Comparing Two Methods for Estimating Network Size [*]

Christopher McCarty

Peter D. Killworth

H. Russell Bernard

Eugene Johnsen

Gene A. Shelley

Author statement:

Chris McCarty is Survey Director for the Bureau of Economic and Business Research at the University of Florida, Gainesville, FL, USA. Peter Killworth is an Individual Merit Scientist at Southampton Oceanography Centre, Southampton, UK. Russ Bernard is Professor of Anthropology at the University of Florida, Gainesville, FL, USA. Eugene Johnsen is Emeritus Professor of Mathematics at the University of California, Santa Barbara, CA, USA. Gene Shelley is at Georgia State University, Atlanta, GA, USA. 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 the 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.

Abstract

In this paper we compare two methods for estimating the size of personal networks using a nationally representative sample of the U.S. 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, co-workers). The results

[*] Christopher McCarty, Peter D. Killworth, H. Russell Bernard, Eugene Johnsen and Gene A. Shelley. "Comparing Two Methods for Estimating Network Size", Human Organization 60:28-39. (2000).

demonstrate a remarkable similarity between the average network size generated by both methods (approximately 291). Similar results were obtained with a separate national sample. Other attempts to corroborate our estimates, such as replication among a population we suspect have 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. Efforts to estimate the impact of a power law associated with the cognitive process are still underway. We conclude that both methods for estimating network size yield valid and reliable proxies for actual network size, but questions about accuracy remain. Each method has advantages and disadvantages that lead us to recommend the use of both methods concurrently whenever possible.

Introduction

Since 1986, we have been working on a method for estimating the distribution of 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 network size, including the reverse small world studies (Bernard et al. 1982, and Killworth and Bernard 1978), the telephone book studies (Freeman and Thompson 1989) and the current scale-up method studies (Killworth et al. 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 both 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 which suggest various possible confounding effects in our methods. These include: number preference by respondents; whether the consistency between methods and between surveys is an artefact; missing data; whether respondents chosen for their large network size modifies the results; whether data from other cultures differ significantly; whether respondents are less able to estimate accurately for large subpopulations; and a preliminary examination of barrier effects.

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 they know, overall. The primitive model is:

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

where m is the reported number of people known in a subpopulation, c is personal network size, t is the size of some larger population T (such as the U.S. population), and e is the size of a subpopulation E of T (rape victims, for example). If we wish to estimate e, we can do so from knowledge of m, c and t. If we wish to estimate c, we can do so if we know m, e and t. (In practice, we use maximum likelihood estimates based on multiple simultaneous measurements across many subpopulations and many 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 the people whom 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. An initial investigation of this effect is presented below. 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: a) the stigma or embarrassment associated with that membership, b) because it is not a common subject of conversation, or c) 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). Transmission effects are being investigated in a separate study. We call a violation of the third assumption the "estimation effect," also discussed below.

Another reason a respondent 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 out of their home in the past year, but they are 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 US

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 contribute to the accuracy of our estimates.

The extent of this error is almost certainly widely varying depending on the particular subpopulation. The range of definitions of "homeless" (cf., for example, the US 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 the data 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 that they are of a specific type, such as medical conditions or well-defined events. This may introduce a barrier effect of sorts 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 know falls into a particular relation type (such as someone they know through a hobby or organization).

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.

TABLE 1 HERE

This survey generated an average network size of 286 (Killworth et al. 1998b). We are confident that 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 in our model suggest a program of research, the goal of which is to minimize the barrier, transmission, and estimation effects and possibly correct for them. This would maximize the accuracy of our estimation of network size and improve our estimates of subpopulation sizes. We report elsewhere on our current efforts to minimize estimation effects (Killworth et al, in preparation). 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, co-workers, etc.). The list of 16 relation types is also shown in Table 1. (Note that the last relation 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 countable subpopulation among all the people they know.

2. This method is quicker to implement as we currently must collect 20-30 countable subpopulations 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, an American Indian, etc. (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 countable subpopulations as a way of checking the accuracy of our estimates of uncountable subpopulations.

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 relation categories since those categories are not countable. The scale-up method is based on countable subpopulations, which provides a way to test the statistical integrity of estimates of c.

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 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.

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 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 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. 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. Perhaps 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[1] 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 the similar estimates for all subpopulations and relation types as the interview progressed, we would expect the line to become more flat. 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.

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 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 which tend to co-vary, 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 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 – continued to raise suspicion about 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 can not determine which of the two methods produces the more reliable and more valid results. We decided that we needed more information about the actual process respondents go through in providing the 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 they say; 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. The 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 people how many gun dealers they know.

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

Number preference

Rounding to prototypical values is the phenomenon in which, when requested to supply a count or estimate in a category, people respond with preferred integers (Huttenlocher et al.,1990; Lichtenstein et al.,1978). Such commonly reported values might be 0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, ..., 100, ..., 200, etc., which show a preference for terminal digits 0, 5, 00, etc. (Myers, 1940; Turner, 1958; Winick, 1962; Stockwell, 1966; Stockwell & Wicks, 1974; Wicks & Stockwell, 1975; Baker, 1992) 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 (Stockwell and Wicks, 1974) and refer to this rounding phenomenon as heaping, with the values which exhibit the 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, etc.

Fig. 4 shows one ubiquitous example of number preference. The figure shows the total number of reports of 0, 1, 2, …, summed over all subgroups, using a logarithmic scale, here including zero for clarity (since the most common response is 0). Selected values which are large compared with those around them, are indicated on the diagram. It is clear that "round numbers" are preferred, especially for responses above 10, and other numbers (e.g., 7) are preferentially unused. Multiples of 5 occur almost invariably above 30.

Could number preference be modifying our results? On the face of it, the effects of number preference are hard to undo, since we only observe their results and not conditions before the numbers were modified by the preference. We have attempted to analyze the effects in two ways: (a) by modeling how the data might change, and how our results would change, by using plausible initial data for the summation data; and (b) by a variety of ad hoc methods to reverse number preference given the final scale-up data.

(a) Modeling how the data might change

There is strong evidence (given below) that the number preference in Fig. 4 has little effect on our scale-up estimates. 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 of the number of responses (not shown). It is monotonically nonincreasing, starting off like a negative exponential but slowly flattening out so that there is a nontrivial tail out to the highest value. It is an idealized distribution, similar to those encountered with responses for the relation categories. For these 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 Fig. 4, but with occasional high values due to heaping. The mean response for this distribution is 16.07, while the mean response for the 16 relation categories is 19.54.

Now, suppose that 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 a respondent has a "true" number of people known in the subpopulation (according to the idealized distribution) which 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. We assume that 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 that the heap values here are those less than or equal to 100 which occur for at least half of the 16 relation categories, and estimate the effect of heaping on average m. For the following 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% of the original 16.07, for an error of about ˜6.2%. The resulting estimated value of c would thus be low by a little over 6%.

Certain other combinations of heap values from the above list for the relation categories yield similar errors. This example also incorporates the effects of an artefact which 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 heap value are rounded downwards, producing an artificially low value of c for those relational categories where this occurred. In this example, the artefact is responsible for around two percent of the six percent error due to the overall heaping. Other scenarios for heaping are being investigated and will be reported in more detail later (Johnsen et al., in preparation).

(b) Ad hoc methods to reverse number preference 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% 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 estimates of subpopulation size changed by at most 5%, usually much smaller. Since this change was stronger than simply modifying round-numbered responses alone, we conclude that number preference has only a small effect on our results, which are thus robust to this form of error.

Nonetheless, when and where number preference takes place can have an effect. Consider (3) in Killworth et al., (1998b: 293) for each respondent's c; this is simply proportional to a sum of m's over a sum of e's. We could assume that informant responses (the m's) were unreliable for values above some cutoff value k. To model this in a simple fashion[2], we down-weight that component of both the numerator and denominator sums to 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% decrease. In other words, to reproduce our findings, the large responses (which correspond to informants with large c's and 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 artefact

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 sizes of 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 disprove 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%. 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% but in the opposite direction. Again, subpopulation estimates changed by a similar amount. We then made a smaller, but again systematic, change. Reported knowledge above 10 was replaced by 10. The average c dropped to 245, a change of 16%.

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) which produce a smaller change, at least in the mean c. Thus we can safely reject the suggestion that "any reasonable reports of knowledge of large subpopulations would yield similar answers with this method."

Missing data

As with any survey data, it is often the case that respondents cannot or refuse to answer some questions versus others. 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 organizations 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 to date), when adding an m value and an e value for some subpopulation in the denominator, ignored both contributions if the data for that subpopulation were missing. 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% 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% of the 1,370 respondents to surveys 1 and 2 (25% of these were missing only one or two categories) . By replacing these data with averages for that category, we found an under-reporting of 24, or 8% of the average c. Thus missing data and number preference effects may combine to produce under-reporting 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 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 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 can't buy a sample of recluses or hermits. Subpopulations that 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 towards 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 artefacts of our methodology, at least within the U.S. culture.

Replication in Mexico

We have not ruled out the possibility of an instrument effect, and clearly, we have no independent, objective counts of our respondents' personal network sizes. So far, however, the evidence is strong that both our methods are converging to some useful representation of personal network size. We next tested the possibility that consistent replication of our results across surveys and methods was an artefact of culture or language, by replicating our work in Mexico.

For Mexico we used Spanish translations of the 16 relation categories used in the US surveys for calculating the summation method c, and a collection of eight subpopulations for which we had counts (although the reliability of these counts is not certain). Future replications might benefit from a qualitative process, such as a focus group and/or pile sort, that would tell us what categories are more sensible in Mexican culture.

A significant complication in the Mexico replication is that telephone interviewing is not a viable option. Whereas telephone coverage by household in most areas of the U.S. is 95% or more, it is still not unusual for large areas of Mexico to be without telephone service. Mail surveys, which are a less expensive option than telephone, are difficult because most households in rural Mexico do not have direct postal service. This leaves in-person interviews as an option.

In-person interviews are, for obvious reasons, the most expensive type of interviewing. Although the cooperation rate and response rate tend to be high, the expense of sending interviewers to households usually requires a compromise on the coverage of the sample. For this study, funding forced us to limit the sample to three areas: 911 responses from Mexico City, 250 from urban Oaxaca and 250 from rural Oaxaca. Mexico City was chosen because it represents 20 percent of the population. We also wanted an urban comparison and a rural sample. Oaxaca was chosen because of contacts there that facilitated the logistics of selecting respondents. At present we will report on the Mexico City sample alone as the Oaxaca survey is as yet incomplete.

Our comparison of the results of the scale-up and summation methods between Mexico and the U.S. are mixed. The average summation c for the 911 Mexico City respondents is 223 (SD = 208), as compared to 286 using the summation c results for Surveys 1 and 2 combined. That is, network size in Mexico city is 78% of network size in the U.S. Killworth et al. (1990) found the estimated network size of respondents in Mexico City to be 570 vs 2025 for Orange County, California and 1700 for Jacksonville, Florida. These estimates were generated using a modification of the “phone book method” pioneered by Freeman and Thompson (1989). Thus, using this method the Mexico City estimates were about 28% of those of Orange County and about 34% of those of Jacksonville; much smaller than the estimates from the scale-up or the summation method. While all methods suggest that network size among respondents in Mexico is, on average, smaller than respondents in the US, the discrepancy suggests a methodological problem as well. Since there is no theoretical reason to believe that Mexican respondents networks are fully 70 percent lower than those in the US, we suspect that the phone book method may be prone to more error than the summation or the scale-up methods.

One explanation is that telephone coverage in Mexico City is far lower than the US. Some estimates are that less than 60 percent of households in Mexico City have telephone service, compared to around 95 percent for the US. It is not unlikely that there is a bias to this coverage such that those with certain names are not represented in the phone book. These problems are compounded by the likelihood that the Mexico City phone book is not updated as regularly as is typical of US phone books. Thus, the numbers used in calculating network size may not be representative of the actual counts for the names listed. If either or both of these problems exist, they could easily account for a significant underestimate of network size in Mexico City using the phone book method.

The scale-up method in Mexico produces somewhat smaller numbers, with a mean c of 151. The ambiguity effect described above almost certainly has dire consequences with the scale-up method here. Since we were able to find fewer countable subpopulations for Mexico than for the US, the contribution of estimates for each subpopulation in Mexico is much more than for the scale-up method in the US. For Mexico, one subpopulation that is subject to much ambiguity is Mexican Indians. Unlike the US, a large proportion of the population of Mexico is indigenous. Further, there has been significant interbreeding between the Spanish and the Indians in the more than 400-year history of that country. Given this interbreeding, it is difficult to say just exactly what constitutes an Indian in Mexico (official estimates differ by factors of 2.5 depending on source - and hence on definition). This no doubt contributes to inaccuracies in our estimates.

In addition, for most subpopulations, respondents were asked to estimate the number of people they know in Mexico, *not* the number of people they know in Mexico City. However, since the sample was restricted to residents of Mexico City, the argument could be made that we should use Mexico City counts. Using the eight subpopulations for which there are counts, the scale-up c is 800 when limited to Mexico City as compared with 151 when using the country as a whole.

The difference between these two estimates is clearly due to the e's we use. Arguments can be made for both. Only the calculation using the whole country subpopulations is close to the U.S. Unfortunately, the subpopulations used in the Mexico study are fewer and of different types, mostly, than those used in the U.S., again because of the availability of counts. One type used in both is the number of diabetics known. The mean number of diabetics known differs significantly ($p = 0.0001$ by t-test) between the U.S. (with a mean of 3.4) and Mexico (with a mean of 2.8), lending support to the conclusion that network size in Mexico is smaller than the U.S. Indeed, the mean number of diabetics known to Mexicans is 82% of the number known to Americans on average. This is close to the comparison of summation method network size between the two countries.

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 number preference, 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. Arguing that 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 a minimum and a maximum number between which they were confident the answer lay. We suspected that respondents' answers would echo our beliefs: small subpopulations would be reported relatively accurately – i.e., have low fractional ranges – while large subpopulations would be inaccurate, and possess high fractional ranges. To our surprise, this was not the case. Fig. 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 accurate (as a fraction of subpopulation size) for small or large subpopulations, and we can discount this as a possible source of error.

Barrier effects

Finally, we return to the first potential source of error in our assumptions: barrier effects. There are many possible barriers to exposure to certain subpopulations; an obvious one being geographical, and others needing elucidation. Provided that a survey is representative, most if not all barrier effects should cancel out. We have begun an investigation of barrier effects by examining how respondents' reports of the number of people they know in subpopulations varies with properties of the respondents. An example is shown in Fig. 6a, which plots the mean number of Native Americans known to respondents in each of the 50 U.S. states. Figure 6b shows the fractional number of Native Americans in each state. The two diagrams are extremely similar ($r = 0.58$, $p = 0.0001$). Thus respondents who live in a state which contains a high proportion of Native Americans report knowing more of them. This is in direct contradiction to assumption 1.

However, the difficulty this could present is mitigated if a survey is truly representative across the country, or if many (and different) subpopulations are used. Surveys confined to small geographical areas may be subject to geographical barrier effects (for example, consider the survey limited to Florida in Killworth et al. 1998a, which predicted a much higher number for the HIV positive subpopulation than the national survey reported here).

TABLE 2 HERE

It may be more difficult to remove barrier effects if they are less easy to predict. Table 2 shows the results of analyses of variance across a combined national dataset across almost 3,000 respondents (produced by combining the datasets from all three replications of the scale-up method). The class variables used in the anovas were sociodemographic variables on the respondents. Most of the analyses showed significant differences in the number known in a subpopulation due to the variable analysed. This holds true even for apparently innocuous subpopulations such as "people named Michael."

In other words, the results of our methods are subject to a variety of barrier effects; again, we assume that by using a representative sample the effects can be overcome. A further paper will address this issue more fully.

Discussion

In this paper 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.

However, our analysis has 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. Although we have not as yet discovered the mechanism, further analysis shows that the process of estimating the number of people known in subpopulations and relation categories is apparently subject to a power law: the mean number known in a given subpopulation (i.e. those in Table 1) varies as the square root of the subpopulation size. This is in direct contradiction to (1) which would predict a linear variation. A companion paper (Killworth et al., in preparation) discusses this potentially serious difficulty more fully.

One possible consequence of a power law would be that our estimates of personal network size may not be accurate, even though they are proxies. This may be critical for our procedure to estimate hard-to-count populations because that procedure assumes that the resulting estimates are reasonably accurate. At this stage of our research we simply do not know how accurate these estimates are.

However, we are confident that our estimates can be used in other research where an investigator wishes to use network size as an explanatory or dependent variable. Although we cannot say that a respondent with an estimated c of 400 has a personal network twice as large as one with an estimate of 200, we can say that their network is larger. This suggests a new program of research into the predictive value of network size, and the attributes of respondents that explain how it varies.

References

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. pp. 159-175 in: The Small World, ed. M Kochen, Ablex, Norwood N.J., 382pp.

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.

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, ed. M Kochen, Ablex, Norwood N.J., pp. 147-158.

Killworth, Peter D. and H. Russell Bernard

1978 The reverse small-world experiment. Social Networks, 1, 159-192.

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

1990 Estimating the size of personal networks. Social Networks, 12, 289-312.

Killworth, Peter D., Eugene C. Johnsen, H. Russell Bernard, G.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.

Killworth, Peter D., Chris McCarty, H. Russell Bernard, Gene A. Shelley, and Eugene C. Johnsen 1998b. Estimation of seroprevalence, rape and homelessness in the U.S. using a social network approach. Evaluation Review, 22, 289-308.

Captions

1. Frequency distribution of network size (binned by 60s) for surveys 1 and 2 (circles and squares respectively). Closed symbols show size computed from the scale-up method; open symbols from the summation method.

2 Scatterplot of network size estimates from survey 1, showing scale-up estimates against summation method estimates.

3. Coefficient of variation of average reported number known subpopulation and relation type estimates in interview order.

4. Histogram showing number of occurrences of reported number known in any subpopulation across all informants. The y-axis is scaled logarithmically, but includes zero.

5. Mean fractional error of estimated (i.e., range / mean) reported number known in subpopulations for the first survey, as a function of the subpopulation size. There is no significant correlation.

6. (a) Mean number of Native Americans reported as known by respondents, by the state of residence of the respondents. (b) Fractional population of Native Americans for each state in the U.S. The two datasets are well correlated (r = 0.58).

Table 1. Average number of people known for subpopulations and relation types (combined surveys with a total of 1370 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	Co-workers	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 know 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		
Women who adopted a child in past 12 months	0.3	People who provide a service	7.7		
Widow(er) under 65 years old	3.2	Other	3.9		
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				

Table 2. Significant differences between levels of class for estimates of a given population (significance cut-off = 0.05).

Population	State	Sex	Race	Age	Education	Marital status	Work status	Religion	Political Party
Michael		•		•	•	•	•	•	•
Christina		•		•	•	•	•	•	•
Christopher		•		•	•	•	•		•
Jacqueline			•	•	•		•	•	
James		•	•		•	•	•	•	•
Jennifer			•	•	•	•	•	•	•
Anthony	•	•	•		•		•	•	•
Kimberly	•		•	•	•	•	•	•	•
Robert		•		•	•	•	•	•	•
Stephanie				•	•	•	•	•	•
David		•		•	•	•	•	•	•
Nicole				•	•	•	•		
Native Americans	•		•			•			
Gave birth in past 12 months			•	•	•	•	•	•	
Adopted a child in past year					•			•	
Widow(er) under 65 years				•		•	•	•	•
On kidney dialysis								•	
Postal worker		•	•					•	
Commercial pilot		•		•	•				•
Member of Jaycees	•	•			•	•	•	•	•
Diabetic					•	•	•	•	
Opened a business in year				•	•	•	•		
Have a twin brother or sister	•	•	•	•	•		•	•	•
Licensed gun dealer		•							
Came down with AIDS				•	•	•	•		
Males in prison		•	•	•			•		
Homicide victim in past year	•		•		•	•			
Suicide in past year			•		•				
Died in wreck in past year	•			•			•	•	
Women raped in past year				•	•				
Homeless			•			•			
HIV positive				•	•	•	•		

Figures

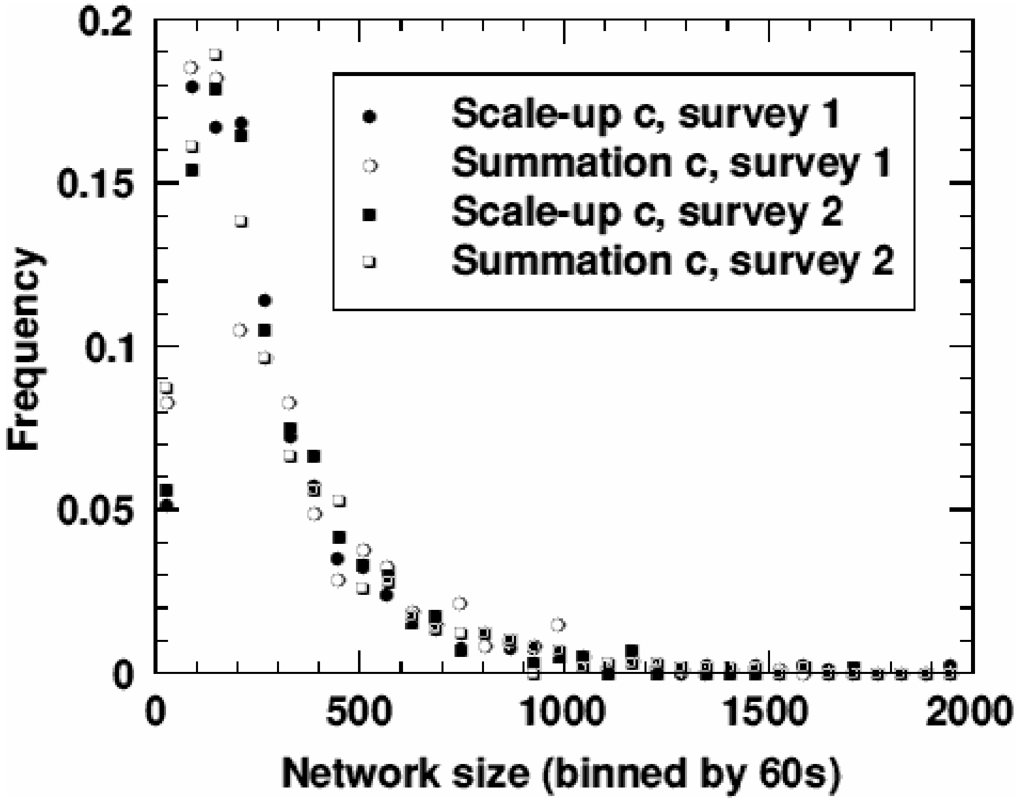

Fig. 1. Frequency distribution of network size (binned by 60s) for surveys 1 and 2 (circles and squares respectively). Closed symbols show size computed from the scale-up method; open symbols from the summation method.

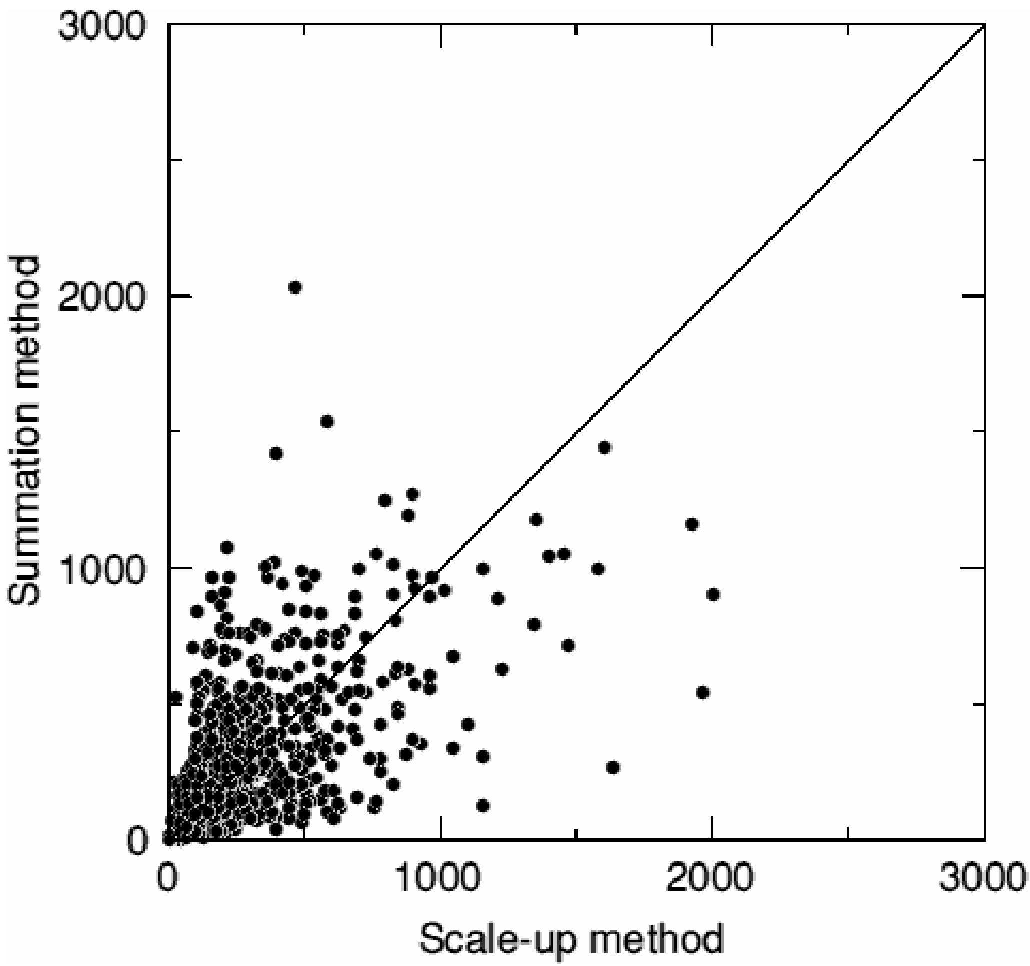

Fig. 2. Scatterplot of network size estimates from survey 1, showing scale-up estimates against summation method estimates.

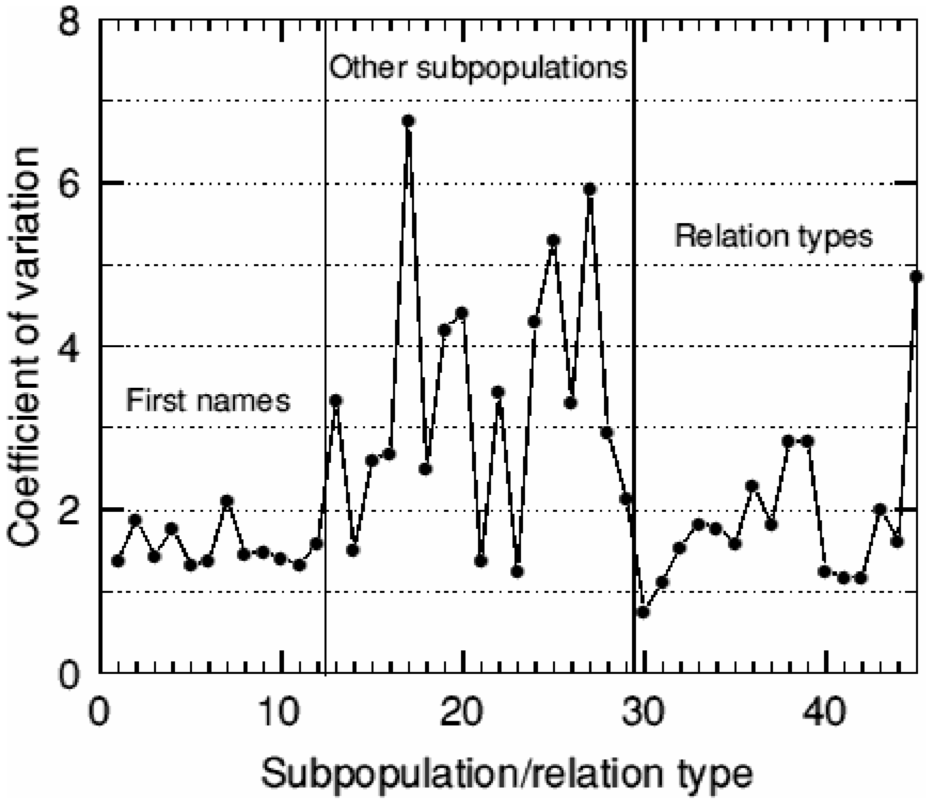

Fig. 3. Coefficient of variation of the average reported number known in subpopulations and relation types, in interview order.

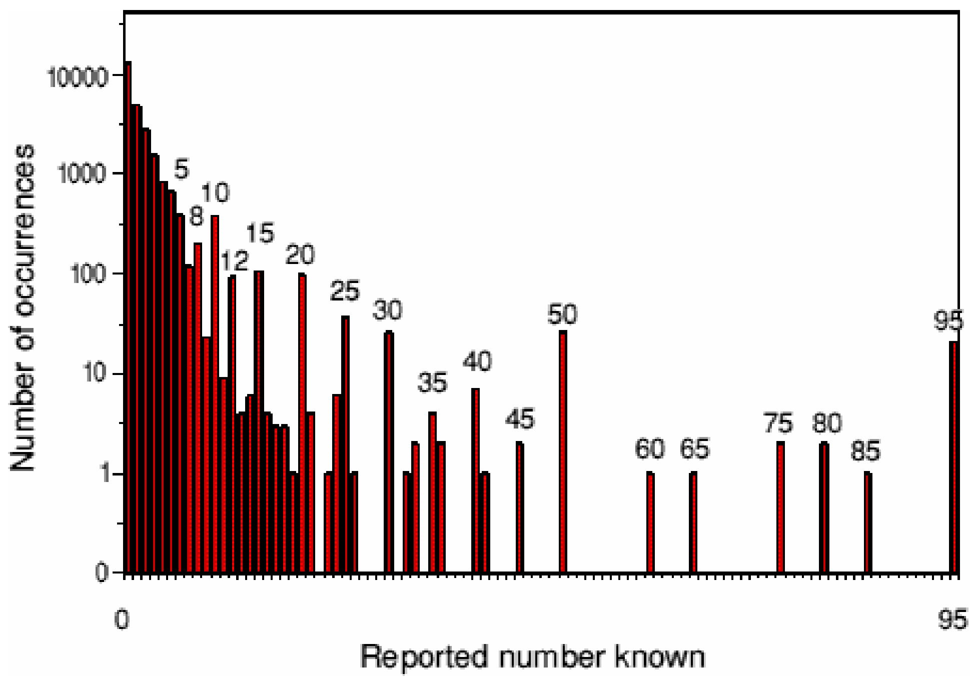

Fig. 4. Histogram showing number of occurrences of reported number known in any subpopulation or relation type across all informants. The y-axis is scaled logarithmically, but includes zero.

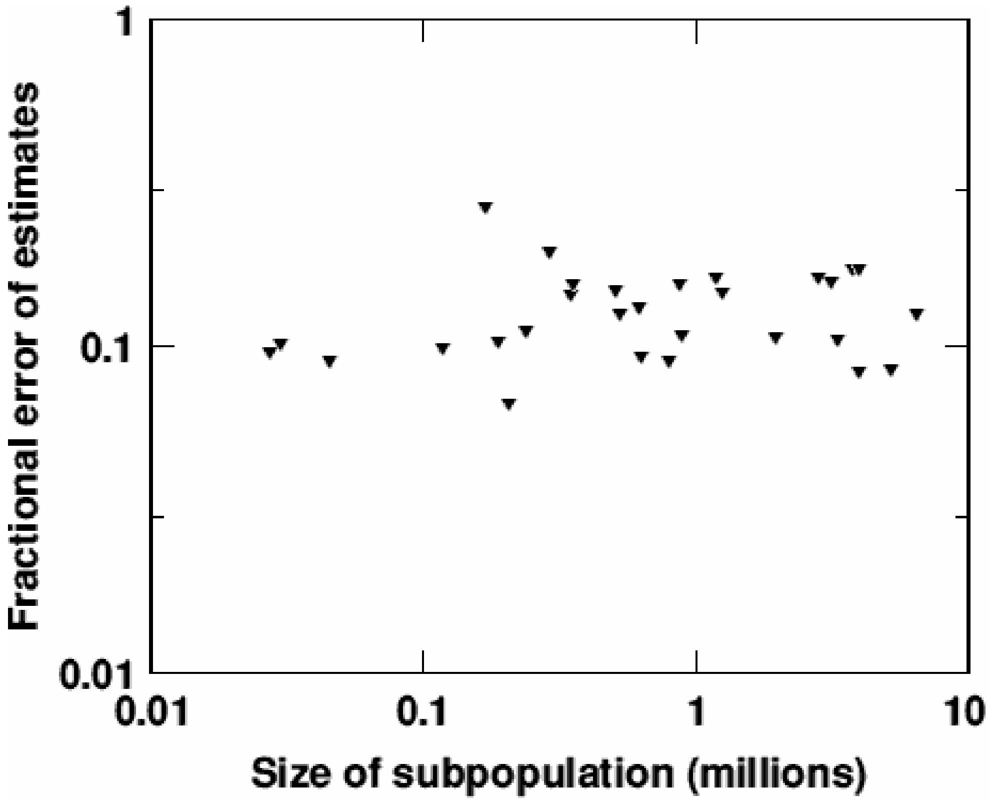

Fig. 5. Mean fractional error of estimated (i.e., range / mean) reported number known in subpopulations for the first survey, as a function of the subpopulation size. There is no significant correlation.

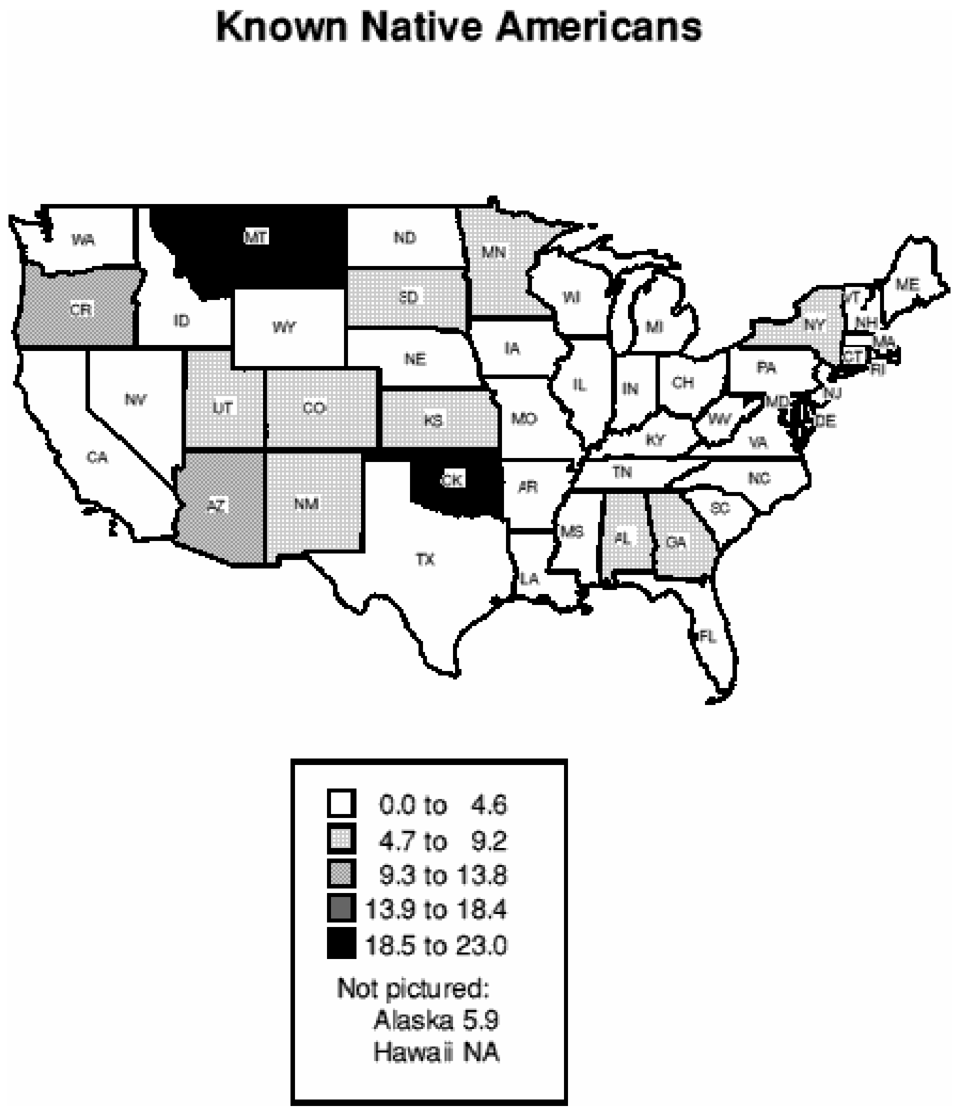

Fig. 6(a). Mean number of Native Americans reported as known by respondents, by the state of residence of the respondents.

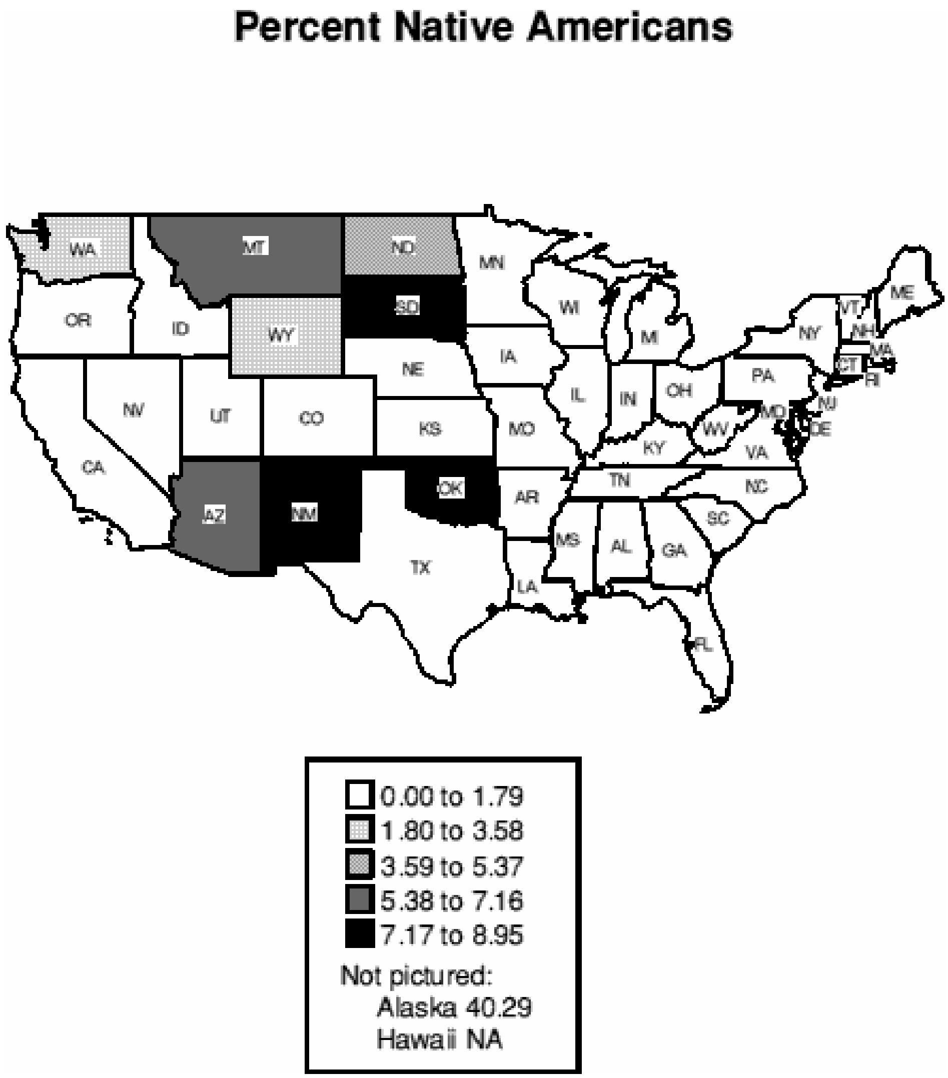

Fig. 6(b). Fractional population of Native Americans for each state in the U.S.

Notes

[1]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.

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