USING NETWORK SAMPLING FOR RARE POPULATIONS: AN APPLICATION TO LOCAL CRIME VICTIMIZATION SURVEYS

Ronald Czaja, University of Illinois at Chicago
Johnny Blair, University of Illinois at Urbana-Champaign

1. INTRODUCTION

A basic problem in survey sampling is the efficient location of rare populations and estimation of their sizes. If the population of interest is a small subdomain of the general population, one can, in theory, simply select a probability sample of the general population and conduct screening interviews to identify the target group. In practice, however, this approach requires samples that are very large, sometimes prohibitively so.

In the late 1950's and early 1960's, in response to recurring problems in surveys of rare populations, Monroe Sirken and his colleagues at the National Center for Health Statistics (NCHS) began experimenting with network sampling. Their initial work focused on estimating the incidence of diagnosed cases of cystic fibrosis (Sirken, Crane, Brown and Kramm, 1959; Kramm, Crane, Sirken, and Brown, 1962). Up to that time, the use of traditional survey samples to estimate rare diseases had resulted in large sample sizes and substantial reporting errors. The advantages of network, or multiplicity designs, quickly became apparent, though these designs bring with them their own particular problems and constraints.

The primary difference between network and conventional surveys is the number of enumeration units to which target respondents are linked. In conventional surveys, the counting rule links each target respondent to one enumeration unit: his place of residence. In network sampling, respon dents are linked to, i.e. can be reported by, additional enumeration units (Sirken, 1970). This linkage is accomplished by using a multiplicity counting rule. The most common types are rules which link individuals to the households of pre-specified relatives. The resulting sample is a probability sample, although all elements do not have the same probability of selection. In data analysis, cases are appropriately weighted to adjust for the different probabilities of selection. The number of households eligible to report the target respondent is elicited in the interview; the inverse of the network size is typically the case weight.

For network sampling to be successful, at a minimum, two broad conditions must be satisfied:

1. The network respondents must be able and willing to report about the event or characteristic of interest; and,
2. They must know the size of the eligible reporting network.

For our research four conditions were relevant:

1. Network respondents must know about the victimization;
2. They must be willing to report about it;
3. They must have a reasonable knowledge of the time period in which it occurred; and,
4. They must know the size of the target persons's (victim's) reporting network.

Which counting or reporting rules to use in a network survey is a critical decision. There is very little guidance from the social science literature on how broad or narrow to make the reporting rules for specific research topics. On one hand, the use of a broad counting rule allows the target person or household to be reported by multiple respondents which in theory minimizes the number of households to contact. However, as one expands the number of potential respondents, increases in reporting errors are also likely.

In the present study, we linked network sampling with telephone interviewing to determine if these two methodologies could bring the cost of conducting local victimization surveys within the means of many local jurisdictions while also increasing the precision of sample estimates. A reverse-record-check victimization survey was conducted in a small mid-western Metropolitian Statistical Area. Three types of crimes were used: robbery, assault, and burglary. In addition, a victimization interview with a random sample of victims' relatives, co-workers and close friends was conducted. In this paper we compare the results from a conventional survey with the results from network surveys using a number of different reporting rules in terms of reporting rates, bias and the mean square error.

2. MULTIPLICITY ESTIMATION

The counting rules which we assess in this paper are the following:

- Rule 1: A conventional rule in which the victims are linked only to their usual residence.
- Rule 2: A sibling rule in which the victims are linked to their usual residence and to the residences of their siblings.
- Rule 3: A parent and children rule in which the victims are linked to their usual residence and to the residences of their parents and children.
- Rule 4: A relative rule in which the victims are linked to their usual residence and to the residences of their siblings, parents and children.
- Rule 5: A close friend rule in which the victims are linked to their usual residence and to the residences of their close friends.
- Rule 6: A combined rule in which the victims are linked to their usual residence and to the residences of their relatives and close friends.

In the following expressions, the multiplicity of an event will be referred to as rule 1-6. The multiplicity model which we are using was first developed by Sirken (1979) and is further elaborated in Casady, Nathan and Sirken (1985) and Czaja, Snowden and Casady (1986). Let N be the size of the population at risk and $\theta = V/N$ be the incidence of victimization for the population at risk. Then, a multiplicity estimator for θ from a simple random sample of m households from a universe of M households for rule r is:

$$\hat{\theta} = \frac{(M/m)}{N} \left[\sum_{i=1}^{M} a_i \sum_{j=1}^{V} B_{rij} / S_{rj} \right], \quad (2.1)$$

$$r = 1,2,3,4,5,6$$

where: V = the total number of victims in the population at risk

$$a_i = \begin{cases} 1 & \text{if household } i \text{ is selected} \\ 0 & \text{otherwise} \end{cases}$$

$$B_{rij} = \begin{cases} 1 & \text{if event } j \text{ is reported at household } i \\ 0 & \text{otherwise} \end{cases}$$

S_{rj} = multiplicity of event j for rule r

For the conventional rule, rule 1, $S_{1j} = 1$ and for rules 2-6, $S_{rj} \geq 1$ is the total number of different households in which the victim and the eligible network members reside.

The expectation for the mean, variance and bias for the estimators are

Conventional

$$E(\hat{\theta}_1) = \theta p_1 \quad (2.2)$$

$$Var(\hat{\theta}_1) = \frac{\theta p_1}{m} \left[M/N - (\theta p_1) \right] \quad (2.3)$$

$$bias(\hat{\theta}_1) = -\theta(1-p_1) \quad (2.4)$$

Multiplicity

$$E(\hat{\theta}_r) = \theta \left[p_r + (p_1 - p_r) \frac{\sum_{j=1}^{V} 1/(S_{rj})}{V} \right], \quad (2.5)$$

$$r = 2,3,4,5,6$$

$$bias(\hat{\theta}_r) = -\theta \left[(1-p_r)-(p_1-p_r) \frac{\sum_{j=1}^{V} 1/(S_{rj})}{V} \right] \quad (2.6)$$

$$var(\hat{\theta}_r) = \frac{\theta(M/N)}{m} \left[p_r \frac{\sum_{j=1}^{V} 1/(S_{rj})}{V} + (p_1 - p_r) \frac{\sum_{j=1}^{V} 1/(S_{rj})^2}{V} \right] - \frac{[E(\hat{\theta}_r)]^2}{m} \quad (2.7)$$

Where p_1 = the conditional probability that a victim is reported when his or her residence is selected in the sample

p_r = the conditional probability that the victim is reported at the residence of their relative or close friend as specified in rule r, r = 2,3,4,5,6

$$MSE(\hat{\theta}_r) = Var(\hat{\theta}_r) + bias^2(\hat{\theta}_r) \quad (2.8)$$

One should note that we have ignored finite population correction factors, that the variance expression assumes that no household reports more than one event, and that the conditional probability of reporting an event for the combined rules, rules 4 and 6, are constant across rules. On this latter point, we are assuming this to be true for purposes of simplifying the testing of combined rules even though Table 3 shows that, as we discuss later, reporting by siblings and children/parents is not constant from household to household. In addition, it should be noted that the data we present below are not adjusted for people who had multiple crime victimizations.

3. DESIGN AND IMPLEMENTATION OF THE STUDY

A. Design

The logic of the design of this feasibility survey is quite simple. A sample of known crime victims was selected from the police records of a small Illinois MSA. A general crime victimization telephone interview was conducted with each victim. In the interview, respondents were asked both about crimes that had happened to them personally and about crimes that had happened to members of their pre-specified network. At the end of the interview, the name and telephone number of a randomly selected member of the victim's defined network was elicited. These network members were called and the same interview was conducted with them. This design provided for simple comparisons of self and network reports. The victim and network member samples were combined with a general population (decoy) sample to mask the source of the original list from the interviewing staff and to provide anonymity to respondents.

The sample was selected using a disproportionate stratified probability sample with systematic random sampling within strata. The stratification was by type of respondent (victim, network member, and decoy) and by type of victimization (robbery, burglary, and assault).

B. Procedures

The crime victims were chosen from police department records for the period February to September 1986. The names, addresses and telephone numbers of network members were, of course, provided by victims with whom interviews were completed. The decoy sample was selected from current telephone directories covering the MSA. Data collection was conducted primarily by telephone, with face-to-face interviewing used for only a few respondents who were not reachable by telephone.

The sample frame for the victimization respondents consisted of two parts: First, we used a tape of 2,640 robbery, assault, burglary and petty theft cases which occurred in the jurisdiction of the police department from February to September 1986. Second, we used the actual police reports, including the narrative report, for all sample cases. The sample frame provided on tape was sorted by type of victimization and then systematic random samples were selected. The corresponding police reports were then pulled from the police department files. A decoy sample of 160 telephone numbers was selected from current MSA telephone directories.

Once the target respondent was contacted, the interview began with a series of general questions about satisfaction with their neighborhood and the area in general. The next section elicited the names of network members beginning with friends and coworkers followed by parents, siblings and adult children living in other households. Only after the names of all network members had been elicited, were the questions about victimizations asked. This ensures that respondents would not simply give names of people who had been victims of crime, but would, in fact, give the names of their closest work and nonwork friends. Then the questions about victimizations of the respondent were asked. The questions eliciting the various types of victimization were adapted from the National Crime Survey.

The same set of victimization questions were then asked in regard to coworkers, close friends and relatives. A set of standard demographics about the respondent was then asked. And, finally, one relative and one co-worker or friend were randomly selected. For these two randomly selected individuals, complete name, address and telephone contact information was asked. When the respondent was unable to give complete information to contact the network member, the name address and telephone number of a secondary source who might be able to provide that information was also elicited. Interviews were conducted from the Survey Research Laboratory (SRL) Telephone Center. All refusals were reworked.

4. FINDINGS

A. Survey Completion Rates

A sample of 688 crime victims was selected from police records. Of these cases, 307 were burglary victims, 148 were robbery victims, and 233 were assault victims (Table 1).

Interviews were obtained with 254 burglary cases (82.7 percent). Of those interviews, 204 were classified as in-scope and 50 were classified as out-of-scope. A case was dispositioned as in-scope if the correct household was contacted, the respondent was the crime victim according to the police record and the crime occurred within the designated recall period.

Any of the following reasons resulted in a case being classified as out of scope: (1) the respondent was victimized in a commercial location such as a bank or gas station (2) the interview was conducted in the wrong household (3) when, during coding, it was discovered that the question was not appropriate to elicit reporting of the target crime (4) the reference period did not encompass the date of the incident or (5) the "seeded" victim was not mentioned as a relative or close friend by the network respondent.

TABLE 1

INTERVIEW COMPLETION RATES FOR THE VICTIM SAMPLE

	Burglary	Robbery	Assault	Total	Rates
Sample	307	148	233	688	100%
Interviews	254	109	196	559	81.3
In-scope	204	69	110	383	--
Not in-scope	50	40	86	176	--
Refusals	32	14	17	63	9.1
Other	21	25	20	66	9.6

Of the 148 robbery cases, interviews were completed with 109 (73.7 percent)--69 in-scope and 40 out-of-scope. Interviews were completed with 196 assault cases (84.1 percent)--110 in-scope and 86 out-of-scope.

Each victim was asked to nominate one relative and one friend/co-worker. A total of 132 relative nominations and 127 friend/co-worker nominations were obtained from the victims (Table 2). Of the 132 relative cases, interviews were conducted with 113 (85.6 percent)--52 in-scope and 61 out-of-scope. The same basic criteria applying to the victim out-of-scope interviews apply to relatives. Of the 127 friend/coworker cases, interviews were conducted with 108 (85.0 percent)--35 in-scope and 73 out-of-scope.

B. Response Model Findings

The reporting rates of victims by their own households and the households of the elgible relative and close friend networks are shown in Table 3.

For the total sample of victim households, 66% of the target crimes were reported, whereas in the network households the rates varied from a low of 26% for siblings to a high of 59% for the child-parent rule. There was significant variability by type of crime especially for the victim households. Burglary and robbery were reasonably well reported, 84% and 72%, respectively, but assaults were grossly underreported (29%). The results for burglary and robbery are similar to those found in the San Jose reverse-record study (Turner, 1972) where 90% of the burglaries and 76% of the robberies were reported. For assault, the results are similar in that it was poorly reported in both studies, however, our rates are much lower than the 48% report rate in San Jose.

In general the victim households were the better reporters but for a number of subdomains (Table 3) one or more of the counting rules had a comparable or higher reporting rate. This occurred for assaults, victims who were female or nonwhite and for both age groups. The child-parent rule had the highest reporting rates with the friend rule being comparable in many instances. Clearly, the sibling rule had the poorest reporting rates.

TABLE 2

NETWORK COMPLETION RATES

	Bur-glary	Rob-bery	Assault	Total	Rates
Relative Nominees	51	33	48	132	100.0%
Interviews	41	31	41	113	85.6
In-scope	26	10	16	52	--
Not in-scope	15	21	25	61	--
Refusals	10	0	3	13	9.8
Other	0	2	4	6	4.6
Friend Nominees	63	26	38	127	00.0%
Interviews	55	23	30	108	85.0
In-scope	21	9	5	35	--
Not in-scope	34	14	25	73	--
Refusals	7	1	5	13	10.2
Other	1	2	3	6	4.7

The expressions for bias are a function of the estimated rate of nonresponse and of multiplicity parameters (see 2.6). In Table 4 we present the ratios of the estimates of the bias for the multiplicity and the conventional counting rules. The pattern of results is essentially the same as for the reporting rates. That is, the conventional rule in most instances has a smaller bias than the other rules and the child-parent rule has the smallest bias ratio among the multiplicity rules. The results are due primarily to the higher reporting rates for these two rules. The pattern is not similar for the sibling rule where its bias ratio is comparable to the friend rule.

In Table 5 we present the estimated sample size for which the mean squared error (MSE) of the conventional estimator equals the MSE of the specific multiplicity estimator. The results are shown for an incidence rate of .001, for various subdomains, and for a metropolitan area of about 350,000 population. The results are essentially the same for the U.S. population. The data in the table can be interpreted as follows. Under the child-parent rule at an incidence rate of 1 per 1000 (.001), the estimator based on the child-parent rule has a smaller MSE than the estimator using a conventional rule for a sample of less than 6436 households. If ones sample size requirements necessitate a sample larger than 6436 households, the conventional rule is more efficient. The data in Table 5 indicates that in about three-quarters of the cells, the conventional rule is more efficient. The data in Table 5 indicates that in about three-quarters of the cells, the conventional rule is the estimator of choice. The reason is that while these sample sizes seem large they would not yield many respondents who have been victims of a crime. For example, returning to the child-parent rule for the total sample, if we assume two adults per household, an incidence rate of .001 and an average of five eligible network households to report each crime victim, a sample of 6436 households would yield about 64 victims if there was perfect reporting. Because the event of victimization is so rare in this example, the resulting sample size would not provide an acceptable sampling error for most analyses and estimation tasks.

The table cells with infinity indicate that the multiplicity rule is always more efficient than the conventional rule. This occurs for the nonwhite, under age 35 and the assault subdomains. Among the multiplicity rules, the child-parent is again the most efficient but the pattern is not as dominant as before because a number of the other rules in select instances are more efficient than the conventional rule. In general, for a condition this rare with our results, it is unlikely that network sampling would be the preferred way.

5. DISCUSSION

A major objective of our research was to determine how well network respondents--relatives and close friends of a crime victim--report crime victimizations. In this paper we examined a number of multiplicity counting rules, as compared to a conventional counting rule, in terms of victim reporting rates and a mean square error analysis.

While it may seem reasonable to conclude that network sampling is not the methodology of choice based upon the data we presented, it is also difficult to draw any final conclusions from this research because there are a number of confounding factors. It is difficult to determine how the reverse record check design and the telephone method of data collection affected our evaluation of network sampling methodology for crime victimizations.

The best multiplicity counting rules in our research were the child/parent rule and the friend rule. The friend rule provides the broadest coverage but it also presents a problem in estimation. Often there is not reciprocity between friends. Person A may name person B as a close friend, but B, when asked the same question, may not name A. This creates a problem in estimation because individuals come into the sample with unknown probabilities. We tried to overcome this by asking for a set number of friends, three, but it did not work. Approximately, 47% of the network friends did not mention the crime victim as one of their three closest friends. Our estimator does not take this into consideration. This issue must be addressed in future work before the friend rule can be an acceptable component in network designs.

Our results notwithstanding, we believe that there is a need for more experimentation with

TABLE 3

ESTIMATED TARGET CRIME REPORTING RATES FOR VICTIM AND NETWORK HOUSEHOLDS, FOR THE TOTAL SAMPLE AND SELECTED VICTIM SUBDOMAINS

Victim Subdomains	Reporting Households					
	VICTIM p_1 (N)	SIBLING p_2 (N)	CHILD, PARENT p_3 (N)	SIBLING, CHILD, PARENT p_4 (N)	FRIEND p_5 (N)	SIBLING, CHILD, PARENT, FRIEND p_6 (N)
TOTAL SAMPLE	.66 (383)	.26 (23)	.59 (29)	.44 (52)	.51 (35)	.47 (87)
CRIME TYPE						
Burglary	.84 (204)	.38 (13)	.62 (13)	.50 (26)	.57 (21)	.53 (47)
Robbery	.72 (69)	.20 (5)	.60 (5)	.40 (10)	.67 (9)	.53 (19)
Assault	.29 (110)	.00 (5)	.55 (11)	.38 (16)	.00 (5)	.29 (21)
SEX						
Male	.68 (159)	.20 (10)	.55 (11)	.38 (21)	.57 (14)	.46 (35)
Female	.65 (224)	.31 (13)	.61 (18)	.48 (31)	.48 (21)	.48 (52)
RACE						
White	.71 (304)	.25 (20)	.55 (22)	.40 (42)	.50 (32)	.45 (74)
Nonwhite	.46 (79)	.33 (3)	.71 (7)	.60 (10)	.67 (3)	.62 (13)
AGE						
<35	.60 (167)	.17 (12)	.69 (13)	.44 (25)	.31 (13)	.39 (38)
≥35	.70 (216)	.36 (11)	.50 (16)	.44 (27)	.64 (22)	.53 (49)

TABLE 4

RATIOS OF THE BIAS FOR MULTIPLICITY COUNTING RULES TO THAT FOR THE CONVENTIONAL COUNTING RULE, BY THE TOTAL SAMPLE AND SELECTED VICTIM SUBDOMAINS

BIAS RATIO $\frac{\text{Bias}\theta_r}{\text{Bias}\theta_1}$

Victim Subdomains	SIBLING	CHILD, PARENT	CHILD, SIBLING, PARENT	FRIEND	SIBLING, CHILD, PARENT, FRIEND
TOTAL SAMPLE	1.36	1.07	1.29	1.29	1.44
CRIME TYPE					
Burglary	1.80	1.41	1.90	2.16	2.52
Robbery	**	**	1.49	**	1.56
Assault	**	0.88	0.94	**	1.01
SEX					
Male	1.44	1.11	1.39	1.23	1.55
Female	1.30	1.03	1.22	1.32	1.37
RACE					
White	1.46	1.17	1.47	1.53	1.75
Nonwhite	**	**	0.87	**	0.78
AGE					
<35	1.37	0.94	1.18	1.52	1.43
≥35	1.32	1.23	1.40	1.15	1.46

** = less than 10 respondents

TABLE 5

ESTIMATED SAMPLE SIZE FOR WHICH THE MSE OF THE CONVENTIONAL ESTIMATE EQUALS THE MSE OF THE MULTIPLICITY ESTIMATOR AT AN INCIDENCE RATE OF .001 FOR SELECTED COUNTING RULES: MSA

	MSE Intersection Sample Size*				
Victim Subdomains	SIBLING	CHILD, PARENT	SIBLING, CHILD, PARENT	FRIEND	SIBLING, CHILD, PARENT, FRIEND
TOTAL SAMPLE	1047	6436	1762	2591	1758
CRIME TYPE					
Burglary	2063	4820	2308	2539	1864
Robbery	**	**	1437	**	2189
Assault	**	∞	∞	**	19,795
SEX					
Male	1937	8340	2791	8585	3185
Female	2269	23,918	4439	3829	3774
RACE					
White	1180	3650	1498	2054	1386
Nonwhite	**	**	∞	**	∞
AGE					
<35	1785	∞	4809	1856	2904
≥35	2629	4336	2854	13,542	4100

** = less than 10 respondents

*The MSE intersection sample size is the estimated sample size where the MSE of the conventional estimator equals the MSE of a specific multiplicity estimator.

network sampling rather than less. We suggest that our research in this area be repeated but with some modifications. We suggest a split method design with half the interviewing conducted by telephone and the other half face-to-face. Between five and ten percent of the respondents reported a crime but not the target crime. We suspect that forgetting is the major reason for these non-reports. Therefore, we also suggest some experimentation with cognitive psychology methods to improve recall and reporting.

REFERENCES

Casady, Robert J., Nathan, Gad, and Sirken, Monroe G. (1985), "Alternative Dual System Network Estimators," International Statistical Review, 52, No. 2, 183-97.

Czaja, R., Snowden, C. and Casady, R. (1986), "Reporting Bias and Sampling Errors in a Survey of a Rare Population Using Multiplicity Counting Rules," Journal of the American Statistical Association, 81, 411-19.

Kramm, E.R., Crane, M.M., Sirken, M.G., and Brown, M.L. (1962), "A Cystic Fibrosis Pilot Survey in Three New England States," American Journal of Public Health, 52, 2041-57.

Sirken, M.G. (1970), "Household Surveys With Multiplicity," Journal of the American Statistical Association, 65, 257-66.

-------- (1979), "A Dual System Network Estimator," in Proceedings of the Survey Research Methods Section, American Statistical Association, pp. 340-42.

Sirken, M.G., Crane, M.M., Brown, M.L., and Kramm, E.R. (1959), "A National Hospital Survey of Cystic Fibrosis," Public Health Report, 74, 764-70.

Turner, A.G. (1972), San Jose Methods Test of Known Crime Victims. Statistics Technical Report No. 1. Washington, D.C.: Law Enforcement Assistance Administration.