Human Organization, Vol. 60, No. 1, 2001

0018-7259/01/010028-12$1.70/1

Comparing Two Methods for Estimating Network Size

Christopher McCarty, Peter D. Killworth, H. Russell Bernard, Eugene C. Johnsen, and Gene A. Shelley

In this paper we compare two methods for estimating the size of personal networks using a nationally representative sample of the United States. Both methods rely on the ability of respondents to estimate the number of people they know in specific subpopulations of the U.S. (e.g., diabetics, Native Americans) and people in particular relation categories (e.g., immediate family, coworkers). The results demonstrate a remarkable similarity between the average network size generated by both methods (approximately 291). Similar results were obtained with a separate national sample. An attempt to corroborate our estimates by replication among a population we suspect has large networks (clergy), yielded a larger average network size. Extensive investigation into the existence of response effects showed some preference for using certain numbers when making estimates, but nothing that would significantly affect the estimate of network size beyond about 6 percent. We conclude that both methods for estimating personal network size yield valid and reliable proxies for actual network size, but questions about accuracy remain.

Key words: network size, hard-to-count populations, telephone survey

Since 1986, we have been working on a method for estimating the distribution of personal network size across the U.S. population and the size of hard-to-count subpopulations, such as those who are HIV positive, the homeless, and rape victims. A major component of our method involves estimating the average size of personal networks for a large sample of people using what we term a "scale-up method." There have been only a handful of studies on estimating personal network size, including the reverse small-world studies (Bernard, Killworth, and McCarty 1982; Killworth and Bernard 1978), the telephone book studies (Freeman and Thompson 1989), our initial attempts (Bernard et al. 1989; Killworth et al. 1990; Bernard et al. 1991; Johnsen et al. 1995), and the current scale-up method studies (Killworth et al. 1998a, 1998b). These methods yield widely varying estimates of network size, due in part to the definition of who should be included in a respondent's network and also to characteristics of the methods themselves.

This paper critically examines our methodology to date and introduces an extension to a second, parallel "summation method." We begin by briefly describing the scale-up method and then define the new summation method. The results of the two methods are then compared and found to be similar. We then discuss findings from focus groups that suggest various possible confounding effects in our methods. These include: number preference by respondents; whether the consistency between methods and between surveys is an artifact produced by our numerical approaches; missing data; whether respondents chosen for their large network size modify the results; and whether respondents are less able to estimate accurately for extremely small and extremely large subpopulations.

The Scale-Up Method

Our method is based on the assumption that the number of people a person knows in a particular subpopulation is a function of, among other things, the number of people known overall. The method begins from an assumption that has received a variety of tests in Killworth et al. (1998a, 1998b). It assumes that, other things being equal, the probability that any member of the respondent's network is in a subpopulation is the fraction of the larger population occupied by the subpopulation. (In other words, if 1/100th of the U.S. population have some characteristic, then on average one would assume 1/100th of any network to possess that characteristic also.)

Chris McCarty is survey director for the Bureau of Economic and Business Research at the University of Florida, Gainesville; Peter Killworth is an individual merit scientist at Southampton Oceanography Centre, Southampton, U.K.; Russ Bernard is professor of anthropology at the University of Florida, Gainesville; Eugene Johnsen is emeritus professor of mathematics at the University of California, Santa Barbara; and Gene Shelley is at Georgia State University, Atlanta. *We are indebted to John Domini, Amber Yoder, and David Kennedy for the help they provided in data collection and during the writing of this survey. The Bureau of Economic and Business Research at the University of Florida provided telephone survey services as well as a place for us to work during analysis and write-up. The work was funded through NSF grant SBR-9710353.*

More formally, let m be the reported number of people known in a subpopulation E (rape victims, for example) of some larger population T (such as the U.S. population), c be the personal network size of the respondent, t be the size of T, and e be the size of the subpopulation. Then the probability that any member of c is in E is e/t, so that the *expected* number known would be given by

$$\frac{m}{c} = \frac{e}{t}. \tag{1}$$

We do not assume that the simple proportionality applies to any specific individual: the number of people reported to be known in the subpopulation has a binomial distribution with probability $p = e/t$ and mean cp. The respondent is asked about many subpopulations, and from the responses given, a maximum likelihood estimate of the respondent's c is computed (i.e., one which best fits the pattern of responses elicited). Once the network size for each respondent has been estimated (and the distribution of this is of great theoretical interest), a similar maximum likelihood estimate is computed for the fractional size of the unknown subpopulations. Each respondent has a c value and reports how many he or she knows in the unknown subpopulation. The maximum likelihood estimate is constructed from the pattern of responses across all respondents.

The scale-up method is based on three further assumptions:

1. Everyone in T has an equal chance of knowing someone in E; that is, everyone in the U.S. has an equal chance of knowing someone who is a diabetic, for example. This chance increases proportionally with the size of E.

2. Everyone has perfect knowledge about all people they know. That is, if someone in your list of social network members is a diabetic, then you know this fact. You also know whether your network members have a twin, have ever had typhoid fever, build houses for a living, have an American Express card, and so on.

3. Respondents can tell us accurately and in a very short time (less than 30 seconds) the number of people whom they know who are, say, Native Americans, diabetics, or golfers. ("Knowing" someone is defined as follows: "you know the person and they know you by sight or by name; you can contact them in person, by telephone or by mail; and you have had contact with the person in the past two years.")

We call a violation of the first assumption the "barrier effect." That is, there are spatial and sociodemographic characteristics of respondents and the people they know that create barriers for some respondents to know some types of people. (Conversely, respondents who are members of some subpopulation may be *more* likely to know other members of the subpopulation. However, in a nationally representative sample of respondents, and for the sizes of subpopulations used here, it is easy to show that this aspect has a negligible effect on the computations.)

We call a violation of the second assumption the "transmission effect." That is, we know that the fact of membership in a subpopulation is not transmitted with equal probability to all network alters (people they know). This is due either to: 1) the stigma or embarrassment associated with that membership; 2) because it is not a common subject of conversation; or 3) it is personal and private and usually not discussed even with friends (we usually don't divulge the amount of our personal wealth even though it may be the subject of speculation among our friends and acquaintances). Barrier and transmission effects are being investigated in a separate study. We call a violation of the third assumption the "estimation effect" and discuss this below.

Another reason respondents may not be able to report accurately about the number of people they know in a particular subpopulation is that they are unsure about the subpopulation boundary; that is, the term is ambiguous. For example, a respondent may know someone who started doing Web page design at home in the past year, but s/he is unsure whether that person counts as someone who "opened a business in the past 12 months." Indeed, the definition used by those who compiled the U.S. Statistical Abstract is very specific and would reflect data available from sources such as records of those who have incorporated. Since respondents are not usually privy to these definitions, or likely to be able to interpret them consistently, this may affect the accuracy of our estimates.

The extent of this error almost certainly varies widely depending on the particular subpopulation. The range of definitions of "homeless" (see, for example, the U.S. Bureau of the Census) demonstrates how difficult it would be to define this subpopulation unambiguously for respondents. The same is true of other subpopulations, such as Native Americans, or in the case of Mexico, its Indian population. On the other hand, some subpopulations are easy to define, such as diabetics or women who have given birth in the past year.

The problem of ambiguity of subpopulation boundaries in the eyes of respondents is, to a great extent, unavoidable. We are limited to subpopulations for which we have counts. And often the definitions used by the government entities that gather them are too specific for a respondent to grasp. Further, if we rely on subpopulations that are unambiguously defined, we may find they are of a specific type, such as medical conditions or well-defined events. This may introduce a source of error since certain people may be more likely to know people in these specific subpopulations. In other words, the benefit of using subpopulations of different types may be compromised.

Ambiguity of survey questions is certainly not unique to our research. Most texts on survey research and the construction of questionnaires stress the importance of avoiding ambiguous terms and phrases. Research has shown that in some cases the effect can be significant (Fowler 1992:218). We should note that the ambiguity effect is relevant for both the scale-up and the summation methods. In some cases respondents may be unsure (or disagree) whether someone they

Table 1. Average Number of People Known for Subpopulations and Relation Types
(combined surveys with a total of 1,370 respondents)

29 Populations for Scale-up Method		16 Relation Types for Summation Method		3 Populations to be Estimated	
Michael	4.8	Immediate family	3.5	HIV positive	0.7
Christina	1.3	Other birth family	24.0	Women raped in past 12 months	0.2
Christopher	1.8	Family of spouse or significant other	12.3	Homeless	0.7
Jacqueline	0.7	Coworkers	35.6		
James	3.4	People at work but don't work with directly	62.1		
Jennifer	2.3	Best friends/confidantes	4.3		
Anthony	1.7	People known through hobbies/recreation	12.3		
Kimberly	1.4	People from religious organization	43.4		
Robert	4.1	People from other organization	17.1		
Stephanie	1.3	School relations	18.3		
David	3.5	Neighbors	12.8		
Nicole	1.1	Just friends	22.6		
Native Americans	3.5	People known through others	22.6		
Gave birth in past 12 months	3.6	Childhood relations	6.8		
		People who provide a service	7.7		
Women who adopted a child in past 12 months	0.3	Other	3.9		
Widow(er) under 65 years old	3.2				
On kidney dialysis	0.6				
Postal worker	2.2				
Commercial pilot	0.7				
Member of Jaycees	1.1				
Diabetic	3.3				
Opened a business in past 12 months	1.1				
Have a twin brother or sister	2.0				
Licensed gun dealer	0.5				
Came down with AIDS	0.4				
Males in state or federal prison	1.0				
Homicide victim in past 12 months	0.2				
Committed suicide in past 12 months	0.2				
Died in auto accident in past 12 months	0.5				

know falls into a particular relation type (such as someone they know through a hobby or organization).[1]

In the first national telephone survey of 1,554 respondents across the U.S. in which we applied our scale-up method, we asked respondents to estimate the number of people they knew in 29 subpopulations of known size, such as people with a particular first name, victims of motor vehicle accidents, and diabetics. Table 1 shows a list of the 29 subpopulations.

This survey generated an average network size of 286 (Killworth et al. 1998b). We are confident the scale-up method produces a useful and reasonable estimate for network size, but there are problems. For one thing, there is an apparent tendency for respondents to overreport the number of people they know for small subpopulations and to underreport for large. This is due, in part, to reliance on the three assumptions above.

The Summation Method

The assumptions of our model suggest a program of research, the goal of which is at best to estimate accurately, and at least to minimize, the barrier, transmission, and estimation effects and try to correct for them. This would improve the accuracy of our estimation of network size and our estimates of subpopulation sizes. Our current efforts along these lines will be reported elsewhere. To minimize the transmission and barrier effects, we have tested a completely different method for estimating network size. Instead of asking people to count their network alters who are members of various subpopulations, we ask people to count the number of those alters who stand in various relations to the respondent (kin, coworkers, etc.). The list of 16 relation types is also shown in Table 1. (Note that the category "Other" is a catch-all for network members who have not been elicited otherwise.)

Using relation types to estimate network size instead of countable subpopulations offers several potential advantages:

1. It may be easier for respondents to make smaller estimates than to think about who fits into a known subpopulation among all the people they know.
2. This method is quicker to implement as we currently must collect 20-30 subpopulations of known size to estimate network size, c.
3. The scale-up method relies on accurate counts for some subpopulations. These are difficult to obtain, particularly in developing countries.
4. It virtually eliminates the transmission and barrier effects from the estimate of c. Respondents almost always know who is and who is not of a particular relation type (e.g., family relation, work relation, etc.) where they do not necessarily know if a network member is a diabetic or an American Indian (see assumption 2, above). Also, we do not expect spatial or sociodemographic barriers to knowing network members of a particular relation type, and this eliminates the barrier effect.
5. The summation method is independent of the scale-up method for estimating c, which lets us use respondents' estimates of subpopulations of known sizes as a way of checking the accuracy of our estimates for subpopulations of unknown sizes.

Brewer (1993, 1995a, 1995b) and Brewer and Yang (1994) also present evidence that respondents would find relation types an easier set on which to report.

Using relation types to estimate network size has some potential disadvantages:

1. There is no way to check the validity of respondent estimates of the number of people they know in various relational categories since the sizes of those categories are not known. The scale-up method is based on countable subpopulations, which provides a way to test the statistical integrity of estimates of c. (See Sudman, Bradburn, and Schwarz 1996 for their discussion of counting versus estimation in surveys.)
2. Counting a network member more than once (say as a work relation and as someone with whom they socialize) may be common. This type of systematic error would result in an inflation of c and a deflation in the estimate of any unknown subpopulations.

Survey Design

We tested and compared the scale-up and summation methods for estimating personal network size across four national telephone surveys in the U.S. (Survey 1, n = 796, January 1998, cooperation rate = 41%; Survey 2, n = 574, January 1999, cooperation rate = 35%; Survey 3, n = 159, June 1999, cooperation rate = 54%; and Survey 4, n = 426, June 1999, cooperation rate = 44%). In each survey, respondents were presented with both methods for estimating c and also provided some demographic information. The scale-up method took an average of 7 minutes while the summation method took 5 minutes. On average each estimate took 15 seconds per subpopulation using the scale-up method versus 18 seconds for the summation method.[2] On all occasions, the scale-up questions preceded the summation questions; this is a potential shortcoming which will be remedied in future surveys.

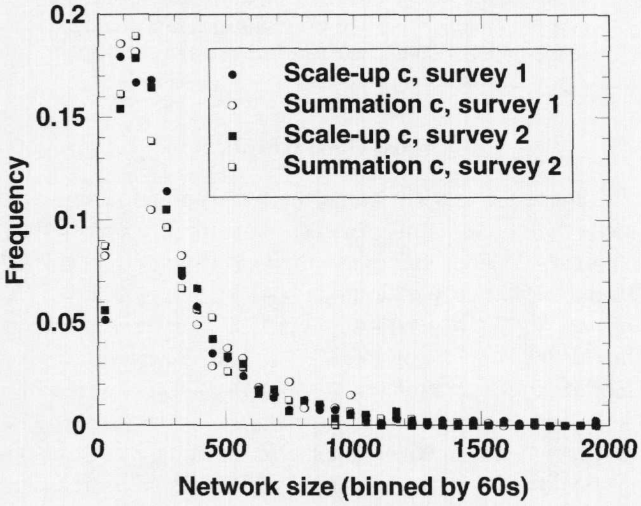
For surveys 1, 2, and 4 telephone numbers were generated by random digit dialing using a database that eliminates telephone banks that are primarily commercial. Random digit dialing has the advantage of including households with unlisted numbers in the sample. Respondents within the household were chosen by first asking for the youngest male, and failing that, the oldest female in the household. This method tends to balance a natural bias toward a disproportionate number of female respondents. Surveys were conducted in English and Spanish. Telephone numbers were finalized as unproductive after 10 calls, and refusals were normally called twice.

Survey 3 was a listed sample of clergy purchased through a nationally known sampling service. Clergy included priests, reverends, and rabbis. Respondents were randomly selected from the list using a random number assigned by the Statistical Analysis System. Given the listed sample there was no randomization of respondent selection within the household. This survey was conducted only in English.

Comparison of *c* Values

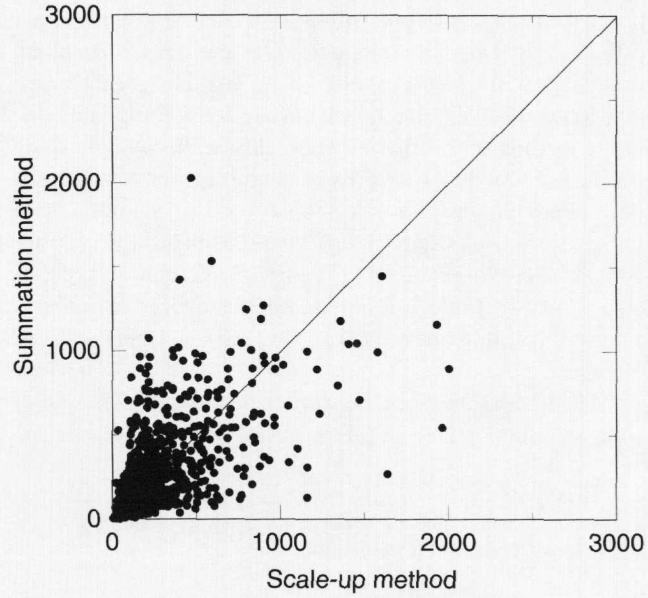
Figure 1 shows the distribution of network size using the two methods. The closed circles represent the distribution for the scale-up method in our original survey of 796 respondents (Survey 1), while the open circles represent the distribution for the summation method (the closed and open squares refer to the replication survey of 574 respondents [Survey 2], to be discussed later). A Kolmogorov-Smirnov test of equality shows these distributions to be statistically different. Visually, however, these distributions appear

Figure 1. Frequency of Network Size for Surveys 1 and 2



Note: Closed symbols show size computed from the scale-up method; open symbols from the summation method.

Figure 2. Scatterplot of Network Size Estimates from Survey 1, Showing Scale-up Estimates Against Summation Method Estimates



remarkably similar; indeed, for Survey 1, the scale-up method yielded a mean network size of 290.8 (SD = 264.4) compared to 290.7 (SD = 258.8) for the summation method.

The similarity of the means is so striking that we were led to ask whether there might be some instrument effect. If, for example, respondents put little or no thought into their estimates of how many people they knew in each of the subpopulations or relation categories (and simply made up similar results), we would expect scale-up and summation estimates to be similar for individual cases. Figure 2 plots the two network size estimates for Survey 1. If the two estimates were really similar, we would expect little dispersion about the diagonal. There appears, however, to be significant variability at all levels of network size.

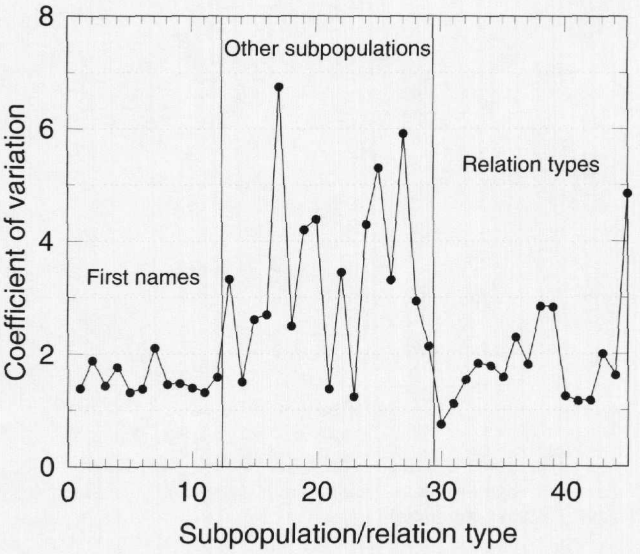
Indeed, the correlation (0.56 at p = 0.0001) implies variation between the two estimates. In other words, it is often the case that one estimate is low while the other is high. Yet over the entire sample of 796 respondents for Survey 1 this balances out to yield similar estimates. It appears, then, that the methods independently corroborate the estimate of about 291 for mean network size, for the survey's definition of networks.

Another possibility is that respondents might become tired, less focused, and prone to simple repetitions as the interview progresses.[3] Recall, too, that respondents first gave estimates of how many people they knew in the various subpopulations and were then asked to estimate the number of their network alters in the various relation types. By the time they got to the latter estimates, respondents may have fallen into a pattern of repetition. This might not cause the estimates to be the same, but it is a cause for concern.

Figure 3 shows the coefficient of variation[4] of the reported number known (i.e., *m*) for each subpopulation and relation type in the order the questions were asked. If respondents had fallen into a pattern where they tended to give similar estimates for all subpopulations and relation types as the interview progressed, we would expect the line to become flatter. With the exception of a relatively flat line beginning with the eighth first name and ending with the last first name, the line shows considerable variance for these estimates. This is particularly true of the relation category "Other," and countable subpopulations such as "people who are on kidney dialysis" or "victims of a homicide." It appears, then, that respondents do not fall into a pattern of repeating similar estimates as the interview progresses. As further proof, we computed the mean coefficient of variation within each of the three types (names, other subpopulations, and categories) for each respondent and averaged these across respondents. The three means (standard deviation) were 1.09, 1.90, 1.55 (0.45, 0.66, 0.56), which are all highly significantly different.

Given the striking similarity of the average network size generated by the two methods, we decided to replicate the survey with a separate sample (n = 574), also shown in Figure 1. In this sample, the estimate for network size for the scale-up method was 291.2 (SD = 259.3) and that for the summation method was 281.2 (SD = 255.4). Again, scale-up and summation methods yield essentially identical distributions, both between themselves and compared with Survey

Figure 3. Coefficient of Variation of the Average Reported Number Known in Subpopulations and Relation Types, in Interview Order



1. This appears to be striking confirmation of the reliability of the methods. Based on the similarity of the distributions in Figure 1 and on the consistent means for network size across the two surveys, we combined the data from the two surveys (n = 1370) for further analysis.

Table 1 shows the results. Across the 32 subpopulations representing the scale-up method in Table 1—including three subpopulations for which independent corroboration of size is not available—Pearson's r for the mean reports for the two surveys is 0.99. Similarly, using reported numbers in the 16 subpopulations representing the summation method in Table 1, Pearson's r for the two surveys is 0.99. (Note, of course, that we are correlating items that tend to covary, so that much of this correlation may be expected.)

Estimates can also be made for subpopulation sizes using the methods of Killworth et al. (1998b), which showed that the estimates scale proportionally to the mean number reported over the mean personal network size. Since both of these quantities are almost identical in the two surveys, it follows that back-estimates of network size are also identical (and hence not shown).

The consistency between the two methods—in the visual distribution of network size and the estimates of subpopulation sizes—suggested we continue to examine the process by which people make these estimates when we ask the question. In the absence of a standard instrument for measuring network size and distribution, we cannot determine which of the two methods produces the more reliable and more valid results. We decided we needed more information about the actual process respondents go through in providing estimates. An obvious way to acquire information is focus groups. The focus group is a valuable tool for eliciting text about process because it takes advantage of interactions within the group—as one respondent discusses a topic, other respondents hear what is said; this triggers discussion and counter-responses by other respondents.

Focus Groups

We conducted two focus groups in Gainesville, Florida: a group of nine men and a group of seven women. Ages in both groups ranged from the early 20s to over 60. Respondents to the focus groups were recruited by an undergraduate assistant at parking lots and a shopping mall. Respondents were asked if they would be willing to come to the university, respond to a survey, then be interviewed for approximately two hours. Those who agreed to participate were paid $75.

Participants first answered the survey in person with one of the interviewers who had conducted the survey over the telephone. We wanted participants to experience the process in the same way as did the telephone respondents. During the focus groups we concentrated on the following questions: "What were your general impressions of the survey. Did it appear strange to you?"; "Were you uncomfortable answering any of the questions?"; "Describe the process you went through to estimate the number of people you know named Michael. Do you think the estimate was accurate? Do you think you missed any?"; "Did you have enough time to provide these estimates?"; "Were some groups more difficult to estimate than others?"; "Is the definition of knowing someone reasonable to you? Does the definition leave out important people?"; and "Do the relation types used conform to your personal network? Are some categories too big to estimate?"

Overall, the focus group participants were interested in the study. The 29 countable subpopulations in Table 1 were chosen because they represent a variety of subpopulation types for which reliable counts are collected annually. Some of the men, however, were suspicious of our intentions in asking how many gun dealers they know.

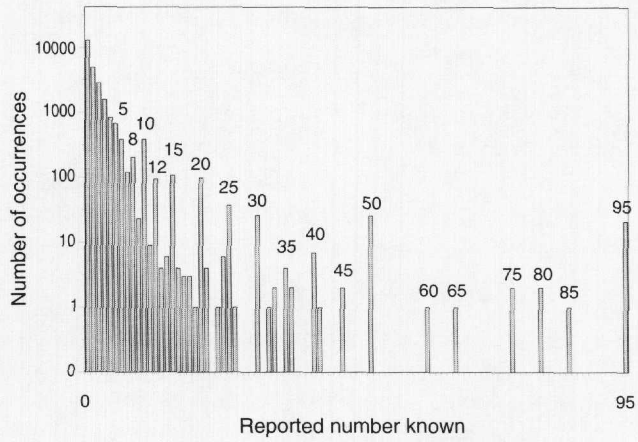
Most participants felt that our definition of "knowing" was appropriate, although a few were concerned some of their important network alters would be left out given the two-year cutoff. Even though they had not talked to some people for more than two years, some participants said they could pick up the relationship immediately where it left off. Most participants, though, also agreed that these contacts did not greatly affect their lives.

The most valuable insights into the process of answering our survey questions came when we asked focus group participants to recall the process they went through in estimating specific subpopulation sizes in their personal networks. All participants agreed that for some subpopulations, primarily the large ones (people named Michael, for example), estimation was difficult in the short time available.

Several people mentioned the difference between counting and estimating (a topic discussed by Sudman et al. 1996). For relatively small groups, like gun dealers, they said they enumerated; for others, like Native Americans and people named Michael, they came up with an estimate. Some participants said explicitly that they relied on their "feel" for how large the group was and how likely it is that they knew someone in that group. When asked what numbers they used to operationalize those feelings, it became apparent those numbers varied widely across participants.

The finding from the focus groups that respondents guessed for some subpopulations, and that some respondents told us they used the same value repeatedly, led us to examine the survey data for evidence of number preference. If respondents select systematic guessing over counting for many of the subpopulations, and if the guessing values vary between respondents, this could explain similar estimates for network size between replications of the scale-up method and between the scale-up and summation methods. Although they appear fundamentally different, both methods rely on respondents estimating the number of people they know in a given category. Systematic estimation that is unrelated to the network size would be a serious error. On the other hand, guessing could be systematic, but related to network size.

Figure 4. Number of Occurrences of Reported Number Known in Any Subpopulation or Relation Type Across 796 Informants



Note: The y-axis is scaled logarithmically, but includes zero. This inflates some small values visually, but is unavoidable given the wide range of values present.

Number Preference

Rounding to prototypical values is the phenomenon in which, when requested or implicitly invited to supply a number, such as a count or estimate in a category, people respond with preferred integers (Myers 1940; Turner 1958; Zelnick 1961, 1964; Winick 1962; Stockwell 1966; Stockwell and Wicks 1974; Wicks and Stockwell 1975; Huttenlocher et al. 1990; Baker 1992). Such commonly reported values might be 0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50...100...200, which show a preference for terminal digits 0, 5, 00. More generally, however, the set of preferred values may include other numbers, such as 2, 8, and 12. For succinctness we shall adopt a simple previously used term (Myers 1940; Turner 1958) and refer to this rounding phenomenon as heaping, with the values that exhibit this phenomenon called heap values. Heaping on preferred numbers generally occurs in conjunction with the underrepresentation of certain other numbers, such as 7, 9, 11, 13, 14, 16. In simple estimation this latter phenomenon typically exhibits the avoidance of terminal digits 1 and 9, sometimes 3 and 7, and possibly 4 and 6 (Turner 1958).

Figure 4 shows an example of heaping. The figure shows the total number of reports of 0, 1, 2, 3, 4,..., summed over all subgroups from the first survey, using a logarithmic scale, here including zero for clarity (since the most common response is 0). Selected values that are large compared with those around them are indicated in the diagram. It is clear that numbers with terminal digits 0 and 5 are preferred, especially for responses above 10, and other numbers (e.g., 7, 9, 11) are preferentially unused. Numbers above 20 are preponderantly multiples of 5. Although Roberts and Brewer (2000) present some measures of heaping in network data, in our data the evidence is already clear that heaping is occurring.

Could heaping be modifying our results? On the face of it, the effects of heaping would seem hard to undo, since we only observe their results and not the conditions before the numbers were modified by the preference. And the focus groups had given us little insight into how the heaping process was carried out. We have attempted to analyze the effects of heaping in two ways: 1) by modeling how the data might change, and how our results would change, by using plausible initial data for the summation data; and 2) by a variety of ad hoc methods to reverse heaping given the final scale-up data.

Modeling How the Data Might Change and Our Results Would Change

There is strong evidence (given below) that the heaping in Figure 4 has little effect on the estimates from the scale-up method. However, the same need not be true for the relational categories, since they are summed to provide a c estimate, and so errors can compound. To investigate possible effects, we constructed an ad hoc frequency distribution $f(k)$ of the number of responses for each value of k, $0 \leq k < 134$, as shown in Table 2.

The total number of respondents here is 800, which is similar to that for Survey 1. This frequency function is monotonically nonincreasing, starting off like a negative exponential but slowly flattening out so that there is a nontrivial tail

Table 2. Frequency Distribution for an Ad Hoc Distribution of Responses for a Typical Relational Category

k	$f(k)$
0	268
1	135
2	69
3	37
4	21
5	13
6	10
7	9
8	8
9	6
10-12	5 each
13-20	4 each
21-35	3 each
36-68	2 each
69-134	1 each

out to the highest value. It is an idealized distribution, similar to those encountered with the responses for the relational categories in Survey 1. For these latter subpopulations, the early values are similar, the distributions decrease in a similar manner, and there are nontrivial tails going out at least as far as in Figure 4, but with occasional high values due to heaping. The mean response for this idealized distribution is 16.09, while the mean response for the 16 relational categories is 19.54.

In our surveys the imposed ceiling values to encode the responses for the different subpopulations were set to 95 for the 12 name and 20 attribute subpopulations, and varied over 99, 100, 150, and 500 for the 16 relational categories. It appears that some actual responses may have exceeded those ceiling values, since the rather large frequencies for some of them seem out of proportion to the responses for lower heap values and thus may be artifactual. This effect will also be modeled using this ad hoc distribution.

Now, suppose this idealized distribution is the true underlying distribution for the respondents, but that respondents default to heap values. How does heaping occur? The following mechanism is a possible cognitive model for this phenomenon. Suppose we have the two adjacent heap values 10 and 25, where respondents do not report any numbers in between but rather default to the closest heap value. If an informant has a "true" number of people known in the subpopulation (according to the idealized distribution) that lies in the first half of the interval 11–24, namely 11–17, it is reported as 10, and if it lies in the second half, 18–24, it is reported as 25. If the interval has an odd number of values, we assume the middle value defaults half the time to the lower heap value and half the time to the upper one. We now assume the above defaulting occurs between every pair of adjacent heap values, and that "true" values which are already heap values do not change but are reported accurately.

We compared the mean response for the default distribution containing heap values to that of the idealized distribution from which it was derived. For fixed values of t and e the percentage change in average m will approximate the percentage change in the estimated value of c from (1). We assume the heap values here are those less than or equal to 100, which occur for at least half of the 16 relational categories in Survey 1, and estimate the effect of heaping on average m. For the heap values 0, 2, 5, 8, 10, 12, 15, 20, 25, 30, 40, 50 and 100, we find that the average m is 15.08, 93.8 percent of the original 16.07, for an error of about –6.2 percent. The resulting estimated value of c would thus be low by a little over 6 percent.

Certain other combinations of heap values from the above list for the relational categories yield similar errors. This example also incorporates the effects of the artifact that occurred in a few cases within the relational data; namely, imposition of ceiling (heap) values which were too low for encoding the actual responses. In this case all responses above the ceiling value are rounded downward, producing an artificially low value of c for those relational categories where this occurred. Other scenarios for heaping are being investigated and will be reported in more detail later (Johnsen et al. n.d.).

Ad Hoc Methods to Reverse Heaping in the Scale-Up Method

We tried to undo any effect of guessing (whether rounded or not) by changing each reported answer to a random number uniformly distributed in the range 50–150 percent of that reported answer. In other words, we assume that respondents guess in an unbiased way. Almost no change in the mean c was found (291.0), and no estimate of any subpopulation size was changed by more than 5 percent, usually much smaller. Since this change was stronger than simply modifying round-numbered responses alone, we conclude that heaping has only a small effect on our results, which are thus robust to this form of error.

Nonetheless, when and where heaping takes place can have an effect. Consider (3) in Killworth et al. (1998b: 293) maximum likelihood estimate for each respondent's c:

$$c_i = t \cdot \frac{\sum_{j=1}^{L} m_{ij}}{\sum_{j=1}^{L} e_j} \quad (2)$$

where the suffix i refers to respondent and j to each of the L subpopulations (so m_{ij} is the number of people respondent j reports knowing in subpopulation j, whose size in the total population is e_j). We could assume that informant responses were unreliable for values above some cutoff value k. To model this in a simple fashion,[5] whenever an m_{ij} is above k, we multiply both that m_{ij} and the respective e_j in numerator and denominator sums by an arbitrary value of 0.2 (rather than an implicit unity). We find that the results are strongly dependent on this change to the method. Even with a cutoff of 10, the mean c drops to 233, a 20 percent decrease. In other words, to reproduce our findings, the large responses (which correspond to informants with large c's and to large subpopulations) must be given full weight. The dependence of solutions on the subpopulations chosen for study will be discussed more fully in a later paper.

Consistency Between *c* Estimates Is Not an Artifact

We have reported findings that show great consistency between two different methods of computing c, and, therefore, also in back-estimates of subpopulation size. However, the two methods both share the necessity of asking respondents to estimate the number known in various subpopulations: in one case, selected subpopulations (scale-up) and in the other, relation categories (summation). We were concerned, therefore, that the agreement might be illusory: essentially almost any response to the tasks given might, after the data were processed, have yielded similar answers. To test this, we examined the effects of changes to the data on the results, mainly the mean of c.

We have argued that respondents may guess their responses once their estimated number rises above some cutoff value. Now, we computed that if respondents estimated numbers at least 5, the average number estimated was 5.24. We thus changed reported m values at or above 5 to a value of 5 precisely. This produced large changes, with the mean c dropping to 206, a change of 29 percent. Changes to the subpopulation size estimates were of a similar order, varying in size and direction. (Estimates of small subpopulations rose, since reported knowledge of these was usually unchanged in the data and only the mean c estimate had changed.)

This was repeated, replacing values at least 5 by 10 (i.e., nearly doubling the average value reported above 5). There was little change in the mean c, at 284. A similar answer was found setting values at least 5 to a uniformly distributed random value between 5 and 15. We repeated the random change (5–15) above, but only for large subpopulations (with $e > 1$ million). This increased the mean c to 402, a change of 38 percent but in the opposite direction. Again, subpopulation estimates changed by a similar amount. We then made a smaller, but again systematic, change. Subpopulation estimates above 10 were replaced by 10. The average c dropped to 245, a change of 16 percent.

These experiments show that some changes to the data can produce large changes in the results. Apparently small changes (replacing m by 5 when it is above 5, when the existing average of such data is 5.24) gives a large change in the output. Conversely, there are apparently large changes in the data (replacing m by 10 in such cases) that produce a smaller change, at least in the mean c. Thus we can safely reject the suggestion that "any reasonable estimates of large subpopulations would yield similar answers with this method."

Missing Data

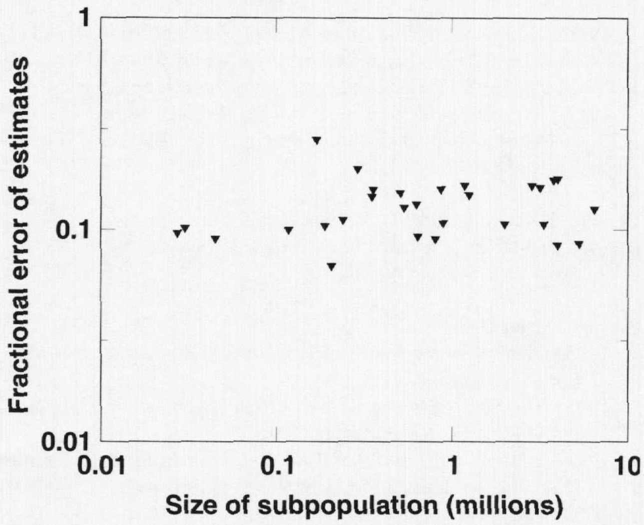
As with any survey data, it is often the case that respondents do not know, or refuse to answer, some questions. For example, depending on the background and characteristics of a given respondent, they may be unable to estimate the number of people they know in one population (such as people in their organization or people named Michael) while they can for others (such as people in their immediate family and people named Nicole). We estimated what effect this might have on our results in several ways. First, we compared estimates of c and back-estimates of subpopulation size using the scale-up method which treated missing data in two ways. The first method (used in all data reported in our previous papers), when adding an m value and an e value for some subpopulation in the denominator, ignored *both* contributions in (2) above if the data for that subpopulation were missing (i.e., that subpopulation was simply ignored for that respondent).[6] The second method continued to add the e value in the denominator, but made no change to the numerator. The resulting change was 1/3 percent for any respondent's c (though reaching as much as 30% for a very few respondents). Thus missing data had a negligible effect on the scale-up method.

The same was not the case for the summation method. Data were missing in one or more relation categories used in the calculation of the summation c for 35 percent of the 1,370 respondents to Surveys 1 and 2 (25% of the 1,370 were missing only one or two categories, the other 10% were missing more than two). The summation c results quoted above simply added the nonmissing values together (i.e., ignoring missing data). By replacing these data with averages over all respondents' responses for that category, we found an underreporting of 24, or 8 percent of the average c. Thus missing data and heaping effects may combine to produce underreporting in the summation method, and it is necessary to ensure that respondents answer as fully as possible when this method is used.

Respondents with Atypically Large Networks

Another way to test the validity of our method for estimating personal network size is to restrict our respondents to a subpopulation whose network size we expect to be very large or very small. As we have already pointed out, the literature on estimating personal network size is limited, and virtually nothing exists on network size of particular subpopulations. Thus we must rely on experience and common sense to suggest appropriate subpopulations for this test. Another requirement for the subpopulation is that there must be an available sample. Although we would like to, we cannot buy a sample of recluses or hermits. Subpopulations that

Figure 5. Mean Fractional Error of Estimated Reported Number Known in Subpopulations for Survey 1, as a Function of Known Subpopulation Size



Note: There is no significant correlation.

we think have large networks include politicians, labor organizers, diplomats, and clergy. We purchased a list of telephone numbers for a representative sample of members of the clergy nationwide and used this to conduct Survey 3 described above.

For these data, mean network size from the scale-up method was 598 (SD = 504) and for the summation method, 948 (SD = 1223). The difference between network size for clergy compared to the general population is significant (p = 0.0001). As we expected, the average network size produced by both methods were larger, by far, than those for the general population.

There appears to be a big difference between mean network size of clergy generated by the two methods. A t-test reveals that these two estimates are significantly different (p = 0.0001), whereas they were statistically the same in both Surveys 1 and 2. This is almost certainly caused by larger values for one or two of the relation categories used in calculating the summation network size, specifically those asking respondents how many people they know through religious and organizational affiliations. Note, however, that the estimate for network size from the scale-up method is also larger for clergy than for the general population, and this method has no subpopulations that are obviously biased toward eliciting large estimates from clergy. Note, also, that the fact that some classes of respondent can yield larger network sizes indicates that our consistent findings on nationally representative samples are not artifacts of our methodology, at least within the United States.

Range Data

So far we still have no knowledge of respondents' ability to report accurately the information we request. The focus groups, and our work on heaping, indicated that respondents handled the problem of providing us with an estimate in two ways, depending upon the size of the number they were estimating: when small, they enumerated, and when large, they estimated. Since we believed it would be useful to at least ask respondents how accurate they thought their answers were, we ran Survey 4.

The change made was that rather than being asked how many people were known in a subpopulation, respondents instead provided a range, consisting of a minimum and a maximum number, between which they were confident the answer lay. We suspected respondents' answers would echo our beliefs: small subpopulations would be reported relatively precisely—i.e., have low fractional ranges—while large subpopulations would be imprecise and possess high fractional ranges. To our surprise, this was not the case. Figure 5 shows the fractional range of reported numbers known in each subpopulation for the first survey as a function of the true subpopulation size. There is no significant variation in the fractional range. Thus, respondents did not think they were more or less precise (as a fraction of subpopulation size) for small or large subpopulations, and we can discount this as a possible source of error.

Discussion

We have described two methods for estimating the size of personal networks and shown that they yield very similar distributions. These results are consistent across independent replications. Further, we were unable to find any instrument effect that would explain these similar and consistent results. Thus far we must conclude that both methods yield estimates that, at worst, are proxies for personal network size as we have defined it.

So how might these methods be applied? First, they can be used to estimate personal network size by itself. To the extent that network size affects (or is related to) individual attributes and behaviors, these methods have immediate applications. Network size represents access to information and resources. For example, those with large personal networks may be more knowledgeable about options in pursuing health care. Those whose networks are loosely connected and consist mostly of acquaintances and coworkers may be quicker to access public assistance than those whose networks are densely connected sets of relatives and close friends (cf. the "strength of weak ties" argument of Granovetter 1973).

The second application is the estimation of hard-to-count populations (see Killworth et al., 1998b for a more complete discussion of this method). We can imagine many uses for such a method. There are many hidden populations in developed countries that are elusive, such as heroin users or the homeless. In developing countries many populations are hard to count because the infrastructure precludes conducting what would be a simple survey process in the United States or Europe. A method that allows researchers to estimate the size of hard-to-count populations quickly, cheaply, and with reasonable accuracy would be invaluable.

Our method for estimating the size of hard-to-count populations requires not only a proxy for network size, but an accurate estimate of network size. While we are confident that our estimates are valid proxies,[7] based on the analyses presented here, recent findings have called into question the accuracy of these estimates. Other parts of our analysis (not reported here) have raised questions about the cognitive process that a respondent goes through in making estimates of the number of people they know in various subpopulations and relation categories. Specifically, the mean number known in a given subpopulation varies approximately as the square root of the subpopulation size. This pattern violates assumption (1) which would predict a linear variation, and we are currently studying this. Until the effects of this power law relationship are understood, caution must be used in the application of the methods for estimating network size for estimating hard-to-count populations.

We conclude that both methods for estimating network size yield valid and reliable proxies for actual network size, but questions about accuracy remain.

Notes

[1]For the summation method, it is unimportant whether respondents agree on who is or is not within a relation category, as long as they sum to the total. For the scale-up approach, it is crucial that respondents all use the same definition of a subpopulation.

[2]One reviewer expressed concern over the low cooperation rates for the four surveys. The danger of low rates is that those who failed to respond may be different from those who did. Although this is a valid concern, we must keep two things in mind. Response and cooperation rates have declined sharply over the past two decades. The rates reported here are not unusual for a national study using random digit dialing, with no incentive and no introductory letter. Second, there appears to be no correlation between estimated network size and the number of attempts it took to get the completed survey. We conclude that more calls would not have changed our results.

[3]Since the order of questions remained constant (i.e., respondent categories were always asked last), this could have serious consequences.

[4]The coefficient of variation is the standard deviation divided by the mean. This produces a "normalized" standard deviation that can be compared across subpopulation estimates.

[5]A more accurate statistical approach would involve error variance ratios, which are unknown here.

[6]The mathematics leading to the maximum likelihood formula (2) show this is the correct approach.

[7]Respondents believed personal network size is another valid proxy, since it is generally monotonically increasing with probability of knowing someone in an event subpopulation (Bernard et al. 1989).

References Cited

Baker, Michael
1992 Digit Preference in CPS Unemployment Data. Economics Letters 39:117-121.

Bernard, H. Russell, Eugene C. Johnsen, Peter D. Killworth, and Scott Robinson
1989 Estimating the Size of an Average Personal Network and of an Event Subpopulation. *In* The Small World. Manfred Kochen, ed. Pp. 159-175. Norwood N.J.: Ablex Publishing Corporation
1991 Estimating the Size of an Average Personal Network and of an Event Subpopulation: Some Empirical Results. Social Science Research 20:109-121.

Bernard, H. Russell, Peter D. Killworth, and Chris McCarty
1982 INDEX: An Informant-Defined Experiment in Social Structures. Social Forces 61:99-133.

Brewer, Devon D.
1993 Patterns in the Recall of Persons in a Student Community. Social Networks 15:335-359.
1995 The Social Structural Basis of the Organization of Persons in Memory. Human Nature 6:379-403.
1995 Patterns in the Recall of Persons in a Department of a Formal Organization. Journal of Quantitative Anthropology 5:255-284.

Brewer, Devon D., and Bihchii L. Yang
1994 Patterns in the Recall of Persons in a Religious Community. Social Networks 16:347-379.

Fowler, Floyd Jackson, Jr.
1992 How Unclear Terms Affect Survey Data. Public Opinion Quarterly 56:218.

Freeman, Linton C., and Claire R. Thompson
1989 Estimating Acquaintanceship Volume. *In* The Small World. Manfred Kochen, ed. Pp. 147-158. Norwood N.J.: Ablex Publishing Corporation

Granovetter, Mark S.
1973 The Strength of Weak Ties. American Journal of Sociology 78:1360-1380.

Huttenlocher, Janellen, Larry V. Hedges, and Norman M. Bradburn
1990 Reports of Elapsed Time: Bounding and Rounding Processes in Estimation. Journal of Experimental Psychology: Learning, Memory and Cognition 16:196-213.

Johnsen, Eugene C., H. Russell Bernard, Peter D. Killworth, Gene A. Shelley, and Chris McCarty
1995 A Social Network Approach to Corroborating the Number of AIDS/HIV+ Victims in the U.S. Social Networks 17:167-187.

Killworth, Peter D., and H. Russell Bernard
1978 The Reverse Small-World Experiment. Social Networks 1:159-192.

Killworth, Peter D., Eugene C. Johnsen, H. Russell Bernard, Gene A. Shelley, and Chris McCarty
1990 Estimating the Size of Personal Networks. Social Networks 12:289-312.

Killworth, Peter D., Eugene C. Johnsen, Chris McCarty, Gene A. Shelley, and H. Russell Bernard

1998a A Social Network Approach to Estimating Seroprevalence in the United States. Social Networks 20:23-50.

1998b Estimation of Seroprevalence, Rape and Homelessness in the U.S. Using a Social Network Approach. Evaluation Review 22:289-308.

Myers, Robert J.

1940 Errors and Bias in the Reporting of Ages in Census Data. Transactions of the Actuarial Society of America 41: 395-415.

Roberts, John. M., and Devon D. Brewer

n.d. Measures and Tests of Heaping in Discrete Quantitative Distributions. Journal of Applied Statistics. In press.

Stockwell, Edward G.

1966 Patterns of Digit Preference and Avoidance in the Age Statistics of Some Recent National Censuses: A Test of the Turner Hypothesis. Eugenics Quarterly 13:205-208.

Stockwell, Edward. G., and Jerry. W. Wicks

1974 Age Heaping in Recent National Censuses. Social Biology 21:163-167.

Sudman, Seymour, Norman M. Bradburn, and Norbert Schwarz

1996 Thinking About Answers: The Application of Cognitive Processes to Survey Methodology. San Francisco: Jossey-Bass.

Turner, Stanley H.

1958 Patterns of Heaping in the Reporting of Numerical Data. Proceedings of the Social Statistics Section, American Statistical Association, Washington D.C.:248-251.

Wicks, Jerry W., and Edward G. Stockwell

1975 Age Heaping in Recent National Censuses: An Addendum. Social Biology 22:279-281.

Winick, Charles

1962 Preference for Individual Digits. Journal of General Psychology 67:271-281.

Zelnick, Melvin

1961 Age Heaping in the United States Census: 1880-1950. Milbank Memorial Fund Quarterly 39:540-573.

1964 Errors in the 1960 Census Enumeration of Native Whites. Journal of the American Statistical Association 59:437-459.