

Use of Employee Interviews to Estimate Worker Populations:  
Comparison of Multiple Recapture Methods

by T. Fischbach, National Institute for Occupational Safety and Health,  
J. Wittes, National Heart, Lung, and Blood Institute,  
E. Dacey and S. Shilling,  
National Institute for Occupational Safety and Health

Information about the size of worker populations is very useful for occupational health surveillance. For example, when the number of workers exposed to a suspected hazard can be reliably estimated, rates of illness, injury, and/or death can be estimated to obtain epidemiological evidence of the associated occupational risk. Plant records can be used to determine the number of workers when these have been maintained. However, very often such records are either not available for the entire period of interest or else they are too incomplete to use to estimate the desired population at risk.

An overlooked source of information for estimating worker population size is the current workforce. Currently employed workers, when interviewed individually or in groups, can provide lists of both current and former co-workers which might be useful for estimation of the desired or target population. Recognizing the potential of this source of information, the National Institute for Occupational Safety and Health (NIOSH) has attempted to develop and test multiple recapture methods to analyze worker generated data for population at risk estimation.

When the lists can be assumed to be independent, and if each lister recalls or includes each member of the desired population with the same probability, then methods developed by Wittes(1) can be used. If the latter assumption holds but the assumed independence of sampling does not, then methods developed by others, for example Fienberg(4), can be used. However, great difficulties arise when any given lister recalls members of the population with differing probabilities. These difficulties can be overcome if the population is stratified into groups whose members are recalled with nearly equal probability by each lister (inter-lister variations in recalling a particular population member do not pose a difficult problem but variations among population members for a given lister do).

In an earlier small scale study, Wittes interviewed fellow mathematics department staff members at Hunter College to estimate the number of current and former staff. Perceived period of employment was used to stratify the desired population and the results seemed promising. The present study is intended as a larger pilot study to test the method on a larger scale yet in a locale susceptible to more control than might be encountered in target applications.

Lists may be obtained by interviewing workers individually or in groups. Group members may stimulate each other to produce a larger list. Moreover, group members may check each other's memories and filter out the names of persons not in the desired population to produce lists of better quality. On the other hand, group interviewing results in fewer lists than would be realized if the same number of individuals were interviewed separately. Wittes(1) has shown that the variance of the estimated size is a decreasing function of the number of (independent) samplings or lists. However, this difference would be offset if inclusion probabilities for group lists are sufficiently large.

Potential sources of information are not limited to workers who are members of the desired population but include those working in close proximity to the desired group. Obtaining lists from such workers may be advantageous if the current workforce in the desired population is small or if it is scattered among several worksites. However, any advantage depends on the usefulness of the information from such workers. This may depend greatly on the type and frequency of contacts between the two groups.

Group output may be susceptible to optimization by the researcher. If group members are interviewed individually before the group interview, will this increase output during the group interview? The individual interviews could stimulate memories. If workers who are not members of the desired population are used, should they be grouped separately or with members of the population? If members and non-members are grouped together, does this result in less output than for groups of members only?

The major purpose of the pilot study was to gain experience in applying multiple recapture methods to this problem and to compare results from individually generated lists to group generated lists.

OBJECTIVES OF THE  
PILOT STUDY

The objectives of the pilot study are as follows in order of priority:

1. To compare worker population size estimates obtained with multiple recapture methods, using lists generated from a sample of workers and with period of employment for stratification, to the

- corresponding worker population sizes obtained from official records. A sub-objective is to evaluate the usefulness of worker list data for population estimation when the sources are drawn from a different department or work group than the target.
2. To compare estimates from group interview generated lists both to estimates from individual interview generated lists and to sizes from official records when the number of subjects is the same for both individual and group interviews.
  3. (Low priority) To determine if group output is related to:
    - a. Prior individual interviewing of members.
    - b. Heterogeneity of the group in terms of target population membership.

## STUDY DESIGN

The study design and protocol are as follows:

1. Two branches within a NIOSH division in Cincinnati were selected for convenience, to contain costs, and because of the availability of good records. The division selected had not undergone major reorganization since 1976. Two of the three available branches seemed likely to be in contact during the normal course of work and were selected for the study. The population to be estimated was the professional, non-professional, permanent, and temporary staff employed in each branch at any time during the period 1976 to November 1982 (the date of the interviews). The size of each branch staff was to be estimated as a logically distinct problem so that a person who had worked in one branch and then worked in the other during the period would count separately in each population.
2. Ten (10) subjects, hereafter termed listers (groups are also termed listers), were randomly selected from each branch for interviewing to produce lists of workers and:
  - a. Eight were randomly assigned for individual interviews and of these
    - Four were assigned to an all branch group for group interviews,
    - Two were assigned to a group with two persons from each branch for interview, and
    - Two were not assigned to group interviews.
  - b. Two were randomly assigned to group interviews only, and the groups had members from both branches.
3. Individual interviews were done in two days. Order of interviewing was randomized. Listers were not told whether or not they would be group interviewed. Listers were asked not to discuss interviews with others.

4. The order of group interviews was randomized. The interviewer did not know who was to be interviewed until the interview began.
5. Listers were first asked to identify themselves, their branch of employment, and when they were first employed in their current branch. Next, the lister was asked to name all persons he/she/they could recall who had worked in the same branch at some time since 1976. They were asked to describe persons who they could recall by description but not by name. The lister was also asked to recall when the person had first been employed in the branch and when he/she had left if the person was not currently employed. Finally, the lister was asked to estimate how well that person was known by his/her members of the branch. After the lister could not recall any more persons from his own branch, he was asked to recall persons who had been employed in the other branch and provide the information described above.
6. Listers were reinterviewed, when necessary, to identify the names of those persons recalled by description but not by name. They were shown lists of names recalled by other listers and asked if any corresponded to those whose names they had not remembered. However, persons not recalled during the initial interview could not be added to the list.

This design resulted in

- Sixteen individual interviews, eight per branch,
- Four group interviews with four persons per group as follows:
  - Four from branch one; all previously individually interviewed,
  - Four from branch two; all previously individually interviewed,
  - Two from each branch, hereafter termed a mixed group; all previously individually interviewed,
  - Two from each branch, hereafter termed a mixed group; none previously individually interviewed,

The NIOSH personnel office prepared a list of staff for the two branches for the period studied. The investigators did not view this list until the analysis had been completed.

## ANALYSIS OF RESULTS

Editing

The data required extensive editing to resolve:

- differences in name spelling,
- multiple names for the same person, e.g. women who were recently married and referenced by their maiden or married names or both with a hyphen,
- persons recalled by description only by

one or more listers and possibly by name by one or more other listers, and -persons apparently listed twice by the same respondent.

The principle followed for editing was to use only the current staff to resolve such problems.

The names of all individual listers were deleted from all individually generated lists. Thus, each lister samples from the same population. This would not occur otherwise as no lister included himself. Similarly, the names of all group members were deleted from all group lists.

## Methods

### Objectives 1 and 2

Accomplishment of objectives 1 and 2 required generation of estimates of the size of each branch using, respectively, individual interview generated and group interview generated data after stratifying each branch by period of employment as described below. These estimates were compared as follows:

- the estimates from each interview method to the sizes from NIOSH records, hereafter termed the official sizes;
- the estimates obtained from the two methods of interviewing were compared to each other.

The basis for stratification was the period the listed person was employed in his/her branch during the period from 1976 to November, 1982 as perceived by the listers. The data for this were the listers' recollections of the person's branch of employment, the year he/she was first employed, and the year he/she left if the person was not currently employed. For each interview method the medians of the dates of entry and departure, respectively, were determined for each person listed by at least one lister. These were the perceived periods of employment for the persons listed. The number of strata and their boundaries were chosen to maximize the sizes of the resulting groups and to achieve groups which, in the judgment of the investigators, would be most plausibly homogeneous in terms of probability of recall.

Estimates were generated by methods described by Fienberg(1) using the loglinear frequency analysis program, LOGLIN, written by Oliver and Neff(3), as follows. The data were the frequencies that any listed person was listed or not listed by each combination of listers. These data were arrayed in a k factor contingency table in which each factor corresponds to a lister (thus, k is 16 for the individual interview method and k is 4 for the group interview method) and has two levels, "listed" and "not listed." The frequency for

the cell corresponding to "not listed by all listers" was, of course, unknown (a structural zero). With LOGLIN a loglinear model can be fit to the observed cells. This model was then assumed to fit the entire array and it was used to estimate the frequency of persons not listed by any lister. The estimate of the total population was the sum of the number of persons listed by at least one lister and the estimate of those listed by none of the listers.

The following forward stepwise procedure was used to empirically and parsimoniously - regarding the number of parameters included - fit a hierarchical model to the data. At any step the goodness of fit of the model under consideration was tested with a chi-square test using the value of a likelihood ratio chi-square statistic available as output from LOGLIN. If the result was not significant at the .05 level, that model was used. If the result was significant, a different model was considered. The model initiating the procedure was that positing independence among all k listers. The subsequent model at any step, if necessary, was that hierarchical model with the fewest additional parameters which decreases the likelihood ratio chi-square the most. If the model with all k-1 factor interactions was rejected, then no estimate was possible with this method. Unfortunately, this method does not yield an estimate of the variance of the estimator of the population size. Fortunately, an alternative method exists to produce confidence interval estimates and this is described below.

The following method suggested by Regal(2) was used to find a 95 percent confidence interval estimate for any desired population size. A model was chosen by the procedure described in the previous paragraph. First, lower and upper boundary points for a 95 percent confidence interval estimate of the number of persons not listed by any lister were found. These were the values that satisfied the following condition: such a value inserted for the frequency of persons not listed by any lister resulted in an increase of 3.84 for the chi-square for lack of fit for the previously chosen model to the full k-way table. The lower bound was the largest such value which was smaller than that estimated using the procedure described previously. The upper bound was the smallest such value which was greater than that estimated using the procedure described previously. The observed number of persons listed by at least one lister was then added to those boundary values to obtain the boundary values for the 95 percent confidence interval estimate of the desired population.

Estimates were compared to official values in two ways: (1) The 95 percent confidence interval either did or did not include the official value; and (2) The point estimate was either lower than, equal to, or higher than

the official value. The first comparison tends to indicate the accuracy of the method while the second tends to indicate the bias of the method, especially if a definite pattern is apparent for several such comparisons involving the same interview method. Two methods were compared by comparing the results of their respective comparisons to official values and by comparing the widths of the 95 percent confidence intervals. However, comparison of widths of confidence intervals can be misleading as small widths may be the result of greater departure from the assumptions, particularly that regarding equal sampling probabilities over the population for any lister.

The usefulness of including listers from outside the target population was evaluated in two ways: (1) By comparing estimates using individual interview generated lists including the outside listers to those computed without using the outside listers; and (2) By comparing the numbers of persons listed, i. e. the sizes of the lists, by listers for their own branch to the corresponding sizes of lists for the other branch. Both sets of comparisons must be interpreted cautiously. The first because the estimates made with the outside listers were based on twice as many lists as those made without the outside listers. The second because list size is only one factor affecting the precision of the resulting estimates.

### Objective 3

The effect of prior interviewing on group output was estimated as the difference between the total number of persons listed by the mixed group whose members had been interviewed individually before the group interview and the corresponding total for the mixed group whose members were not individually interviewed. The effect of including persons from both branches in the same group on group output was estimated as the difference between the sum of the total number of persons listed by the two mixed groups and the corresponding total sum for the remaining two groups.

## R E S U L T S

### Completion of the interviews

Individual interviews typically required about one hour and the 16 interviews were completed in two days. A third day was used for reinterviews to secure names of recalled but unnamed persons in the original interviews. Group interviews typically required 2 hours and the four group interviews were completed in two days.

### List sizes

The list sizes, or the numbers of persons recalled during the 16 individual interviews

are shown in Table 1 by perceived branch (that is, the branch of the recalled person as recalled by the lister) of the listed persons and by the branch of the lister. The corresponding information for the group generated lists is shown in Table 2. In either case, as might be expected, it is evident - with the single exception of lister 6 in branch 1 - that list size or list output is greatest when the lister recalls persons from his/her/their own branch as opposed to the other branch. However, this does not necessarily imply that "outside" staff are poor sources of information. Also,

as was expected, the list output of the groups was higher than that for the individuals. This does not, however, imply superiority of group generated lists as there are only 4 group lists for each branch while there are 16 individually generated lists for each branch.

The group output data provide point estimates which are relevant to the third objective. First, the mixed branch group whose members had been previously interviewed produced larger lists than the mixed branch group which had not been previously interviewed. This suggests that interviewing persons individually prior to the group interview tends to increase output, which seems reasonable. Second, the two mixed branch groups produced larger (combined) lists than the two homogeneous groups, although much of this difference is attributable to the low output of one of the latter groups. Thus, the evidence - which could be only suggestive at best - does not provide an answer to the question of whether group output is increased or decreased if its members are selected from an "outside" work population in addition to those from the population of interest.

### Stratification

Stratification was done independently with the group and with the individual data. Because of the small output for both individuals and groups, it seemed desirable to keep the number of strata as small as possible so two were tried. Stratum 1 was defined as all current staff, or all those persons with a median termination date of 1982. A few persons actually terminated in 1982 were included as "current staff". Stratum 2 was defined as all former staff or those with a median termination date of 1981 or before. This choice was motivated in part because estimation of the size of the current staff can usually be done by several methods while estimation of former staff is usually more difficult. This produced strata that seemed large enough for analysis.

### Estimation of Branch Staff Size by Stratum

The estimates produced by analysis of the data are shown in Table 3 by branch and stratum and by method of data generation, i.e. group generated or individually generated

lists. The third set of estimates shown in Table 3, labeled "Mod. Individual," are those obtained from individual interview generated data using only listers from the target branch. The sizes of the target populations obtained from NIOSH personnel records are also shown in Table 3. The last column of Table 3 provides a brief descriptor for the model fitted to the data. "Indep." indicates that independence among the listers was not rejected. Otherwise, the complexity of the model fitted is denoted by the number and type of interactions in the model.

In every case for the individually generated data the 2 to the 16th power layout, which has more than 65,000 cells, was too large for practical analysis with LOGLIN or almost any standard contingency table analysis program. However, with so many cells and so few cases, analysis was feasible only if independence among the 16 samplings was assumed. An approximate test of that model was developed as follows. The 8 listers from each branch were randomly divided into two sets of four each. A model assuming independence among the four listers was fit to the four-way layout for each set. Lack of fit to independence was not statistically significant at the .05 level for any set of four listers for either branch or stratum. This made it possible to analyze the data using the model that each of the 16 listers 'conducted' independent Bernoulli Censuses of the target populations. A description of the method is found in Wittes(1). The resulting point and 95 percent confidence interval estimates are shown in Table 3. In all cases the lower bound of the confidence interval is higher than the size indicated by NIOSH records.

In all cases except one the lower bound of the 95 percent confidence interval estimates generated from the group lists is higher than the size indicated by NIOSH records. The exception is the current staff of branch 1 for which the interval includes only the one integer value indicated by NIOSH records.

#### Comparison of Group and Individually Generated Estimates

If the population size provided by NIOSH records is used as the target, it appears that both methods tend to overestimate the population size. It also appears that in general the individually generated list data produces estimates which appear to have less bias and variance than that produced by group samplings.

#### Results from Modified Individually Generated Data

When the "outside" listers are not used the estimates appear improved in all cases: point estimates are closer to the values indicated by NIOSH records and confidence intervals are narrower.

#### Further Analysis of Lists

The comparison of estimates to target values, as indicated by NIOSH records, was not the only source of information available. Each method produces a list of persons listed at least once for each target population. These lists were compared to the corresponding lists of staff prepared from NIOSH records to develop information on the sensitivity and specificity of the lists of the two methods.

The results are shown in Table 4. Persons on any generated list could be either "true" cases of members of the target population or they could be "bogies," i.e. persons not indicated as being members of that population in NIOSH records. Members in the target population may either be included in the generated list or they may be "missed." The results in Table 4 suggest that the group and individually generated lists tend to miss about as many members of the target populations (although the group lists are slightly better). However, with the one exception of Branch 1, Stratum 1, the group samplings produce more "bogies." Thus, some of the greater group output was the production of more bogies.

Excluding the "outside" listers appears to reduce the number of bogies. The values shown in Table 4 indicate that while the number of missed population members is no smaller than before (it could not be), it is no greater, and the number of "bogies" is smaller for the modified individual method. Thus, the outside target branch listers seem to contribute to higher "bogie" counts and do not reduce the number missed.

#### Output and Tenure of Staff

A reasonable expectation is that the staff employed for the longest period would have better recall, especially of former staff. Results from this study and those from Wittes' previous study at Hunter College do not provide support for this expectation. In this case the list output for persons employed an intermediate period, neither the longest nor the shortest, tended to be greatest. Wittes also found that "bogie" production was often greater for the "old timers."

#### C O N C L U S I O N S

While stratification by period of employment does appear to be useful, neither group generated nor individually generated list data appear to result in accurate estimates of population size. Perhaps other bases for stratification or other methods for screening contributors would improve the results. Size estimates from lists from individuals within the target branch seem to be least biased and most precise. Those developed from individual listers appear better by those criteria than those from group

lists. Group and individuals miss about as many persons in the target populations. Groups tend to list more non-members as members than individuals and, among the latter, individuals not in the target branch list more non-members as members than individuals from within the target branch.

It does not appear useful to include staff from outside the target populations as listers. While list output seems increased by

so doing, much of this is "bogey" production or the inclusion of staff not members of the target population. Of course, this may be related to the specific characteristics of the worker groups chosen for this pilot study. Results for other groups may be different.

## REFERENCES

1. Wittes, Janet Turk. Estimation of Population Size. Unpublished dissertation, Harvard University, Cambridge, Mass. 1970.
2. Regal R. Private communication to one of the authors.
3. Oliver D. and Neff R. Loglin -- Computer Program to Analyze Multidimensional Contingency Tables. Computing Center of the Harvard School of Public Health.
4. Fienberg, S.E. The Multiple recapture census for closed populations and incomplete 2 (k) contingency tables. Biometrika 59:591-603, 1972.

Table 1: Numbers of Persons Listed by Branch by Individual

Lister		Number Persons Listed	
Individual	Branch	Perceived Branch	
		1	2
1	1	20	11
2	1	10	0
3	1	18	15
4	1	22	12
5	1	30	14
6	1	22	29
7	1	33	12
8	1	23	12
1	2	4	21
2	2	13	23
3	2	8	27
4	2	17	26
5	2	9	29
6	2	11	28
7	2	18	22
8	2	5	21

Table 2: Numbers of Persons Listed by Branch by Group

Group	Number Persons Listed	
	Perceived Branch	
	1	2
All Branch 1; All Previously Individually Interviewed	36	24
All Branch 2; All Previously Individually Interviewed	23	45
Both Branches; All Previously Individually Interviewed	38	41
Both Branches; Not Previously Individually Interviewed	24	51

Table 3: Estimated Population Sizes by Population and By Method.

Population Branch	Stratum	Method	Estimates		Number Distinct Cases	Model	
			Population Size Point Est.	95% Confidence Interval			
1	1	NIOSH Records Group	15	-	--	indep. 2 parameters	
		Individual	22	15 - 22	22		
		Mod. Indivd.	18	18 - 18	18		
1	2	NIOSH Records Group	27	-	--	indep.	
		Individual	37	33 - 46	34		
		Mod. Indivd.	33	29 - 36	32		
2	1	NIOSH Records Group	20	-	--	1 two-factor interaction	
		Individual	24	24 - 25	22		
		Mod. Indivd.	22	22 - 22	22		
2	2	NIOSH Records Group	20	-	--	indep.	
		Individual	21	51 - 83	46		
		Mod. Indivd.	67	34 - 47	36		
						29	indep.

\* Stratum 1 is all currently employed staff;  
 Stratum 2 is all former staff after 1976 but not employed as of November 1982.

Table 4: Numbers of Identified, Missed, and "Bogies" by Population and by Method of Data Generation.

Population Branch	Stratum**	Method	Total	Numbers			Actual size
				Listed True	"Bogies"	Missed	
1	1	Group	15	15	0	0	15
		Individual	22	15	7	0	15
		Mod. Indivd.	18	15	3	0	15
1	2	Group	34	19	15	8	27
		Individual	32	20	12	7	27
		Mod. Indivd.	30	20(21)*	10(9)*	7(6)*	27
2	1	Group	24	19	5	1	20
		Individual	22	19	3	1	20
		Mod. Indivd.	20	19	1	1	20
2	2	Group	46	17	29	4	21
		Individual	36	15	21	6	21
		Mod. Indivd.	29	15	14	6	21

\* One listed case could be a member of the population.

\*\* Stratum 1 is all currently employed staff;  
 Stratum 2 is all former staff after 1976 but not employed as of November 1982.