

THE SIP GENERAL SURVEY SAMPLE

Barbara A. Anderson
Brian D. Silver

Working Paper #21
July 1986

Data for this study were produced by the Soviet Interview Project. This project was supported by Contract No. 701 from the National Council for Soviet and East European Research to the University of Illinois Urbana-Champaign, James R. Millar, Principal Investigator. The analysis and interpretations in this study are those of the author, not necessarily of the sponsors.

Funding for the research which resulted in this report was provided by the U. S. Department of State which, however, is not responsible for its contents or findings.

Copyright© Soviet Interview Project

THE SIP GENERAL SURVEY SAMPLE

BARBARA A. ANDERSON
Department of Sociology
University of Michigan

BRIAN D. SILVER
Department of Political Science
Michigan State University

July 1986

For Publication in
Politics, Work, and Daily Life in the USSR:
A Survey of Former Soviet Citizens
James R. Millar, Editor

THE SIP GENERAL SURVEY SAMPLE*

The respondents to the General Survey of the Soviet Interview Project are former Soviet citizens who emigrated to the United States.¹ Surveys of Soviet emigrants have relied on a variety of sampling techniques. Some have used snowball samples in which early respondents to the survey help to recruit later respondents. Others have used quota samples in which *a priori* target numbers of respondents with certain specified combinations of characteristics are established, and the sampling stops when the targeted number of interviews is completed.

The SIP General Survey I used a *stratified random sample*, based on the characteristics of the emigrants when they lived in the Soviet Union. Individual respondents were selected from a list that contained information about all eligible persons, defined by explicit eligibility criteria. The probability that given individuals were selected depended on the educational, regional, nationality, and city-size strata in which they fell. An effort was made to complete an interview with every selected individual. This method of sampling is less susceptible to self-selection by the respondents into the survey than snowball sampling or quota sampling, and it permits greater control over sample composition.

This chapter describes how the SIP General Survey I sampling frame and sample were defined. It analyzes the response rates and describes the basic demographic characteristics of the sample. And it discusses the issue of representativeness of the

* We would like to thank Mike Coble and Amy Hsu for the graphic work, Cynthia Buckley and Victoria Velkoff for research assistance, and Robert Lewis and Michael Swafford for helpful advice.

¹ In this appendix, we describe the sample for the *first* SIP General Survey. A follow-up survey, based on Soviet emigrants who arrived in the United States between May 1, 1982, and December 31, 1985, has also been conducted. When it seems necessary to avoid confusion, we shall refer to the first SIP General Survey as General Survey I.

respondents -- to what referent Soviet population can the results of the survey be generalized?²

The Sampling Frame

Designing a sample for the General Survey required the specification of a sampling frame: the set of emigrants from which the sample of prospective survey respondents was to be drawn.³ The sampling frame was defined as all Soviet emigrants who arrived in the United States between January 1, 1979, and April 30, 1982, and who were between ages 21 and 70, inclusive, at date of arrival. This range of dates of arrival includes the peak emigration year of 1979.

Only recent emigrants were included in the frame in order to minimize problems of recall and because the main purpose of the survey was to study *Soviet* life, not the processes of emigration or adjustment to life in the United States.⁴ In addition, most questions in the survey focused on the respondents' "last period of normal life in the USSR," a period that ended from a few months to several years before their arrival in the United States.⁵ On average, the month of arrival of the actual survey respondents was March, 1980, and the end of their "last normal period of life in the USSR" was December, 1978, a difference of fifteen months. Because the field work for the survey took place in 1983, with May 1983 the "average" month, the average length of time

² Several aspects of the question of representativeness and bias are addressed in Millar (1983), Silver (1985), and Bahry (1985, 1986).

³ The sampling frame for the SIP General Survey was also used for developing samples for the specialized or "S" projects, which involved interviews of people with special experiences, such as in economic planning, local government administration, and law.

⁴ The survey did contain a large number of questions concerning emigration experience and immigrant adjustment. These were designed primarily as controls for potential response bias.

⁵ The "last normal period of life in the USSR" was defined operationally in the survey as the five years preceding the major disruption in their lives associated with the decision to emigrate. For most respondents, this disruption was the act of applying for permission to emigrate.

between the end of the "last normal period of life" and the interview was fifty-three months.⁶

To develop the frame, we constructed a list of nearly all adult Soviet emigrants to the United States who arrived during the appropriate period. The list was based on information obtained from family service organizations in the United States, and included an abstract of basic biographical information on each individual: date of birth, country of birth, date of arrival in the U.S., sex, nationality-religion, education in the USSR, occupation in the USSR, city of last residence in the USSR, and military service and military rank in the USSR.

Biographical abstracts were completed for 37,156 individuals, of whom 33,618 met the final eligibility criteria for General Survey I.⁷ The biographical abstract data were important not only for sampling but also for defining the make-up of the emigrant population itself. The information the emigrants could give us about the parent population from which they came depended in part on the mix of backgrounds and experiences of the emigrants. Although the predominant ethnic-religious makeup of the respondents, their overall high levels of educational attainment, and their origination predominantly from the European parts of the Soviet Union was known in advance, the number of emigrants with specific combinations of characteristics, such as young persons with less than secondary education or non-Jews from small cities, was not known.

A large size for the sampling frame was desirable to increase the possibility of including in the sample respondents whose backgrounds were relatively rare among the

⁶ Despite this time lag, the respondents appear to have had excellent recall of life-history events. For discussion, see Anderson and Silver (1986b).

⁷ Armenian emigrants from the USSR to the USA were excluded from the sampling frame because 60 percent of the Armenians on whom biographical information was gathered were not born in the USSR but instead were individuals who repatriated to the USSR after World War II (primarily from Middle Eastern and Mediterranean countries), and most of the other Armenians were members of their families. Thus, it seemed likely that much of their *Soviet* experience would not be typical even of most Soviet Armenians.

emigrants as a whole. This would increase the diversity of any sample that could be drawn. Also, what *part* of the Soviet population the emigrants could represent depended on the characteristics of the individuals in the frame.

Characteristics of the Sampling Frame Population. We shall now describe the characteristics of the sampling frame and make some comparisons between the sampling frame, the sample, the respondents, and the Soviet population.

Column 1 in Panel A of Table 1 reports the number of eligible persons by year of arrival in the United States. Normal sampling error and the use of sample stratification criteria that intentionally favored the selection of individuals with particular backgrounds led to differences between the characteristics of people in the sample and people in the frame. The distribution by year of immigration for the sample (column 2) and for the General Survey I respondents (column 3) is less concentrated in the peak emigration year of 1979 than it is for people in the sampling frame (column 1).

The distribution of the frame population by age at arrival in the United States is presented in Column 1 of Panel B of Table 1. The overwhelming majority of persons in the frame, and hence also in the sample, had completed their education, and a large majority had considerable employment experience by the time they left the USSR. The distribution by age in the sample and among the actual survey respondents is similar to that for the sampling frame.

Ninety-nine percent of the eligible population was born in the USSR or in territories, such as the Baltic states, that are currently part of the USSR. Of the 168 people reported as born outside the USSR, 114 reported that they were born in Rumania or Poland; it is likely that most of these 114 also were born in parts of Rumania or Poland that were subsequently annexed to the USSR.⁸

The distribution of the sampling frame population by sex (Panel C of Table 1) reflects the numerical superiority of women over men in Soviet society, a product of differential war losses and the higher rates of mortality for men than for women during

⁸ For further information, see Anderson and Silver (1986a).

peacetime.⁹ The distribution in the sampling frame also reflects the fact that recent emigrants from the Soviet Union have primarily come as members of families. Seventy-eight percent of the respondents were married in their last period of normal life in the USSR. Of those who were married, 97 percent emigrated with their spouse, and 91 percent of the married couples emigrated with one or more of their children.¹⁰ Eighty-nine percent of the SIP General Survey I respondents emigrated with either their spouse, their children, or their spouse and children. Of the 2,389 respondents who had children at the end of their last normal period in the USSR, 84 percent emigrated with *all* of their children, and only 4 percent emigrated with *none* of their children.

The emigrants came overwhelmingly from cities. The urban origin of the emigrant population is not surprising. At the time of the 1979 Soviet census, 99 percent of Soviet Jews, 74 percent of Soviet Russians, and 62 percent of the entire Soviet population lived in urban areas.¹¹ Furthermore, compared to the Soviet urban population, the emigrants come primarily from large and medium-sized cities (see column 1 in Panel D of Table 1). Ninety-seven percent come from cities that had populations of 100,000 or more in 1979. By comparison, only 38 percent of the entire Soviet population, and 60 percent of the Soviet urban population, lived in cities of 100,000 population or more in 1979.

Moreover, 88 percent of the emigrants in the sampling frame came from Soviet

⁹ For further discussion, see Anderson and Silver (1986c).

¹⁰ The few exceptions when the spouse did not emigrate with the respondent are divided roughly evenly into four categories: 1) spouse was denied an exit permit (or held a sensitive job); 2) spouse stayed with relatives; 3) spouse was too ill to emigrate; 4) spouse "did not want to go."

¹¹ Figures that we cite for the Soviet population in 1979 are based on the 1979 Soviet census. We either derive them directly or calculate them from data published in USSR, TsSU (1984).

cities that had populations of 500,000 or more in 1979.¹² In contrast, only about 20 percent of the total Soviet population, and about 32 percent of the urban Soviet population, lived in cities of 500,000 population or more in 1979.

The big-city origins of the emigrants does not mean that only the Soviet population from large cities can be represented in the survey. There were enough people in the sampling frame from medium-sized cities to permit purposive oversampling of people from medium-sized cities. The proportion of respondents from medium-sized cities is approximately twice as large as the proportion of persons in the sampling frame who emigrated from medium-sized cities (see Panel D of Table 1).

The emigrants in the sampling frame came primarily from the European parts of the USSR, especially the Soviet West (Ukraine, Belorussia, and Moldavia) and the Russian Republic (RSFSR) (see Panel E of Table 1). For purposes of sample design, however, there was a sufficient number of people in each of five major Soviet regions (groups of republics) to approximate in the sample the distribution by region of the Soviet population that lived in cities with populations of 100,000 or more (column 6 of Panel E).

Accordingly, the sample was designed so that the proportion of the people in sample who originated in the RSFSR would be almost twice as large as the proportion of people in the sampling frame who originated in that republic. The proportion from the Soviet West was reduced correspondingly to about half the proportion of the sampling frame that had come from that region (compare columns 1, 2, and 3 of Panel E of Table 1). Also, to assure adequate regional diversity in the sample, minimum target sample sizes were established for the Baltic and Transcaucasia.

Emigrants from the RSFSR and the West came predominantly from a few cities. The seven cities providing the largest numbers are Kiev (7,384), Odessa (4,881), Moscow (3,781), Leningrad (3,760), Minsk (2,133), L'vov (1,493), and Kishinev (1,286). Those

¹² For purposes of sampling, we included the four republic capitals (Ashkhabad, Dushanbe, Tallinn, and Vilnius) that were less than 500,000 in population in 1979 with the cities of 500,000 or more. See Anderson, Silver, and Lewis (1986).

who came from Central Asia, the Transcaucasus, and the Baltic, came overwhelmingly from the largest cities in those regions -- especially the republic capital cities of Riga (1,328), Tashkent (991), Baku (547), Tbilisi (348), Vilnius (240), and Dushanbe (90) (see Anderson and Silver, 1986a).

Seventy-five percent of the survey *respondents* came from the republic capitals (including Leningrad). In every region except the Soviet West, over 80 percent of the respondents came from republic capitals. In the West, only 50 percent of the respondents came from republic capitals (Minsk, Kiev, and Kishinev).

Panel F of Table 1 summarizes the data on educational attainment.¹³ Compared to the Soviet urban population, the emigrants as a whole are highly educated, in line with the high average educational levels attained by Soviet Jews. They also have a somewhat higher average educational level than the Soviet population residing in large cities. Forty-four percent of the people in the sampling frame had achieved at least some higher education (column 1). In contrast, in the adult Soviet population in the republic capital cities taken together in 1979, less than 30 percent had attained that level of education. The sample was designed to select people from the frame in proportions that approximated the estimated distribution by education of the Soviet adult population in large cities in 1979.

Recent Soviet emigrants are primarily Jews or members of families that included Jews. As is shown in Panel G of Table 1, 98.4 percent of all people in the sampling frame were Jews. To maximize the ethnic diversity of the sample, all known non-Jews were included in the sample, so that about 85 percent of the emigrants in the sample were Jews.

¹³ For further details, see Anderson and Silver (1986a).

The Referent Soviet Population

Building the sampling frame was one step in identifying a pool of potential survey respondents whose life histories would shed light on Soviet experience. At the same time, the characteristics of the pool of emigrants determined which segment of Soviet society *could* be represented in the sample. For example, since there are almost no people from rural areas in the sampling frame, it is not possible to draw a sample from the emigrants that represents the experiences of the rural sector of Soviet society. Similarly, no sample drawn from this sampling frame could represent the experiences of most of the major non-Russian nationalities, especially the Moslem nationalities.

Recent emigrants are diverse, however, with respect to education, occupational experience, and geographic origins in the USSR, and most of their everyday experiences in the Soviet Union preceding the traumatizing experiences associated with emigration are likely to be reflective of the experiences of an important sector of Soviet society. By using the information about individuals in the sampling frame to select a sample that maximized the diversity of backgrounds of the survey respondents, the sample could approximate some aspects of the demographic composition of the "adult European population in large and medium-sized Soviet cities." We term this the *referent Soviet population*.

The main purpose of identifying the referent population was to clarify the parts of the Soviet population that the survey respondents could *not* represent, and to identify a sector of Soviet society that the survey respondents *could* represent if the respondents were appropriately selected from the frame and if the survey instrument provided information to test for various forms of response bias.¹⁴ The concept of a referent population was thus a guide for the sample design and for interpretation of the survey results, not an exact blueprint to be executed in the sample.

¹⁴ The main sources of response bias with which we were concerned were the effects of emigrant selection and experience, that is, the fact that most respondents were Jews, and the accuracy of recall.

The Sample

Size. The initial sample size was set at 3,750 under the expectation that 80 percent of the individuals in the sample would complete the interviews, yielding 3,000 completed interviews, or respondents. The target of 3,000 respondents was established so that each of the three survey supplements would have 1,000 respondents.¹⁵

Sampling Procedure. The SIP General Survey sample was designed to approximate the educational and regional composition of the referent Soviet population. It was also designed to diversify the sample on the basis of nationality and size of city compared to the distribution in the frame, but not to approximate those distributions in the referent Soviet population.

To accomplish these goals, the sample was stratified. Although each person in the frame had a known probability of being selected into the sample, the probability varied with the individual's nationality, education, size-of-city, and region of origin within the USSR. First, targets were set for the overall educational, regional, and city-size distributions. None of these was modelled to match the referent Soviet population exactly, but they were made to be much more like the referent Soviet population than like the frame population. Second, all eligible non-Jews were selected into the sample.¹⁶ Third, an iterative, random-selection procedure was used to draw the remainder of the sample (the Jews) so that target distributions by education, city-

¹⁵ On the structure of the survey instrument, see Millar's introductory chapter to this volume. All respondents completed a common set of "Core" questions; respondents were then assigned randomly to receive one of the three variant "Supplements" so that about one-third of the respondents completed each of the supplements. The actual numbers completing the three supplements, which were designated by the color of their face-sheets as orange, blue, and green, were 926, 933, and 922, respectively. Twelve respondents completed no supplement.

¹⁶ The information about the nationality of the persons in the sampling frame was not complete. Based on the data obtained in the survey itself, the nationality of the actual respondents could be determined with greater precision.

size, and region were achieved.¹⁷ The distributions of all these variables in the sample are shown in column 2 of Table 1.

Had the General Survey I sample been a simple random sample of eligible individuals (the frame), it would have differed much more sharply from the referent Soviet population. By stratifying the sample, persons in the frame whose educational level was "completed secondary education or less" were more likely to be chosen than persons who had attained higher education. Persons in the frame from medium-sized Soviet cities were more likely to be selected than persons from large cities. And persons from the RSFSR, the Baltic, and Transcaucasia were more likely to be chosen than persons from the Soviet West (Belorussia, Ukraine, Moldavia) or Central Asia.¹⁸ Minimum sample sizes were established for the Transcaucasus and the Baltic --exceeding their relative proportions in the referent Soviet population -- to permit multivariate analyses based on the individuals from each of these regions.

Modelling the composition of the sample on the demographic composition of the referent Soviet population reduced the unrepresentativeness of the sample. In two respects, however, no sample of recent Soviet emigrants could match the referent Soviet population. First, any sizeable sample had to consist mostly of Jews. Second, for obvious reasons, all persons in the sample were emigrants.

All eligible non-Jews were included in the sample not in order to mimic the referent Soviet population but rather to provide a comparison or control group for assessing the effect of ethnic differences on patterns of survey responses. A similar rationale applies to the effort to increase the number of respondents from medium-sized cities. This permits researchers to test for the effects of city size on response patterns, particularly reports of economic behavior.

The main control for bias linked to the self-selection or to special experiences of the respondents as emigrants was in the design of the questionnaire, not of the sample.

¹⁷ See NORC (1985): Appendix E.

¹⁸ In addition, *within* the West, a maximum of 100 persons was to be selected from Odessa.

For this purpose, a series of questions was included concerning the respondents' motivation for emigration, their role in the decision to emigrate, and their adjustment to life in the United States.¹⁹

Response Rates. The final General Survey I sample was comprised of 3,738 individuals selected from the sampling frame.²⁰ Of these, 187 were subsequently dropped for one of three reasons: a) they were deceased; b) they were too ill to participate in the survey; or c) they were no longer residing in the United States. Because these people did not refuse to participate in the survey, we interpret them as "ineligibles" rather than as "refusals." Individuals whose addresses were never confirmed are treated as eligible, since some of these individuals may have actively avoided participation in the survey by not responding to letters of inquiry or other efforts by the interviewers or the National Opinion Research Center to contact them.²¹

Of the 3,551 persons remaining in the sample, 2,793 completed the interview, for a response rate of 79 percent. This rate compares favorably with that in most other sample surveys conducted in this country.

Participation in the survey was voluntary, and respondents were assured that both their answers and their participation in the survey would be confidential.²² Of those who completed the interview, 221 participated only after initially stating that they did not want to participate or after they did not respond to initial inquiries.

¹⁹ Analysis of the General Survey I data indicates that responses to questions related to religious behavior and to perceptions of discrimination are very sensitive to the ethnic or religious affiliation of the respondent, but responses to questions dealing with most other issues are not sensitive to the respondent's ethnic or religious background. See Bahry (1985).

²⁰ The initial size of 3,750 was reduced to 3,738 when it was discovered that 12 "ineligible" persons had inadvertently been included, before any contacts were made with potential respondents.

²¹ The effort to obtain current addresses for persons in the sample began only after the sample was drawn. It would have been wasteful and prohibitively expensive to gather this information for all 33,618 persons in the sampling frame. Of the 3,738 persons in the final sample, 91 could not be located.

²² For a description of the steps taken to assure confidentiality, see NORC (1985): 40-42 and Millar's introductory essay to this volume.

Of the 758 persons in the sample (of 3,551) who did not participate in the survey, 647 either "refused" to participate or broke off the interview before completing it. Another 91 persons could not be located. And 20 were not interviewed for some other reason.

Table 2 shows the response rates for various groups of people. There was no difference in the response rates of Jews and non-Jews. As is true of many surveys, persons with higher education were more likely to agree to participate in the SIP General Survey than the less educated. Although younger people were slightly more likely to complete the interview than older people, the difference in the response rate associated with education is not a function of age. Instead, as is shown in Figure 1, more highly educated respondents had higher response rates than less educated respondents in each age group.

Figure 1 also reveals that the differences in the response rates associated with age are negligible, once differences in education are taken into account. The only sharp deviation is among persons age 21-30 who had less than complete secondary education, but only 37 persons in the sample (21 of whom completed the interview) who were in this category.

Men in the sample were more likely to complete the interview than women (see Table 2). This difference is not a function of the difference in educational attainment of men and women, for at each educational level men were more likely to complete the interview than women.

In summary, the differences in response rates among educational groups had more of an impact on the composition of the final respondents than differences by age, sex, or nationality. The differential by education moved the composition of the respondents more toward that of the sampling frame, and away from that of the referent Soviet population (compare columns 1-3 in Table 1). Overall, however, the response rates did not vary greatly with social background.

Weighting the Cases

Purpose. Weights are used in statistical analyses so that the weighted respondents will resemble more closely the population to which the researcher hopes to generalize the results than would the unweighted respondents. When a simple random sample is drawn from a population of interest, weights are generally not necessary. Given normal sampling error, the characteristics of the respondents will be identical to those of the population from which the sample is drawn. If the sample is disproportionate, so that individuals in the population do not have an identical probability of being selected into the sample, the characteristics of the unweighted set of respondents will not match those of the population of interest. The use of weights has the effect of counting some cases more heavily than others in the analysis, thus compensating for the initial disproportionate sampling.²³

Stratifying the sample drawn from the list of eligible emigrants helped to bring the characteristics of the sample more into line with those of the referent Soviet population than would have been true of a simple random sample drawn from that list. For several reasons, however, further adjustments to the composition of the respondents are necessary to make it more similar to the referent population. First, as discussed earlier, the stratification procedure did not bring the sample completely into line with the referent population. Second, information about respondent backgrounds that was known in advance of the survey was less accurate and less complete than information obtained in the survey itself, particularly regarding the respondents' educational attainment. Third, how closely the actual respondents would match the characteristics of the referent population depended on how the response rates varied among different groups of respondents.

²³ For readers who are not familiar with how weights are applied in practice, it may be useful to note that major statistical software programs, such as SPSS^X, have built-in routines that automatically weight the cases at the user's option. The user need only designate the name of the variable that is to be used to weight the cases.

Method. Information on the respondents' Region (five categories), education (three categories), and age (five categories) was used in devising the weights. As a preliminary step in constructing the weights, we estimated the three-way distribution of age-by-education-by-region in the referent Soviet population -- to define the appropriate share that each of the resulting 75 population categories should represent among the weighted survey respondents.

The most formidable problem in estimating the education-by-age distribution of the referent Soviet population is that neither age distributions nor education distributions by age have been published for the most recent Soviet census year, 1979. This census date corresponds most closely to the date of the "last normal period of life" of respondents to the first SIP General Survey. Therefore, we estimated the distributions indirectly using an iterative fitting procedure from 1970 and 1979 Soviet census data, based on the characteristics of the populations in republic capital cities.²⁴

The three-way crosstabulation of the region, education, and age variables defines 75 population categories to which weights were assigned. Each category can be represented as a proportion of the total population -- so that the sum of the proportions across all 75 categories is 1.000. The weight assigned to survey respondents in any given cell is calculated as the proportion of the referent Soviet population in that cell divided by the proportion of the respondents in that cell.

Thus, if the proportion of respondents in that cell is smaller than the proportion of the referent population in that cell, the weight assigned to respondents in that cell will be greater than unity, thus causing respondents in that cell to count more heavily than they would otherwise. If the proportion of respondents in the cell is larger than the proportion of the referent population in that cell, the weight assigned will be less than 1.00 -- to reduce the relative contribution of those respondents to the overall distributions. The actual weights for the first SIP General Survey vary from 6.28 to

²⁴ A detailed discussion of the method of development of demographic estimates for the weights is given in Anderson, Silver, and Lewis (1986a).

0.22. Thus, if analysts choose to use the weights, a respondent with a weight of 6.0 would "count" as six respondents; a respondent with a weight of .25 would count as one-fourth of a respondent.

In some of the cells of the age-by-education-by-region distribution, there were very few respondents. No respondents from the Baltic and only one respondent from the RSFSR, for example, fell into the age range 21-30 at date of arrival and had less than complete secondary education. To avoid assigning extraordinarily high weights to young persons with less than complete secondary education for some regions, we collapsed the cells across regions for persons who had less than complete secondary education for each of the three age categories 21-30, 31-40, and 41-50. Had we not done this, no weight could have been assigned to respondents age 21-30 with less than complete secondary education from the Baltic, and a weight of 57 would be needed for the one respondent from the RSFSR who fell into that age-by-education cell.

The main consequence of collapsing across regions -- for those with less than completed secondary education and who were age 50 or under at date of arrival -- is that the regional distribution of the weighted respondents does not match the estimated regional distribution of the referent Soviet population. This is shown in the distributions in columns 5 and 6 in Panel E of Table 1. But the target distributions for education and age separately as well as in combination are matched exactly (see Panels B and F).

Average weights for respondents by age and education are shown in Table 3.²⁵ Cases that are weighted most heavily are younger persons with less than secondary education. Accordingly (as shown in Table 3), the 15 actual respondents who were age 21-30 at date of arrival in the U.S. and had less than completed secondary education would count as 94 respondents if the data were weighted, and the 84 respondents who

²⁵ These are averages, because they do not reflect the differences in the weights related to the respondents' region of residence in their last period of normal life in the USSR.

were age 61-70 at date of arrival and had some higher education would count as 60 respondents in a weighted data analysis.

The *total* number of respondents (the *N*) is the same for the weighted cases and the unweighted cases. By multiplying the number of actual cases in each cell by the weight applied to each case, the total number of respondents for the weighted cases comes to 2,793.

Weights and the Referent Population. Most analyses of the SIP General Survey are not likely to use weighted data. This is because most statistical analyses will focus on the *relationships between variables*, rather than on either the overall frequency distributions or the "average" score or answer found among all respondents. When the focus is on the relationships between variables, whether one uses the weights will seldom affect analytic results.²⁶

If one is interested, for example, in how the level of support for the Soviet regime varies with the respondent's education, it does not matter whether 42 percent of the respondents had some higher education or 27 percent had some higher education (which is the adjustment in the proportion with some higher education that would result if one shifted from using unweighted data to using weighted data). If one were interested, however, in measuring the *average* level of support for the regime among *all* respondents, then using weighted data would increase the apparent overall level of regime support because respondents who have higher education are less supportive of regime norms than are respondents with secondary or lower education (see Silver, 1985).

Thus, the weights are an auxiliary tool that may be useful for some types of analysis of the General Survey data. But they are not mandatory for all analyses, particularly those which focus on the relationships between variables rather than on

²⁶ A major exception is when the analysis focuses directly on the relation between age and education -- two variables whose relationship is most severely adjusted in the weights -- or on the relation between age, education, and another variable that is correlated with both age and education, such as income. See Anderson (1986) for an analysis of the relation between age, education, and income among SIP General Survey respondents.

univariate distributions or measures of central tendency for the entire set of respondents.²⁷

The Question of Generalizability

The logic that applies in determining whether or not to weight the responses in analyses applies also to whether it is important that respondents exactly match the demographic characteristics of the referent Soviet population. The validity of any generalization from the survey to the referent Soviet population requires more than a mechanical matching of the socio-demographic characteristics of the respondents as a whole and the referent population.

It is more important to establish that survey respondents with specific socio-demographic backgrounds are similar to persons with the same background who did not emigrate from the USSR, or who were not Jewish. This is not just a sampling issue. Many researchers have compared the distributions on variables of interest in the SIP General Survey with analogous distributions for the Soviet population in official Soviet publications. When these distributions are similar, one can have greater confidence in the results of multivariate analyses using the SIP data.

Diversification of the sample, especially by nationality, coupled with the use of a stratified random sample based on a list of the eligible population, provides another basis for assessing the sensitivity of responses to potential bias. The concept of a referent Soviet population is relevant not because it represents the population from which the sample is drawn and against which the sampling error could be determined in

²⁷ The main effect of using the weights when one engages in multivariate analysis is on the amount of variance in the dependent and independent variables. Hence, whether one uses the weights will have a much greater effect when one employs correlation coefficients or standardized regression coefficients in statistical analyses than when one uses unstandardized coefficients. For most purposes it is probably preferable to use unstandardized coefficients for analyzing both the weighted and unweighted SIP data, because the amount of variance among the respondents is substantially affected by *a priori*, and inevitably somewhat arbitrary, decisions about the composition of the sample.

precise statistical terms. Rather, it is important because it provides a referent sector of Soviet society with whose experiences and behavior the SIP General Survey respondents are most likely to correspond.

References

- Anderson, Barbara A. 1986. The Life Course of Soviet Women Born 1905-1960. *Soviet Interview Project Working Papers*, No. 11.
- Barbara A. Anderson and Brian D. Silver. 1986a. Descriptive Statistics for the Sampling Frame Population. *Soviet Interview Project Working Papers*, No. 2, Revised.
- _____. 1986b. The Validity of Survey Responses: Interviews of Multiple Respondents in a Household from a Survey of Soviet Emigrants. *Soviet Interview Project Working Papers*, No. 14.
- _____. 1986c. Sex Differentials in Mortality in the Soviet Union: Regional Differences in Length of Working Life in Comparative Perspective. *Population Studies*, 40 (July): forthcoming.
- Barbara A. Anderson, Brian D. Silver, and Robert A. Lewis. 1986. Demographic Estimates for the Post-Sampling Weights of the SIP General Survey. *Soviet Interview Project Working Papers*, No. 4, Revised.
- Bahry, Donna. 1985. Surveying Soviet Emigrants: Political Attitudes and Ethnic Bias. Manuscript, Department of Politics, New York University.
- _____. 1986. Politics, Generations and Change in the USSR. *Soviet Interview Project Working Papers*, No. 20.
- Millar, James R. 1983. Emigrants as Sources of Information about the Mother Country: The Soviet Interview Project. *Soviet Interview Project Working Papers*, No. 5.
- NORC [National Opinion Research Center]. 1985. *Soviet Interview Project Methodological Report*.
- Silver, Brian D. 1985. Political Beliefs of the Soviet Citizen: Sources of Support for Regime Norms. *Soviet Interview Project Working Papers*, No. 6.
- USSR, TsSU [Tsentral'noe statisticheskoe upravlenie]. 1984. *Chislennost' i sostav naseleniia SSSR: Po dannym Vsesoiuznoi perepisi naseleniia 1979 goda*. Moscow: Finansy i statistika.

TABLE 1 -- Page 2

	Sampling Frame (N=33618) (1)	Final Sample (N=3551) (2)	Respondents [Frame Data] (N=2793) (3)	Respondents [Survey Data]		Referent Soviet Population Estimate (6)
				Unweighted (N=2793) (4)	Weighted (N=2793) (5)	
F. Education						
Some Higher	44.1%	33.8%	36.8%	41.9%	27.3%	27.3%
Complete Sec.	38.5	45.3	44.8	40.7	40.6	40.6
< Comp. Sec.	17.4	20.9	18.4	17.4	32.1	32.1
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
G. Nationality^f						
Jews	98.4%	85.7%	85.7%	82.8%	83.1%	
Non-Jews	1.6	14.3	14.3	17.2	16.9	
Total	100.0%	100.0%	100.0%	100.0%	100.0%	

^a Figures for the sampling frame and sample (columns 1 and 2) are derived from the "frame data" -- the biographical abstracts developed for sampling. Figures for the respondents in column 3 are also derived from the pre-survey biographical abstracts. Figures for the respondents in columns 4 and 5 are based on the SIP General Survey results, unless otherwise noted. Figures for the referent Soviet population (column 6) are derived from Soviet census data.

^b All arrivals in 1982 were in the first four months of the year.

^c Age in columns 1-5 is age at arrival in U.S. The age distribution in column 6 is as estimated for 1979. See Anderson, Silver, and Lewis (1986).

^d City sizes are based on the population in 1979. The largest size category includes republic capital cities even if they were less than 500,000 population. City size based on the frame data (columns 1-3) refers to size of city in which persons were last employed in the USSR. City size based on the General Survey data (columns 4 and 5) refers to the size of city in which persons lived at end of their last normal period of life in the USSR.

^e The region categories based on the frame data refer to the region where persons lived when last employed in the USSR. Region based on the General Survey results refers to the region in which persons lived at the end of their last period of normal life in the USSR. Republics included in the multi-republic regions: *West* (Belorussia, Moldavia, Ukraine); *Baltic* (Estonia, Latvia, Lithuania); *Transcaucasia* (Armenia, Azerbaidzhan, Georgia); *Central Asia* (Kazakhstan, Kirgizia, Tadzhikistan, Turkmenistan, Uzbekistan.) The figures in column 6 refer to the regional distribution of the Soviet population in cities of 100,000 population or more in 1979.

^f For columns 4 and 5, persons who were Jewish by self-identified nationality or religion are classified as Jewish; all others are classified as non-Jews. In both columns 4 and 5, if those who were children of Jews (but not self-identified as Jewish by nationality or religion) were counted as Jews, then 87.4% of the respondents would be Jews.

TABLE 2. Percentage of Persons in Sample Completing the Survey, by Education, Age, Sex, Nationality, and Size of City of Last Employment in USSR^a

	Percent Completing Survey	Base Number in Sample (N=3551)	Number Completing Survey (N=2793)
<u>Education</u>			
Some Higher	85.6	1200	1027
Complete Secondary	77.8	1609	1257
Less than Complete Secondary	69.3	742	514
<u>Age At Arrival in U.S.</u>			
21-30	79.1	764	604
31-40	80.1	898	719
41-50	82.6	720	595
51-60	73.9	590	436
61-70	75.8	579	439
<u>Sex</u>			
Men	80.3	1511	1213
Women	77.5	2040	1580
<u>Nationality</u>			
Jews	78.7	3042	2394
Non-Jews	78.4	509	399
<u>Size of City of Last Employment in USSR</u>			
500,000+	79.6	2868	2283
100,000-499,999	74.7	598	447
Less than 100,000	74.1	85	63

^a The characteristics used in this table are from the pre-survey sampling frame data, not the survey results.

TABLE 3. Average Weights Assigned by Educational Level and Age at Arrival in the United States^a

	Education			
	Some Higher	Complete Secondary	Less than Comp. Sec.	All
<u>Age at Arrival</u>				
21-30	.65	1.34	6.28	1.15
(Unweighted N)	(275)	(312)	(15)	(602)
(Weighted N)	(179) ^b	(417)	(94)	(690)
31-40	.69	.95	5.00	1.01
(Unweighted N)	(371)	(311)	(34)	(716)
(Weighted N)	(258)	(296)	(171)	(724)
41-50	.54	1.09	2.65	1.02
(Unweighted N)	(314)	(194)	(83)	(591)
(Weighted N)	(169)	(211)	(220)	(600)
51-60	.76	.75	1.46	.98
(Unweighted N)	(127)	(170)	(142)	(439)
(Weighted N)	(96)	(128)	(207)	(431)
61-70	.71	.54	.98	.78
(Unweighted N)	(84)	(150)	(211)	(445)
(Weighted N)	(60)	(82)	(206)	(347)
All	.65	1.00	1.85	1.000
(Unweighted N)	(1171)	(1137)	(485)	(2793)
(Weighted N)	(762)	(1135)	(897)	(2793)

^a The weights shown in this table are averages because the actual weights vary also according to the region in which the respondent resided in his or her "last period of normal life" in the USSR. See Anderson, Silver, and Lewis (1986) for the weights by age, education, and region.

^b Weighted N's are rounded to the nearest integer.

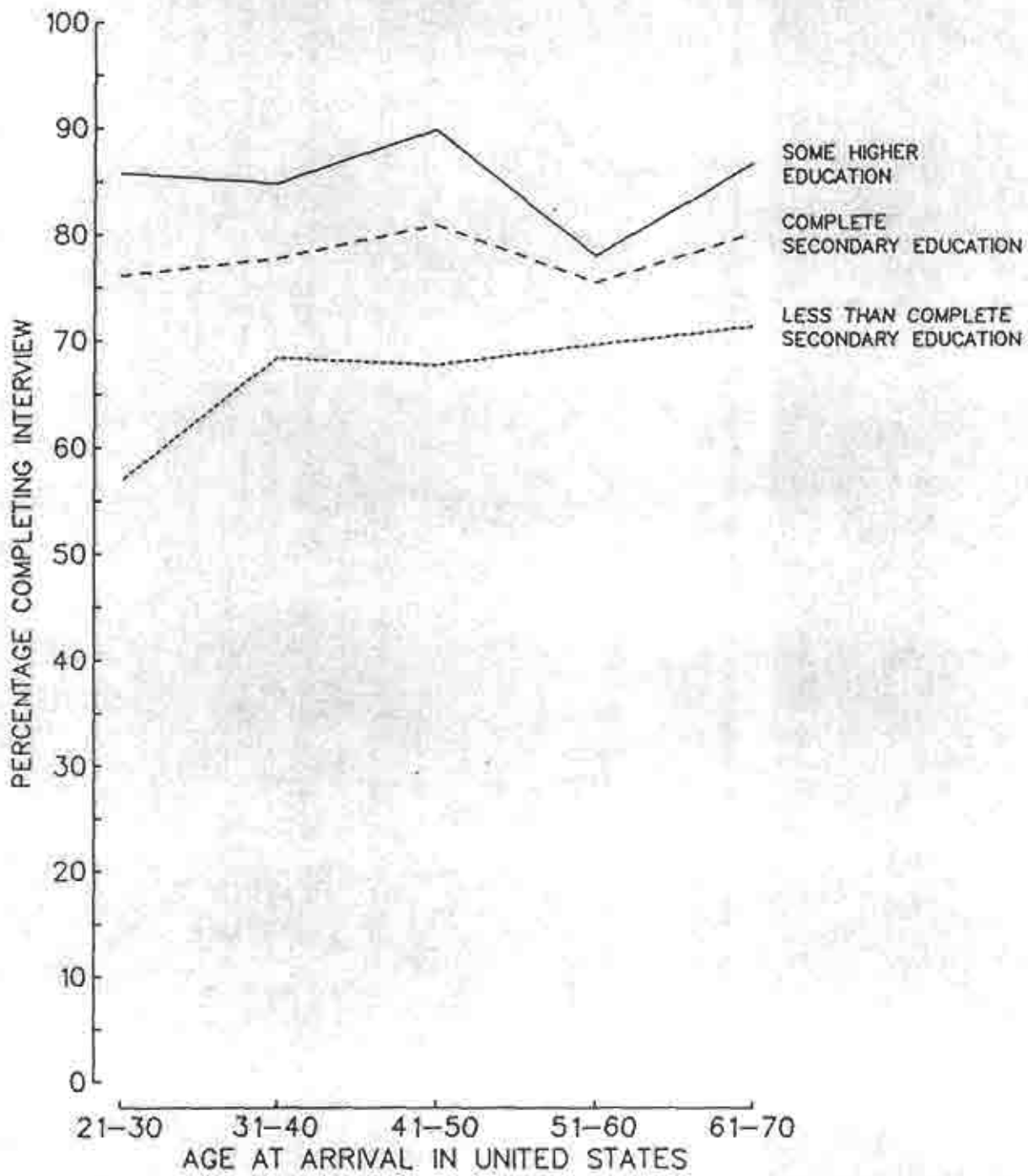


Figure 1. Percentage Completing General Survey by Education and Age at Arrival in the United States

Handwritten text, possibly bleed-through from the reverse side of the page, located along the right edge.