DEMOGRAPHY© Volume 15, Number 3 August 1978

AN ESTIMATE OF A RARE POPULATION GROUP: THE U.S. JEWISH POPULATION

Bernard Lazerwitz
Department of Sociology, Bar-Ilan University, Ramat-Gan, Israel

Abstract—High, middle, and low estimates for a rare population group, the Jewish population of the United States, are presented together with their root mean square errors. These estimates are based upon a national sample whose essential survey design features are outlined. The features indicate that difficult-to-find populations can be sampled in adequate numbers if some sort of a list can be developed with a fair proportion of the population. To this list must be added an integrated area sample.

Those fields of the social sciences whose basic concern is the study of rare population groups, such as people with some type of mental or physical handicap or members of small minority groups, constantly struggle with the need to estimate their sizes and characteristics. This paper presents the essential features of a survey design which was used to obtain an estimate of a rare group, U.S. Jews. It is hoped that this presentation will aid demographers in their efforts to study rare, but important, national population groups.

SAMPLE DESIGN

U.S. Jews compose approximately 3 percent of the country's population. Furthermore, apart from the New York metropolitan area, the Jewish proportion of major metropolitan populations seldom reaches as high as 8 percent. Straight area probability sampling would require screening many thousands of housing units for Jewish occupants at a prohibitive cost in interviewer time and dollars, so some combination of list and area sampling is called for. The lists, to be useful, should contain a fair proportion of the desired type of household. If such a list can be obtained, field screening costs could be considerably reduced.

Fortunately, a sizable proportion of the home addresses of the nation's Jews are known to local Jewish federations. Since this survey was commissioned by the Council of Jewish Federations and Welfare Funds, the roof organization of local federations, there was official access to the lists available in the various local federations. However, there does not exist any list with a sizable proportion of the Jews who live in the New York metropolitan area. Since the Jews of this area do form a considerable percentage of the local population, they had to be sampled by standard area probability methods.

Overall Sampling Fraction

The task of determining an overall sampling fraction was complicated by the obvious fact that the very piece of information required to design the survey's sampling fraction, a solid estimate of the nation's Jewish population, was missing. After studying a variety of estimates of the Jewish population, obtained by several approximate methods, and applying the collective judgment of fellow scientists involved in the overall survey design, we decided to work with an initial estimate of 5,700,000 Jews living in private housing units. With an initial target goal of 12,000 "effective" Jewish interviews, we obtained an overall sampling fraction of 1 in 131. (The word "effective" here means the

number of expected interviews, ignoring reductions in strata sampling fractions as a result of disproportionate stratification and weighting procedures.)

National Stratification

The Council of Jewish Federations and Welfare Funds was able to provide reasonable estimates of the Jewish population of the nation's communities. This information permitted the division of the United States into several strata. Each stratum was divided into a number of primary sampling units. Eighteen primary sampling units were designated as self-representing psu's, 20 primary sampling units represented themselves and other units, and 14 counties represented those counties containing no Jewish population according to the available information. [The details of the sample design including the stratification system and selection of sample primary sampling units can be found in Lazerwitz (1973).]

Actual sample housing units were selected by a two-phase process. First, the indicated lists were obtained from each psu's Jewish federation. Next, area probability samples were selected in each sample psu and integrated with federation lists, so that no area sampling housing unit found on federation lists was permitted to come into the local area samples. Finally, housing units of the combined list and area samples were screened for Jewish occupants who were approached for interviewing.

The Eligible Survey Respondents

Sample housing units were screened for the presence of Jewish respondents by an introductory interview which averaged five to ten minutes. This determined who lived in the housing unit and whether any of the occupants had been born Jewish, had a parent who had been born Jewish, or regarded himself as being Jewish. If any of these points applied to any housing occupant, that unit was eligible for interviewing. Family members temporarily away, say, at college, were assigned to the involved sample housing unit if they were then living in some form of group quarters or an institutional setting. Otherwise, they were dropped from that housing unit's list of occupants and had their chance of entering the survey via their own housing unit. Jews in homes for the aged, prisons, or custodial care in mental hospitals or other equivalent settings were excluded from survey coverage.

Survey Bias

Of course, some survey bias results from screening and from nonresponse. In addition, a small area sample cut-off bias was risked by not sending interviewers to some of the area samples. Portions of some area samples that fell into outlying rural areas or into counties with "no known Jewish population" were just dropped to further reduce heavy field costs. As often as possible, the list approach was extended to such cut-off bias territory, even to the extent of searching local telephone directories for respondents with "distinctively Jewish names."

Sampling Error Model

The sampling error computational model for this design is based upon the system developed at the Survey Research Center of the University of Michigan by Kish, Frankel, and Hess. The details of their computation models are available in Kish (1965, pp. 190–203), Kish and Hess (1959), Frankel (1971), Kish and Frankel (1970), and Kish et al. (1972). Although simple random sample variance formulas can be applied to the list sample data, it is more efficient to link the variance computations for both area and list sample components. This the Kish-Hess-Frankel model permits after the individual list sample housing units are grouped into small clusters.

AN ESTIMATE OF THE JEWISH POPULATION OF THE UNITED STATES

The Estimate From Survey Data

This survey yielded a national sample of 7,179 housing units that contained one or more Jews. After we adjust for the dis-

proportionate sample design features, this becomes a weighted sample of 15,145 Jewish housing units at an overall survey sampling fraction of 1 housing unit in 131. Of these sample Jewish housing units, 79 percent gave interviews and had an average of 2.84 Jews per interviewed housing unit.

We know from a considerable amount of survey research that housing units whose occupants refuse to give interviews or cannot be found at home after a considerable number of visits by interviewers, calling at a variety of times and days of the week, do have somewhat fewer people in them than the interviewed housing units. For example, such nonresponse housing units are occupied by a disproportionate number of young adults, people living alone, people who are single or divorced, people who work long hours, people who live in the central districts of large cities, or occupants of large apartment structures. Hence, it seems best to use a somewhat lower average number of Jews for the nonresponse housing units. However, even if the average number of Jews per nonresponse housing unit falls as low as 2.0, it will have a limited effect upon the average number of Jews per sample housing unit. For example, assuming 2.84 Jews per interview housing unit and 2.0 per nonresponse housing unit gives (.79) (2.84) + (.21) (2.0) = 2.66 Jews per sample housing unit. On the other hand, with 2.65 Jews per nonresponse housing unit, we have (.79) (2.84) +(.21) (2.65) = 2.8 Jews per sample housing unit.

Note the small change of the average figure from interviewed to all sample housing units. The figure of 2.8 Jews per Jewish housing unit is the one used. This permits the incorporation of nonresponse Jewish housing units into our work at the price of using another random variable, average number of Jews per sample housing unit, as a multiplier.

With an average of 2.8 Jews per sample housing unit, a survey sampling fraction of 1 in 131, and an effective sample of 15,145, one gets an estimated U.S. Jewish population of 5,555,186 from the survey data.

The Standard Error of the Survey Data Estimate

The formula used to obtain the sampling error of the survey data estimate is

$$y^2 \text{ variance } x + x^2 \text{ variance } y, \quad (1)$$

where

y = the average number of Jews per sample housing unit (containing one or more Jews), 2.8 here,

and

x = the number of Jewish housing units in the survey sample.

Computer calculations give

$$\text{variance } x = 854{,}819,$$
$$\text{variance } y = .00775.$$

Hence, we have the following variance for the total number of Jews in the sample:

$$(7.84)\ (854{,}819) + (15{,}145)^2\ (.00775) = 8{,}479{,}406.$$

Then, to obtain the variance of the estimated number of Jews in the United States, we multiply the above figure by $(131)^2$ and get 145,515,086,366. Of course, the figure for a 95 percent confidence limit becomes two times the square root of the list figure, or 762,928.

The Coverage Bias

There are several sources of survey bias that need to be considered. When adjusted to the survey rate of 1 in 131, there are 207 sample housing units (weights) and 7,396 area sample housing units (weights) in which the field force was unable to determine if the occupants were Jewish or not Jewish. This is, of course, but a small fraction of the quite large number of housing units screened for Jewish occupants during the course of the survey. Based upon field staff judgment and the extent to which local list sample call-backs found Jewish units, it is judged that a maximum of 100 out of the 207 list sample housing units can be assumed to have had Jewish occupants. However, the undetermined area sample housing units were those not

on any of our Jewish federation lists. In this work, it has been assumed that their percentage with Jewish occupants is the same proportion that the estimated number of national Jewish housing units bears to the number of occupied housing units reported by the U.S. Bureau of the Census for all standard metropolitan statistical areas with 200,000 or more population in 1970. This approach gives

$$\frac{1,983,995}{40,178,000} = .0494,$$

or just about 5 percent of 7,396. As a result, adjusting for those survey housing units at which interviewers were not able to ascertain survey eligibility, but which are now considered to be Jewish housing units, adds 100 + 370 = 470 sample housing units which, in turn, adds 172,400 Jews to the above national population estimate.

An additional source of bias in our estimate derives from the area sample cut-off used to minimize expensive field work. When this limitation is considered, the sample psu's can be classified into the following five groups.

Group 1: New York—completely sampled by an area sample design.
Group 2: 14 psu's fully sampled by a combination of lists and area sampling.
Group 3: 14 psu's with a good list and area sample but which had a cut-off bias for a small portion of their area.
Group 4: 5 psu's that had a good coverage with their list samples but no area samples.
Group 5: 5 psu's that had poor list samples and no area samples.

Only groups 3, 4, and 5 suffer from cut-off biases. For a high level estimate, it has been assumed that the ratio of area sample to total sample found for the psu's of group 2 ought to be just about the same for the psu's of group 3. This ratio is .58 for group 2 and is .64 for group 3 even with the cut-off bias. Since group 3's ratio of area sample to total sample is even larger than that for group 2, and we know their cut-off bias territories to be quite small, it seems best to ignore any numerical loss from group 3.

It was judged that group 4 psu's most closely resembled seven psu's of group 2. These seven psu's have a ratio of area samples to list samples of 1.1. It is then assumed that the same ratio holds for the psu's of group 4. Multiplying the list sample weights of the psu's of group 4 by 1.1 and carrying through our above population estimate approach adds 125,812 Jews onto our survey data national estimate.

Now, we need an upper estimate for the missed Jewish population of sample psu group 5 to complete this struggle with the cut-off bias. For the five psu's of this final bias group, we applied the 1.1 ratio of area sample yield to list sample yield used for bias group 4. Then, telephone directories were obtained for all territories not covered by the limited available list samples. A sample of names was selected from all these directories and the distinctively Jewish names screened for survey eligibility. Then, if we apply the national sample average of 15.5 percent being distinctively Jewish out of all sample Jewish names, we can use the number of Jews from the telephone directory sample who have distinctively Jewish names to get an estimate of the Jewish population in those parts of this psu group beyond list sample territory.

All told, these adjustments to available list samples and application of the "distinctively Jewish names" assumption gave 406 missed Jewish housing units. Applying the 2.8 times 131 standard sample blow-up factors gives an additional 148,921 Jews from the coverage bias of psu group 5.

These estimates are put together in Table 1. The middle bias estimate is assumed to be one-half of the high bias estimate. The low bias estimate is assumed to be zero. The table values can best be

rounded to a high estimate of 6,000,000 U.S. Jews, a middle estimate of 5,780,000 U.S. Jews, and a low estimate of 5,560,000 U.S. Jews.

Root Mean Square Error

The standard error, which is also our low estimate error, is 381,464. The formula for the extension of the sampling error estimates to include the two bias figures is (see Kish, 1965, p. 510):

$$\text{Root mean square error} = [(\text{variance}) + (\text{bias})^2]^{1/2} \cdot \quad (2)$$

For the high population estimate, there is a root mean square error of

$$[(145{,}515{,}086{,}366) + (447{,}133)^2]^{1/2} = 587{,}744.$$

For the middle population estimate, there is a root mean square error of

$$[(145{,}515{,}086{,}366) + (223{,}567)^2]^{1/2} = 442{,}151.$$

SUMMARY

As the details of this article show, it is feasible to design a practical, reasonably efficient national sample of a rare population group which can provide good estimates of its size and characteristics. Such a procedure is most effective through a combination of a basic list sample and an interlocking, nonduplicating area sample.

This design and population estimate approach can readily work on the level of the local community. In fact, the sample design given here is a "richer" form of the rather usual type that restricts itself to a locally available list. (It is dangerous to restrict any sample of a rare population group to available lists as the findings of this survey clearly show.) The estimates have also focused upon the data gap produced by not having good figures on the sizable Jewish immigration to the United States from Eastern Europe and Israel since 1957. Prior arguments over the number of U.S. Jews have relied too much on projections of Jewish birth and death rates from the 1957 Current Population Survey while ignoring such immigration figures.

ACKNOWLEDGMENTS

The author thanks the Council of Jewish Federations and Welfare Funds for permission to use and publish these survey data which he spent so much volunteer time helping to gather. He also wishes to acknowledge the generosity of the computer installations of the University of Missouri and Bar-Ilan University in pro-

Table 1.—High, Middle, and Low Estimates of the U.S. Jewish Population

Source	High Estimate	Middle Estimate	Low Estimate
Sample data	5,555,186	5,555,186	5,555,186
From housing units whose religions could not be ascertained	172,400	86,200	0
Area sample bias of psu group 4	125,812	62,906	0
List and area sample bias of psu group 5	148,921	74,461	0
Total	6,002,319	5,778,753	5,555,186

viding so much free computer time. Finally, credit must be given to Mr. Patrick Corkery, Mr. Randy Meyer, Miss Dahlia Rachman, and Miss Sara Cohen for their outstanding research assistance together with all the clerical assistants who did so much excellent detailed sampling work.

REFERENCES

Frankel, Martin. 1971. Inference From Survey Samples: An Empirical Investigation. Ann Arbor: Institute for Social Research, University of Michigan.

Kish, Leslie. 1965. Survey Sampling. New York: John Wiley and Sons.

———, and I. Hess. 1959. On Variances of Ratios and Their Differences in Multi-Stage Samples. Journal of the American Statistical Association 54:416–446.

———, and M. Frankel. 1970. Balanced Repeated Replications for Standard Errors. Journal of the American Statistical Association 65:1070–1094.

———, M. Frankel, and N. Van Eck. 1972. Sampling Error Program Package. Ann Arbor: Institute for Social Research, University of Michigan.

Lazerwitz, Bernard. 1973. The Sample Design of the National Jewish Population Survey. New York: Council of Jewish Federations and Welfare Funds.