How Many Jews?

Synthesizing Data to Understand American Jewry

by LEONARD SAXE

Developing reliable estimates of the size and characteristics of the American Jewish population has long vexed social researchers. Although we know that 23 Jewish settlers arrived in New Amsterdam (New York) in 1654, ever since then our knowledge of the size of the population has been fraught with error. Part of the problem is the complexity of Jewish identity, but the difficulty is also a function of the methodological intricacies inherent in estimating and understanding "rare" populations.

Jewishness is both an ethnicity and a religion — it's both something one is born into as well as an identity that one has to accept. Many Americans — more than ten million — are Jewish by background and could request citizenship in the State of Israel under the Law of Return. But a substantial number of those who are eligible to be part of the Jewish people do not claim Jewish identity. Recent estimates of the size of the Jewish population in the United States have ranged from around four million to just above six million.

Because the United States Census does not collect information on religion, there are no official estimates. Instead, the Jewish community increasingly has relied on its own surveys. In the last 30 years, dozens of large-scale population studies of American Jewry have been conducted using "random digit dial" (RDD) techniques in which random phone numbers are called and adult members who answer are asked about their religious identity. The largest of these studies are the decennial National Jewish Population Surveys (NJPS) conducted by the federation movement (United Jewish Communities).

Unfortunately, RDD techniques are increasingly difficult to execute properly — response rates have declined dramatically and, as a consequence of the lack of census data, it's not possible to know who is missed. We suspect that Jews, because of socio-economic differences with the population at large, may be more difficult to reach than others. The most recent NJPS (2000-01) produced highly problematic findings, showing, in part, a dramatic reduction in the size of the Jewish population.

Although knowing the size of the Jewish population is not the most critical issue facing the Jewish community, understanding the size and character of the population is nevertheless important. It provides the denominator to assess the extent to which our programs and efforts to engage the community penetrate the population. To the extent that our data are linked to socio-demographic characteristics, it also permits us to understand whom we are impacting and whom we are failing to engage.

The Steinhardt Social Research Institute (SSRI) is developing a new paradigm to understand the demography of American Jewry. Our approach is simple in concept: Instead of doing single, relatively small but very expensive RDD surveys, our focus is to combine and synthesize data from multiple surveys which include questions about religious and/or ethnic identity. Despite the prohibition on collection of religious data by the U.S. Census, the U.S. government (along with major research organizations and foundations) routinely collects such information. By combining data from multiple investigations — most of which contain samples that include only a small number of Jews — one can develop reliable estimates of the overall population.

Leonard Saxe is Professor of Social Policy and Jewish Community Research at Brandeis University and Director of the Steinhardt Social Research Institute.

Many government-sponsored surveys — concerning overall social trends, health care, education, retirement and social welfare issues — include questions about religious and ethnic identity. These surveys often include the same question: "What is your religion? Is it Protestant, Catholic, Jewish or other? In most cases, the data from these studies (with individual respondent identities concealed) are available publicly.

The government invests substantial resources in the conduct of social research and the country's top research centers compete to conduct the studies.

What is clear from our preliminary analysis is that "data mining," combined with synthesis techniques, has the potential to provide valuable information about the size and character of the American Jewish population.

A search of key databases uncovered more than 125 surveys with questions about religion and ethnic identity. The studies range from the "gold standard" National Science Foundation-supported General Social Survey (GSS) to the National Election Study to the Health and Retirement Survey. Also included are surveys specific to religion (e.g., "Religion and Public Life Survey") and various surveys conducted by the Jewish community, including NJPS 2000-01, the American Religious Identification Survey (ARIS) and the Survey of Heritage of Ethnic Identity.

Conceptually, what we are attempting in developing a new paradigm is straightforward: We are simply combining the results of multiple studies. Methodologically, however, the process is exceedingly complex. We need to take account of the characteristics of each survey/sample and we need to understand exactly how the data were collected and weighted. In some cases (e.g., ARIS), the actual data are not available and we need to rely solely on the authors' summary.

The goal of SSRI's synthesis initiative is to produce the equivalent of census data, including data that enables us to do trend analyses over time (back to 1990) of changes in the Jewish population. To develop the analytic procedures, we have focused on the period 2000-2003. Two major Jewish-focused studies were done during this period (ARIS, NJPS), and they will enable us to understand the extent of error in these estimates.

Analytically, we have tested two approaches to combining data across the initial sample of surveys. The first, traditional meta-analysis, involves developing prevalence estimates, based on the weighted estimates of the percentage of respondents who identified as Jewish in each survey. But there is substantial variability across surveys and, in order to understand such differences, we are also using a statistical procedure called multilevel modeling to assess the role of individual and survey-level characteristics. Individual level data consisted of religious identity and demographic characteristics such as sex, age, race and education. We are also testing the impact of study characteristics such as the type of questions used to assess religion, survey purpose and response rate.

Data from the first 22 surveys (excluding NJPS) have been combined and, depending on technique, we have obtained estimates of the adult population between 3.3 and 3.8 million individuals. These are higher than the estimates provided by NJPS, although the variance is substantial. As we add additional studies to the database, we will be able to develop more precise estimates and will be able to understand the factors associated with differences across studies.

What is clear from our preliminary analysis is that "data mining," combined with synthesis techniques, has the potential to provide valuable information about the size and character of the American Jewish population. Potentially, it can transform how we understand the Jewish community. It offers the promise of providing, for a relatively modest investment, reliable estimates of the population and an understanding of how the population is changing over time.

Initial analysis suggests that current data — from studies such as NJPS and ARIS — have underestimated the population. One result is that the denominator for assessing the degree to which communal education and outreach programs and efforts at outreach are successful is too small. This means that we are overestimating the success of our efforts. Thus, for example, NJPS estimated that nearly 30 percent of school-age children receive a day school education. By other data, we know that figure is incorrect.

The issues are not simply "who is a Jew" and how we ask questions about Jewishness. The issue is a more basic methodological problem; it reflects the difficulty of identifying Jews by techniques such as RDD. The costs of doing RDD surveys that obtain adequate response rates may simply be too high.

The synthesis paradigm under development at the SSRI should be seen neither as a panacea nor as the comprehensive solution to information needs of the Jewish community. It is, in fact, only a start. For one, most of the current data focus on religious identity and do not include detailed questions about other ways in which Jews may identify. The surveys also focus on adults and rarely have detailed questions about how children are raised.

Nevertheless, extant data available for synthesis can help us create a broad portrait of American Jewry. Thus, for example, the vast majority of Jewish respondents appear to understand the question about religion as being a broad question about their identity. Even more important, the goal of our effort is to provide the basis to do more detailed and focused studies of segments of the national population and of specific subpopulations.

The Book of *Kohelet* (Ecclesiastes) teaches that a two-stranded cord is stronger than a single-stranded one and that a three-stranded cord is even stronger. Synthesizing data from multiple studies is a far more powerful method of understanding than merely looking at the results of a single study. As we work to understand and foster modern Jewish life, we should allow these studies to direct our efforts and help us to create a vibrant community.