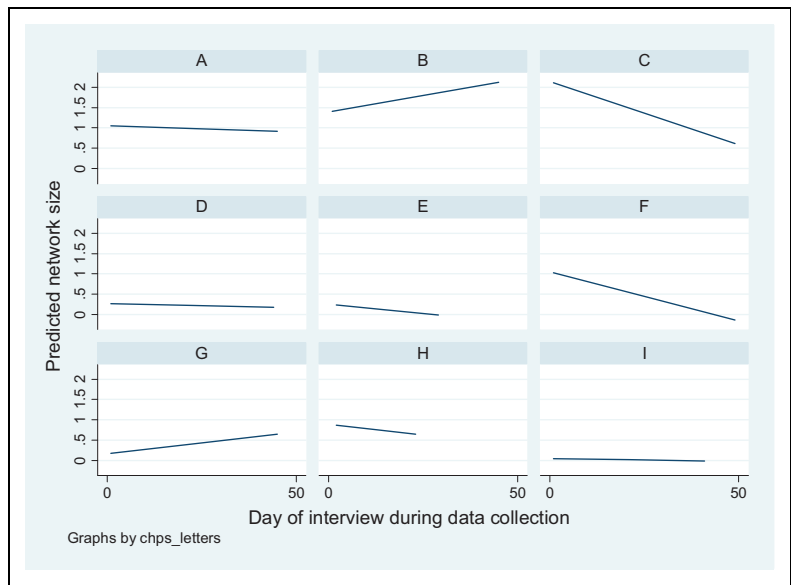
The individual interviewer correlations provided mixed results: Of all, five of the 28 were negative at the  $p < .05$  level and three were positive. Relaxing the probability threshold to  $.10$  to be considered statistically significant resulted in four more interviewers having positive associations between network size and day of interview.

To verify that larger social networks did indeed take longer to conduct, we calculated the length of time for each interview. Some 112 interviews

**Table 2.** Correlations of Network Size and Day of Interview ( $\beta$  Coefficients) and Network Size and Length of Interview, Overall, by District and by Community-based Health and Planning Service (CHPS) Areas.

Panel A: Correlations of Network Size and Day of Interview									
All	.04 (N = 1,606)								
District	One -.21** (N = 625)			Two -.11* (N = 524)			Three .04 (N = 457)		
CHPS	A	B	C	D	E	F	G	H	I
	-.04 (N = 234)	.26* (N = 97)	-.37** (N = 294)	-.06 (N = 111)	-.16* (N = 236)	-.25** (N = 177)	.22** (N = 203)	-.12 (N = 97)	-.12 (N = 157)
Panel B: Correlations of Network Size and Length of Interview									
All	.15**								
District	.28**			.16**			-.045		
CHPS	.27**	.05	.35**	.20*	.10	.33**	-.20**	-.07	-.01

\* $p < 0.05$ ; \*\* $p < 0.01$ .



**Figure 1.** Predicted values of network size by day of interview during data collection for each of nine Community-based Health and Planning Service areas.

**Table 3.** Length of Interview by Network Size.

Network Size	Average Length of Interview in Minutes	Standard Deviation
0	40.6	42.0
1	44.8	37.5
2	58.3	47.0
3	60.6	40.5
All	44.7	42.2

Note:  $F = 14.2$ .  
 $p < .001$ .

took longer than three hours and in some cases took days to complete when it was started on one day and completed on another. We recoded these 112 to 3 hours. (There were also three interviews that were completed in negative time, which is not possible, so these were recoded to the mean of 45 minutes.) Overall, there was a positive correlation between interview length and network size ( $\beta = .15, p < .001$ ). Table 3 reports the average length of time per interview by network size. The average interview length

was 40 minutes with no names and increased to 44.8, 58.3, and 60.6 for one, two, and three names, respectively ( $p < .001$ ).

Given variability by CHPS areas, we regressed network size on interview length separately for the nine CHPS areas. The results (Table 2) showed that the two CHPS areas with the highest negative associations between network size and day of interview (C and F) had the highest positive associations between network size and length of interview. One of the CHPS areas with a positive association between network size and day of interview (G) had a negative significant association. This may explain why learning reduced network size in some areas and not others. In other words, it is possible that some interviewers learned how to collect the network names and complete the survey within the same time, whereas others learned it was easier to record fewer names.

## Discussion

This study provides data from a natural experiment that highlights the challenges of survey data collection when one includes loops for multiple units and fields surveys for which the length may vary depending on responses. The results are mixed: Some communities and some interviewers seemed to have decreased the number of names they were able to elicit in response to an egocentric network name generator. Other communities and other interviewers were able to increase the number of names recorded as they became more experienced.

We find evidence for both individual and group learning with some areas showing an overall trend toward decreasing network size over time, whereas other areas had a positive trend of increasing network size over time. There was considerable interviewer variability in the elicitation of names and in how long it took to complete the surveys. We did not conduct analysis to determine whether there were interviewer characteristics associated with these outcomes as we do not have such data, but future studies should consider recording information on interviewer characteristics in case this variability occurs in other studies.

Our results have other implications. First, we noted that monetary incentives, given for completed interviews, may generate adverse outcomes for some interviewers if they try to maximize income over data quality. Non-monetary incentives such as ability to analyze the data or prizes for data quality may work better in some settings. For example, graduate students who depend on their data for their dissertations would be less likely to sabotage their own research to maximize their hourly rates. Incentives can

be tricky as using them to incentivize interviewers to collect more names may run the risk of eliciting network data that do not exist, with interviewers prompting (or insisting) on names when none, in fact, should or could be provided.

Second, there was considerable training and quality control in this study. Still, there was variability across interviewers, suggesting that some were more effective than others. Perhaps interviewer training should be conducted not only before data collection but also during it. In particular, using those interviewers who were collecting particularly high-quality data as the trainers. In addition, as technology for data collection has improved, daily tracking of interviewer performance can be conducted. For example, software can track the time spent on name generation to determine whether interviewers were spending enough time on the name generation. More time before moving on would suggest that they made an attempt to elicit more names and these data could be used by supervisors to identify interviewers who might need more supervision or retraining. Audio recording the interviews or at least the name generation section would also provide a way to create better management and better data. Making interviewers write notes explaining what the circumstances were and what probes they tried if there are fewer than three names might also reduce the incentive not to probe sufficiently.

Interestingly, in a qualitative scoping study to determine which areas were amenable to the intervention, the three areas deemed most functional and ready for health improvement were the three with the negative associations between size and day. This suggests that perhaps in these communities, there was the potential for more interinterviewer communication as the infrastructure and cultural mood in these communities was conducive to change. It is also possible that these three communities have residents ready for change and being enthusiastic about getting new health services also means they talk among themselves a lot and this resident interpersonal communication is what reduced the number of alters provided.

We verified that larger networks took longer to complete, thus providing an incentive to record none or fewer names. Moreover, the association between network size and length of interview varied by areas: It was positive in those areas with a negative association between size and day and negative in one of the areas that had a positive association between size and day. It seems some interviewers learned how to collect adequate network size numbers without increasing the time needed to complete the survey.

We cannot criticize the interviewers for reducing their and their respondents' burden by failing to prompt for (more) network names. Each additional name increased the number of questions from as few as 16 to a

maximum of 48, a considerable number. This can potentially prompt some interviewers to try less hard to elicit names. Researchers must be aware that this potential response bias exists and take corrective actions to monitor and avoid it. In sum, researchers are well-advised to employ short data collection instruments, and this may be particularly relevant when loops are involved. This is especially true when interviewers can communicate with one another, which they are apt to do.

There are some limitations to this study. First, it was not designed prospectively to test for interviewer bias, so it is a post hoc analysis of the data. Second, we did not measure the networks of the interviewers and could not estimate their potential influence on one another. Third, we did not interview the interviewers to determine if they knew the costs of prompting for more network names to be provided. Despite these limitations, these results have implications for studies using multi-item loops and the collection of egocentric data, namely, that there can be considerable variability in interviewer capability and commitment, which can affect data quality.

### **Authors' Note**

This study is part of a larger study titled The Community Benefits Health (CBH) project, which is part of the project Innovations for Maternal, Newborn & Child Health, an initiative of Concern Worldwide U.S., funded through a multiyear grant from the Bill & Melinda Gates Foundation.

### **Acknowledgments**

We thank our local research partners Kintampo Health Research Center, which implemented the data collection in the field. We thank Rebecca L. Davis for comments on earlier drafts. We also thank three anonymous reviewers who provided important insights.

### **Declaration of Conflicting Interests**

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

### **Funding**

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### **References**

- Bogen, K. 1996. The effect of questionnaire length on response rates: A review of the literature. In *Proceedings of the section on Survey Research Methods*, 1020–25. Alexandria, VA: American Statistical Association.

- Burt, R. 1984. Network items and the general social survey. *Social Networks* 6: 293–339.
- Campbell, K. E., and B. A. Lee. 1991. Name generators in surveys of personal networks. *Social Networks* 13:203–21.
- Converse, J. M., and S. Presser. 1986. *Survey questions: Handcrafting the standardized questionnaire*. Los Angeles: Sage.
- Davis, R. E., M. P. Couper, N. K. Janz, C. H. Caldwell, and K. Resnicow. 2010. Interviewer effects in public health surveys. *Health Education Research* 25: 14–26.
- Eagle, D. E., and R. J. Proeschold-Bell. 2015. Methodological considerations in the use of name generators and interpreters. *Social Networks* 40:75–83.
- Flores-Macias, F., and C. Lawson. 2008. Effects of interviewer gender on survey responses: Findings from a household survey in Mexico. *International Journal of Public Opinion Research* 20:100–10.
- Josten, M., and M. Trappmann. 2016. Interviewer effects on a network-size filter question. *Journal of Official Statistics* 32:349–73.
- Kogovšek, T., and A. Ferligoj. 2005. Effects on reliability and validity of egocentered network measurements. *Social Networks* 27:205–29.
- Kosyakova, Y., J. Skopek, and S. Eckman. 2014. Do interviewers manipulate responses to filter questions? Evidence from a multilevel approach. *International Journal of Public Opinion Research* 27:417–31.
- Marsden, P. V. 1987. Core discussion networks of Americans. *American Sociological Review* 52:122–31.
- Marsden, P. V. 2003. Interviewer effects in measuring network size using a single name generator. *Social Networks* 25:1–16.
- McCarty, C., H. R. Bernard, P. D. Killworth, G. A. Shelley, and E. C. Johnsen. 1997. Eliciting representative samples of personal networks. *Social Networks* 19: 303–23.
- Paik, A., and K. Sanchagrin. 2013. Social isolation in America: An artifact. *American Sociological Review* 78:339–60.
- Rolstad, S., J. Adler, and A. Rydén. 2011. Response burden and questionnaire length: Is shorter better? A review and meta-analysis. *Value in Health* 14: 1101–08.
- Valente, T. W. 2002. *Evaluating health promotion programs*. Oxford, UK: Oxford University Press.
- Valente, T. W. 2010. *Social networks and health: Models, methods, and applications*. Oxford, UK: Oxford University Press.
- Valente, T. W., and D. Vlahov. 2001. Selective risk taking among needle exchange participants in Baltimore: Implications for supplemental interventions. *American Journal of Public Health* 91:406–11.