SOCIAL SCIENCE RESEARCH 20, 109–121 (1991)

Estimating the Size of an Average Personal Network and of an Event Subpopulation: Some Empirical Results

H. RUSSELL BERNARD

University of Florida, Gainesville

EUGENE C. JOHNSEN

University of California, Santa Barbara

PETER D. KILLWORTH

Hooke Institute for Atmospheric Research, Oxford, England

AND

SCOTT ROBINSON

Universidad Autònoma Metropolitana, Mexico City

A compelling problem in a population is to estimate the number of people the average person knows. A consequential related problem is to estimate the size of important subpopulations. A random sample of a population is asked whether they know anyone in a given subpopulation of size e, thus yielding an estimate of the probability that this occurs in the population. Using an equal likelihood probability model, this leads to a lower bound estimate for c, the average number of people a person in the population knows. When the number of people a person knows has a binomial distribution over the population this value is an estimate for c itself. Here we test this method on data from Mexico City, where a large

Presented by invitation at the Annual Meeting of the American Statistical Association, San Francisco, CA, August 17–20, 1987, under the sponsorship of the Section on Survey Research Methods. Supported in part by NSF Grant BNS-8318132 and by a Faculty Research Grant from the Graduate School, University of Florida. We particularly thank Dr. Donald Price, Vice President for Research, University of Florida, for generous support of this research, and student researchers Maria del Carmen Costa, Alejandro Casteneira, Yolanda Hernandez Franco, Patricio Meade, and Miguel Angel Riva-Palacio for their conscientious efforts in the collection of the data analyzed in this paper. Address all correspondence and reprint requests to Eugene C. Johnsen at the Department of Mathematics, University of California, Santa Barbara, CA 93106.

0049-089X/91 $3.00

random sample of people was asked whether they knew anyone in each of several different subpopulations in Mexico City of known and unknown sizes. We develop procedures for obtaining various bounds and estimates for c and determine some of the respondents' attributes on which variation in probability of knowing someone in a subpopulation and variation in personal network size seem to depend. We apply these to the estimation of e for rape victims in Mexico City and the estimation of c from data on AIDS victims in the United States.

An important but vexing problem in social network analysis has been to determine in a population the number of people a person knows, i.e., his/her personal network size, and the mean, range, and distribution of this variable in the population as a whole (cf. de Sola Pool and Kochen, 1978). These data have heretofore defied successful investigation. In an earlier paper (Bernard, Johnsen, Killworth, and Robinson, 1989), we presented a probabilistic method for estimating the average size of a personal network and the size of an event subpopulation in a given total population. We applied it to a first small random data sample of size 400 from the population of the Federal District of Mexico City in order to relate various proposed sizes of the subpopulation of victims of the 1985 Mexico City earthquake to the average personal network size in the Federal District. We give here and in the next two sections a brief summary of this method and these first results. We then apply the method to the data from a second larger random sample from the population of Mexico City proper in order to obtain estimates of the size of the average personal network from various known event subpopulation sizes, and then use this information to estimate the size of an unknown event subpopulation and to compare against results for a known event subpopulation in the U.S.

Consider a population T, of size t, having a subpopulation E, of size $e > 0$, $e << t$, which is the subgroup of T associated with some attribute or event. For each member u of $T-E$ let $k(u)$ denote the number of people in T that u "knows." Here "u knows v" means that u knows v personally, in that u knows v by name, knows where v lives, knows v's occupation, and that v knows the same about u. The people in T whom u knows will be called the *personal network* of u, denoted by $K(u)$.

We allow $k(u)$ to vary with u over $T-E$ and to take its values on a finite interval of nonnegative integers $[n_0, n_0 + n]$, where $n \geq 0$. Regarding average personal network size, we give some results on the general case and the case where $k(u)$ has a binomial distribution. We then address the question of estimating event subpopulation size.

Now, we need to make a fundamental assumption, either about the distribution of the members in the various personal networks $K(u)$ or about the distribution of the members of E, as follows:

A. For a random member u of $T-E$, all subsets of $T-\{u\}$ of size $k(u)$ are equally likely to have been the subset $K(u)$ known by u.

B. All subsets of T of size e were equally likely to have been the subpopulation E.

In some situations (but possibly not some of those discussed in this paper, e.g., the Mexico City earthquake) version B seems plausible. In the case of the earthquake, if all of the downtown buildings in a city were similar in level of earthquake survivability and all socioeconomic strata of the population were randomly represented in the downtown population when an earthquake occurred centered downtown, etc., then this assumption may not be a bad one. Version A is equivalent to assuming that for a random u in $T-E$ the probability that any particular member of $K(u)$ is in E is just the relative size of E in T, e/t, when e is very small compared to t.

AVERAGE PERSONAL NETWORK SIZE

For the general case, where the distribution of $k(u)$ for u in $T-E$ is unspecified, we have the following results (Bernard et al., 1989). We let p denote the proportion of the members of $T-E$ who know someone in E, ln is the natural logarithm, g is a certain real number in the range $1 \leq g \leq \max\{k(u) \mid u \in T-E\}$, and $\varepsilon = e/t$.

LEMMA 1. *Under either of the assumptions* A *or* B, *the value*

$$\alpha \equiv \ln(1-p)/\ln[1 - e/(t-g)] \approx \ln(1-p)/\ln(1-\varepsilon), \qquad (1)$$

determined by the values of e, p, and t and the distribution of k(u), must lie within the range of values $[n_0, n_0 + n]$. *The right hand numerical approximation in (1) is excellent when* $n_0 + n$ *is very small compared to t.*

THEOREM 2. *Under either assumption* A *or* B *and for any probability distribution of the values k(u) on the integer interval* $[n_0, n_0 + n]$, *the value* α *and the average value c of the personal network sizes k(u) must satisfy the inequalities*

$$n_0 \leq \alpha \leq c \leq n_0 + n. \qquad (2)$$

For the one-point distribution, where $n = 0$, *all three inequalities in (2) are equalities. For a distribution with at least two points, where* $n > 0$, *all three inequalities are strict inequalities* $<$.

With an empirical estimate for p and under either of the distribution assumptions A or B we can estimate α, and frequently c, by the right side of (1). For, if $\alpha_1, \alpha_2, \ldots, \alpha_s$ are values corresponding to different event subpopulations $E_1, E_2, \ldots, E_s$ we should have

$$c \geq \max_{1 \leq i \leq s} \alpha_i. \qquad (3)$$

TABLE 1
Values of the Lower Bound Estimate α of Average Personal Network Size c for the Mexico City Earthquake Data with Varying Death Rates e (t = 18,000,000)

e:	7000	12,000	15,000	22,000
α:	664	387	310	211

Our first Mexico City earthquake sample of 400 random respondents did not meet our statistical requirements; however, it was tantalizing to use the data from this sample to find α as a lower bound estimate for c. With $p \approx 91/400 = .2275$, we computed α for the different proposed death counts e to the nearest integer in Table 1 (taking $g = 0$). The right hand approximation in (1) is correct to within .1% here, assuming $n_0 + n \leq 10000$.

We now consider the special case where $k(u)$ is assumed to have a binomial distribution over its range $[n_0, n_0 + n]$. Here n may be viewed as the number of opportunities or encounters (the "trials" of the binomial distribution) that u has with other members of T, over and above a fixed set of n_0 members whom u already knows, each of which has a fixed probability of resulting in u knowing v (i.e., resulting in "success"). For this case we have

$$c \approx \alpha \tag{4}$$

Thus, the values of c for the Mexico City earthquake data when $k(u)$ has a binomial distribution over its range are virtually the same as those given in Table 1, and the approximate equality (4) is practically independent of the value of the fixed trial success probability.

EVENT SUBPOPULATION SIZE

For the probability distribution of $k(u)$ with $\hat{p} \equiv 1 - \mathrm{p}$, $\hat{\varepsilon} \equiv 1 - \varepsilon$, and $q_m = P(k(u) = n_0 + m)$, $m = 0, 1, \ldots, n$, we obtained (Bernard et al., 1989) that

$$\hat{p} = \sum_{m=0}^{n} q_m \hat{\varepsilon}^{n_0+m}. \tag{5}$$

Since $\hat{\varepsilon} > 0$ and $q_m \geq 0$ for all $m = 0, 1, \ldots, n$ with some $q_m > 0$, the derivative of $\hat{p}$ with respect to $\hat{\varepsilon}$ is positive, whence $\hat{p}$ is an increasing function of $\hat{\varepsilon}$ and so p is an increasing function of ε. Thus, if there are event subpopulations $E_1, E_2, \ldots, E_s$, ordered so their corresponding ε_i values satisfy

$$\varepsilon_0 = 0 < \varepsilon_1 < \varepsilon_2 < \cdots < \varepsilon_s, \tag{6}$$

then we also have for their corresponding p_i values

$$p_0 = 0 < p_1 < p_2 < \cdots < p_s, \tag{7}$$

where $\hat{\varepsilon}_0 = 1 - \varepsilon_0 = 1$ and $\hat{p}_0 = 1 - p_0 = 1$ also satisfy (5).

Now let E_x be a new event subpopulation of unknown size e_x and unknown relative size ε_x for which the probability p_x that a random u in T–E_x knows anyone in E_x satisfies in (7)

$$p_{k-1} < p_x < p_k, \quad \text{for some } k,\ 1 \leqslant k \leqslant s. \tag{8}$$

Then, from (6) we have

$$\varepsilon_{k-1} < \varepsilon_x < \varepsilon_k. \tag{9}$$

Thus, if our probability model is reasonably close to correct then, given a sufficiently broad range of ε_i, p_i pairs from previous event subpopulations, we should be able to bound the size of the new event subpopulation between successive values

$$e_{k-1} < e_x < e_k, \qquad \text{for some } k,\ 1 \leqslant k \leqslant s. \tag{10}$$

Clearly, if (8) is true but not (9) and (10) then either the data values e_i, p_i, $1 \leqslant i \leqslant s$, are poor or the original probability model is not completely correct. Thus, if the data are believed to be good we have a negative criterion for the full validity of the underlying probability model.

Assuming in this model that $\hat{p}$ is a differentiable function of $\hat{\varepsilon}$, we have from (5) that

$$\begin{aligned} d\hat{p}/d\hat{\varepsilon}\,|_{\hat{\varepsilon}=1} &= \sum_{m=0}^{n} (n_0 + m)\, q_m\, \hat{\varepsilon}^{n_0+m}\,|_{\hat{\varepsilon}=1} \\ &= \sum_{m=0}^{n} (n_0 + m)\, q_m = c. \end{aligned} \tag{11}$$

Thus, for a fairly large value for c (at least 211 by Table 1) and for $\hat{\varepsilon}$ less than but very close to 1, we see that large changes in $\hat{p}$ correspond to small changes in $\hat{\varepsilon}$. This indicates that whatever the size of the bound within which p_x sits in (8), the corresponding size of the bound for the approximation of ε_x in (9) will be considerably smaller.

Now suppose that for a fixed population T(more precisely, T–E for which e is very small relative to t) we know the average personal network size c. From (1) we derive the relation

$$1 - p = (1 - \varepsilon)^{\alpha}, \tag{12}$$

where α is a function of ε and p and, hence, need not be constant over different pairs ε, p. Now, by (2), we have for $0 < \varepsilon < 1$ that

$$(1 - \varepsilon)^{\alpha} \geqslant (1 - \varepsilon)^{c} \tag{13}$$

TABLE 2
Known Sizes of, Probabilities of Knowing Someone in, and Corresponding Values of α for Six Reference Subpopulations of Mexico City Proper (t = 10,700,000)

Subpopulation	c	Range of p	α	Range of α
Doctors	30,426	.3889 ± .0201	173	162–185
Mailmen	14,728	.1473 ± .0146	116	103–128
Bus Drivers	11,696	.2571 ± .0180	272	250–294
Quake Victims	10,000	.2668 ± .0182	332	306–359
TV Repairmen	4,013	.2619 ± .0181	810	745–876
Priests	1,595	.2854 ± .0186	2254	2082–2431

or, by (12),

$$\varepsilon \geqslant 1 - (1 - p)^{1/c} \equiv \tilde{\varepsilon}. \tag{14}$$

Thus, for E_x of unknown relative size ε_x, with an accurately estimated probability p_x of a person in T–E_x knowing someone in E_x, we have

$$\tilde{\varepsilon}_x = 1 - (1 - p_x)^{1/c} \leqslant \varepsilon_x, \tag{15}$$

which yields a lower bound approximation $\tilde{\varepsilon}_x$ to the true value ε_x. Here, the closer c is to α, or ε_x is to 0, the better the approximation $\tilde{\varepsilon}_x$ is to ε_x. The latter implication corresponds to the fact that the closer $\hat{\varepsilon}$ is to 1 the better the approximation of $\hat{\varepsilon}^\alpha$ by $\hat{\varepsilon}^c$.

As an example, from the first sample data for the Mexico City earthquake with event subpopulation E of assumed size e = 7000 and probability p = .2275, suppose we determined that $c \approx 664$. Now suppose for a new event subpopulation E_x we obtain p_x = .1986. Then the size e_x is bounded by $e_x < 7000$ and underestimated by $\tilde{e}_x = \tilde{\varepsilon}_x \cdot t = [1 - (.8014)^{1/664}] \cdot (18{,}000{,}000) = 6000.67$, so $6001 \leqslant e_x < 7000$. Now, from (5), the graph of $\hat{p}$ as a function of $\hat{\varepsilon}$ is increasing and concave upward; hence, we can do a linear interpolation between the points $(\hat{\varepsilon}_1, \hat{p}_1) = (1, 1)$ and $(\hat{\varepsilon}_2, \hat{p}_2) = (.99961111 \cdots, .7725)$ to obtain the tighter upper bound $e_x \leqslant 6110$, whence $6001 \leqslant e_x \leqslant 6110$.

APPLICATION TO A LARGE DATA SAMPLE FROM MEXICO CITY

In a later second survey we obtained data from a larger random sample of 2260 from Mexico City proper (t = 10,700,000) in the hope of establishing a set of reference value pairs (p_i, ε_i) against which to compare new value pairs (p_x, ε_x) according to (8) and (9) above. The data for the six reference event subpopulations, with the 95% confidence ranges for p and corresponding ranges for α, are given in Table 2.

It is clear from this table that the monotonicity property of the model given by (8) and (9) does not hold, which indicates that either the data are inaccurate, the data are reasonably accurate but not very precise, or

the model is not valid in its simple form for different subpopulations (or possibly a combination of either the first or second and the third).

By the nature of the survey and the data obtained, the first alternative (including its combination with the third) appears untenable. But, before ruling out the model in its simple form, we investigate the data to see whether they are reasonably accurate but not very precise. This suggests analyzing the data at the aggregate level to gain precision and, it is hoped, detect a "signal" amidst the "noise." Since the simple model (1) with the assumption of an approximately binomial distribution of $k(u)$ for u in T (or T–E) produces an approximately constant $c \approx \alpha = \ln \hat{p}/\ln \hat{\varepsilon}$, which says that $\ln \hat{p}$ and $\ln \hat{\varepsilon}$ vary linearly with respect to each other, we attempt to estimate c (via α) by least squares linear regression of each variable against the other. Now, there are four ways to do this, namely,

$$\ln \hat{p} = \alpha \cdot \ln \hat{\varepsilon}, \tag{16}$$

$$\ln \hat{p} = \alpha \cdot \ln \hat{\varepsilon} + \ln \beta, \tag{17}$$

$$\ln \hat{\varepsilon} = \alpha^{-1} \cdot \ln \hat{p} \tag{18}$$

$$\ln \hat{\varepsilon} = \alpha^{-1} \cdot \ln \hat{p} + \ln \gamma, \tag{19}$$

where $\ln \beta$ and $\ln \gamma$ are included in the unconstrained regressions. The unconstrained regressions are included to see how well their regression lines fall naturally into place on the basis of the empirical data alone without imposition of the intercept 0. We obtain the following results, with α to the nearest integer, for the data in Table 2:

$$\alpha = 196, \tag{16.r}$$

$$\alpha = 56, \qquad \beta = .7761, \tag{17.r}$$

$$\alpha = 274, \tag{18.r}$$

$$\alpha = 221, \qquad \gamma = 1.0003. \tag{19.r}$$

The values of α in (16.r) and (18.r) for the constrained model yield the range 235 ± 39, which includes the α values for Bus Drivers and for line (19). The data point for Bus Drivers almost lies on all three lines (17), (18), and (19). Except for the data point for Quake Victims, the other data points do not lie very close to any of these lines. Since $\ln .7761 = -.253$ and $\ln 1.0003 = .0003$, line (17) suggests some problem with the data or the model, whereas line (19) does not. We may also determine the best fit lines (16) and (17) in the sense of least squares distance of the data points from the lines (lines (18) and (19) are then, respectively, equivalent to (16) and (17)). For these we obtain the following results from the data in Table 2:

TABLE 3
Probabilities of Knowing Someone in the Event Subpopulation According to the Partition of the Sample by Age in Years

Subpopulation	<20	20–34	35–49	50–65	>65
Doctors	.3171	.3756	.4484	.4921	.4000
Mailmen	.1111	.1503	.1659	.1746	.1143
Bus Drivers	.2685	.2599	.2646	.2063	.2000
Quake Victims	.2824	.2789	.2399	.2275	.2286
TV Repairmen	.2847	.2798	.2399	.1852	.0857
Priests	.2269	.2720	.3363	.3598	.4000

$$\alpha = 274, \tag{16.s}$$

$$\alpha = 221, \qquad \beta = .9356. \tag{17.s}$$

Since ln .9356 = −.0666 and these best fit lines (which treat the variables symmetrically) are virtually the same as (18.r) and (19.r), respectively, this again supports the corresponding α values 248 ± 27. Both (17.s) and (19.r) suggest that lack of precision could be the culprit in the scatter of the data. With an approximately binomial distribution for $k(u)$ and data of sufficiently high precision, the simple model may be adequate for estimating e from c and vice versa.

For each of the six reference subpopulations the sample (and hence also the total population) was partitioned into subclasses according to (i) zone of survey interview, whether socioeconomically lower, middle, or upper class; (ii) age of respondent, whether <20, 20–34, 35–49, 50–65, or >65 years; (iii) highest education level attained by respondent, whether <4, 4–6, 7–12, 13–16, or >16 years; (iv) socioeconomic class of respondent, whether upper, middle, or lower class; (v) occupation of respondent, whether working at home (home), out of home (oohm), retired (retd), unemployed (unem), students (stud), or a residual class (rsdl); and (vi) respondent's reporting of how many people he/she believes to be in

TABLE 4
Probabilities of Knowing Someone in the Event Subpopulation According to the Partition of the Sample by Education in Years

Subpopulation	<4	4–6	7–12	13–16	>16
Doctors	.2539	.2769	.3735	.5146	.7333
Mailmen	.0933	.1076	.1549	.1683	.2519
Bus Drivers	.3161	.2749	.2833	.1756	.1556
Quake Victims	.1554	.2032	.2853	.3244	.3481
TV Repairmen	.1399	.1096	.2814	.3976	.4444
Priests	.2176	.2470	.2892	.3146	.4074

TABLE 5
Probabilities of Knowing Someone in the Event Subpopulation According to the Partition of the Sample by Socioeconomic Class

Subpopulation	Lower	Middle	Upper
Doctors	.2696	.4644	.6323
Mailmen	.1239	.1743	.1097
Bus Drivers	.3211	.2245	.0710
Quake Victims	.2200	.3066	.2903
TV Repairmen	.1606	.3203	.5097
Priests	.2250	.3285	.3742

his/her personal network, whether ≤100, 100–500, 500–1000, 1000–1500, >1500, or no answer. We present the p values for the subsamples determined by age, education, socioeconomic class, occupation, and number believed known in Tables 3, 4, 5, 6, and 7. Since the full sample was not a quota sample for the different subclasses of T the subsample sizes varied considerably, from 16 in How Many: No Answer to 1158 in Age: 20–34.

We note that there is a great deal of monotonicity in this data, either increasing or decreasing values of p with increasing values of the table variable when the table variable has a natural ordering. For example, the data for Priests show that p increases without exception with increasing age, education, and socioeconomic class and almost without exception with increasing believed personal network size. For Doctors, p increases without exception with increasing education, socioeconomic class, and believed personal network size and almost without exception with increasing age. A similar but weaker version of this occurs for Mailmen.

With regard to all six reference subpopulations, p increases with increasing believed personal network size without exception for Doctors and Quake Victims and almost without exception for Mailmen, Bus Drivers, TV Repairmen, and Priests. Thus, believed personal network size behaves like actual personal network size should behave with respect to

TABLE 6
Probabilities of Knowing Someone in the Event Subpopulation According to the Partition of the Sample by Occupation

Subpopulation	home	oohm	retd	unem	stud	rsdl
Doctors	.3986	.3906	.6957	.1948	.4196	.3391
Mailmen	.1026	.1613	.0870	.1039	.1473	.1845
Bus Drivers	.2243	.2575	.0870	.1948	.2455	.3734
Quake Victims	.1862	.2849	.2609	.2078	.2857	.3133
TV Repairmen	.1432	.2651	.3043	.2078	.3750	.2575
Priests	.3174	.2868	.4348	.2468	.2545	.2790

TABLE 7
Probabilities of Knowing Someone in the Event Subpopulation According to the Partition of the Sample by Believed Personal Network Size (in Hundreds)

Subpopulation	no ans	≤1	1–5	5–10	10–15	≥15
Doctors	.3750	.2634	.3941	.4552	.5282	.5680
Mailmen	.0625	.1156	.1525	.1253	.1897	.2524
Bus Drivers	.0000	.2245	.2585	.2353	.2615	.4272
Quake Victims	.3125	.1801	.2754	.3095	.3385	.3981
TV Repairmen	.3125	.1438	.2472	.3350	.4359	.4320
Priests	.1875	.2124	.2895	.2890	.3436	.4806

all six reference subpopulations. Except for believed personal network size, p decreases without exception for Bus Drivers with increasing socioeconomic class and almost without exception with increasing age and education. For Quake Victims and TV Repairmen the patterns of how p varies with increasing values of the four table variables are very similar. For socioeconomic class and education the patterns of how p varies for the six reference subpopulations are also very similar, indicating that functionally, for our purposes here, these are very similar attributes.

In order to see whether there is more information in this data, we plot the known value of 100ε against the various values of p in Tables 3–7 for each of the six reference subpopulations. For each reference subpopulation this yields a range of p values, from the minimum to the maximum, which we call the *p-spread* for that subpopulation. These are plotted as horizontal lines in Fig. 1. On each p-spread is shown the full sample p value and corresponding α value from Table 2. Now, although there is no simple model curve of the form (12) which comes close to passing through all the points (p, ε) given by the data in Table 2, we try to obtain the next best thing, namely, a simple model curve which intersects a maximum number of these p-spreads. Unfortunately, there is no such curve which intersects all six or even five of the six p-spreads. However, there is a small set of such curves which intersects the p-spreads for Doctors, Bus Drivers, and Quake Victims and comes close to intersecting the p-spreads for both Mailmen and TV Repairmen. The best estimating curve of this set, which just fails to intersect the p-spreads for Mailmen and TV Repairmen by about the same difference in p, has a nearest integer α value of 220. This value is virtually the same as that for the lines given by (19.r) and (17.s) and is easily within the bounds 235 ± 39 for the lines given by (16.r) and (18.r). This lends consistent support to the simple model with $\alpha \approx 220$. We note that the p values where the curve for $\alpha = 220$ intersects the p-spreads for Doctors, Bus Drivers, and Quake Victims are nonunique weighted combinations of the p values for various combinations of the informant subsamples given in Tables 3–7. Each such weighted combination of subsample p values (e.g.,

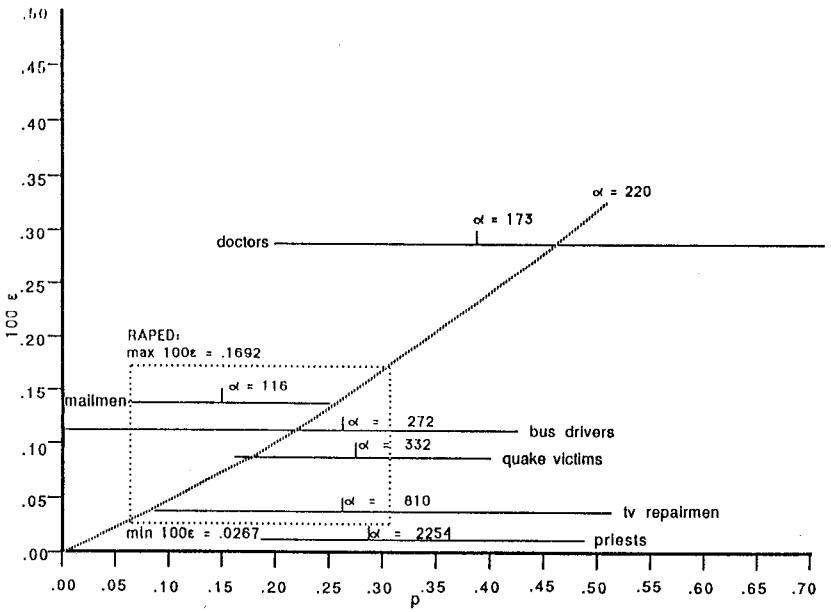

FIG. 1. p-Spreads for six reference subpopulations with best estimating simple model and estimate of relative size range for unknown raped subpopulation in Mexico City proper (t = 10,700,000).

$.35 \cdot p_{(\text{ages } 35\text{–}49)} + .20 \cdot p_{(13\text{–}16 \text{ years education})} + .45 \cdot p_{(\text{middle class})}$) may thus be viewed as an *indicator combination* for the corresponding event subpopulation. The important question here is whether an indicator combination gives consistent results under repeated sampling.

From the number of respondents in each subclass with respect to *believed* personal network size (except No Answer), using the midpoint value as the mean for each of the first four size classes and 1800 for the mean of the last size class, we obtain the average *believed* person network size of 516. This is modestly larger than the values of α obtained relative to quake victims in both Mexican surveys.

We note that for no subclass in Tables 3–7 does the monotonicity relation between p and ε given by (6) and (7) occur even approximately, which tends to disconfirm the simple model in the absence of major errors in the data. It appears that developing from this an accurate and precise method for estimating the size of an unknown subpopulation will require further development of the model and perhaps improvement of the data.

APPLICATION TO ANOTHER EVENT SUBPOPULATION IN MEXICO CITY

We can attempt to estimate the unknown size of the subpopulation of Rape Victims in Mexico City proper. In the second survey, for knowing

TABLE 8
Estimates of the Number of Rape Victims in Mexico City for Various Obtained Values of α

α:	116	196	220	235	274	332	516	810	2254
e:	14,883	8811	7850	7349	6303	5202	3348	2133	766

a rape victim we obtained the estimate $p \approx .1491$ with $t = 10{,}700{,}000$. For the various α values of interest obtained earlier, we obtain the estimates of the subpopulation of Rape Victims given in Table 8.

From the curve for $\alpha = 220$ in Fig. 1 we can estimate bounds for the number of Rape Victims in Mexico City proper according to $\min(p) = .0571$ (for Age > 65 years) and $\max(p) = .3111$ (for Education > 16 years). When the resulting p-spread for Rape Victims intersects this curve at minimum p, ε is minimum, and when it intersects it at maximum p, ε is maximum. As generally shown by the figure, we obtain $\min(\varepsilon) = .000267$ and $\max(\varepsilon) = .001692$, which translates to $\min(e) = 2859$ and $\max(e) = 18110$, so $2859 \leq e \leq 18110$. This range is too large for estimating e, but note that all e values in Table 8 except those determined by $\alpha = 810$ and 2254 lie in this interval. For the α values lying in the previously estimated interval 235 ± 39 or $196 \leq \alpha \leq 274$ we have $6303 \leq e \leq 8811$, which is a considerably better estimate.

APPLICATION TO AN EVENT SUBPOPULATION IN THE U.S.

In a Media General—Associated Press poll of 1304 randomly selected adults across the U.S. taken in April 1987, one of the questions asked was whether the respondent knew anyone with AIDS (cf. Kilman, 1987). Seven percent of them said they did. Using this figure, an estimated May 1, 1987 U.S. adult population (over 17 years of age) of 179,955,000 based on data from the U.S. Bureau of the Census (1987a, b), and the diagnosed number of AIDS victims as of early May 1987 of 35219, we can apply the simple model to estimate the corresponding value $\alpha \approx 371$. For the maximum error of $\pm .027$ in p (at an assumed confidence level of 95%), this gives a range of $223 \leq \alpha \leq 523$. This range is consistent with values and bounds for α which we have determined with this model for the Federal District of Mexico City and for Mexico City proper. Because of the cultural differences between the U.S. and Mexico, however, the similarity of these α values to those for Mexico City may only be coincidental.

REFERENCES

Bernard, H. R., Johnsen, E. C., Killworth, P. D., and Robinson, S. (1989). "Estimating the size of an average personal network and of an event subpopulation," *in* The Small World (M. Kochen, Ed.) pp. 159–175. Albex Pub. Corp., Norwood, NJ.

Kilman, L. (1987). "AIDS Most Feared Of Diseases: Poll," *Santa Barbara News Press*, Santa Barbara, CA, May 12.

de Sola Pool, I., and Kochen, M. (1978). "Contacts and influence," *Social Networks* **1**, 5–51.

U.S. Bureau of the Census (1987a). "Estimates of the population of the United States, by age, sex, and race: 1980 to 1986," *Current Population Reports,* Population Estimates and Projections, Series P-25, No. 1000, Issued February, Table A, p. 2.

U.S. Bureau of the Census (1987b). "Estimates of the population of the United States to May 1, 1987," *Current Population Reports,* Population Estimates and Projections, Series P-25, No. 1007, Issued July, Table, p. 2.