NLSY79

TECHNICAL SAMPLING REPORT (1983)

NLS User Services
614 442-7300
usersvc@postoffice.chrr.ohio-state.edu

National Longitudinal Survey of Labor Force Behavior, Youth Survey (NLS)

Technical Sampling Report

Martin R. Frankel
Harold A. McWilliams

and

Bruce D. Spencer

August, 1983

CONTENTS

1. INTRODUCTION..... 5

2. THE DESIGN OF THE SAMPLE..... 8

3. SELECTION AND IMPLEMENTATION OF THE SAMPLE.....17
Selection of the Cross Sample.....17
Stage I.....18
Stage II.....18
Stage III.....18
Stage IV.....19
Selection of the Supplemental Sample.....20
Stage I.....20
Stage II.....21
Stage III.....22
Stage IV.....23
Special Procedures Used in Both the Cross-Section and Supplemental Samples.....24
Procedures for Inclusion for Unlisted (Missed) Dwelling Units.....25
Procedures to Insure Coverage of the Non-DU Population (College Dormitories and Other Group Quarters) 27
Selection of the Military Sample.....30
Stage I.....31
Stage IIa.....33
Stage IIb.....33

4. WEIGHTING THE SAMPLE.....35
Weighting the Civilian Sample.....35
NLS Baseyear Weights.....35
Weights for Rounds II, III, IV.....44
Weighting the Military Sample.....49
NLS Baseyear Weights.....49
Weights for Rounds II, III, and IV.....50

5. RELIABILITY OF RESULTS.....53
Approximate Standard Errors: Percentages.....53
Approximate Standard Errors: Means.....61

6. POTENTIAL IMPACT OF NONRESPONSE.....63

CHAPTER 1

INTRODUCTION

The National Longitudinal Survey of Labor Force Behavior--Youth Cohort is an annual national survey of 12,686 young people throughout the United States. A one-hour personal interview is conducted each year between January and May. At the present time, six rounds of interviews have been conducted. This report describes the sample design and procedures for the first four rounds, covering 1979-82.

These surveys all have as their base the same 12,686 youth who completed the Round I baseline interview in 1979. In all there are three independent probability samples. Two of these samples were designed to cover the noninstitutionalized civilian population in the age range 14 to 21 as of January 1, 1979. The third sample was designed specifically to cover the 17 to 21 age cohort serving in the military as of January 1, 1979.

The first of the two civilian samples is a cross-sectional sample designed to yield the proper population proportions of various racial, ethnic, and income groups in the 14 to 21 age cohort. The second is a supplemental sample designed to oversample Hispanic, black, and economically disadvantaged non-black, non-Hispanic youth.

The military sample is a clustered probability sample of the 17 to 21 age cohort, stratified by branch of service (Army, Navy, Air Force, Marine Corps) and by geographic location (Eastern United States, Western United States, Europe, the Far East, and Other), in which women were oversampled at approximately six times the rate for men.

The questionnaire centers on the respondents' education, job training and work experiences, with additional questions on a variety of related issues that change from year to year. Other issues have included knowledge of the

world of work, self-esteem, deviant behavior, fertility and contraception, child care, drug use, and consumption of alcoholic beverages, plus a series of questions on the amount of time devoted to school, work, and leisure, and other activities.

Two additional studies have been undertaken to enhance this data set. The first is the collection of high school transcripts for the National Center for Research in Vocational Education at Ohio State University. This project was funded by the U.S. Department of Education. Complete four-year transcripts have been collected for over 8,000 members of the civilian samples. An additional effort is currently underway to collect transcripts of the youngest members of the sample who have graduated from high school.

The second additional study is the Profile of American Youth, sponsored by the U.S. Department of Defense. During the summer of 1980, NORC administered the Armed Services Vocational Aptitude Battery (ASVAB) to 11,914 members of the NLS sample. The ASVAB is used by the military services to screen applicants and assist in determining job placement. The results of testing the NLS samples are being used to create new national norms for the ASVAB as mandated by Congress. Extensive reports on the sampling and testing procedures, the demographics of the test results, and the psychometrics of the ASVAB will be available from the Department of Defense.

The survey is sponsored by the U.S. Departments of Labor and Defense under a grant to the Center for Human Resource Research at Ohio State University. NORC has a subcontract to provide data collection services. The youth cohort is a continuation of the National Longitudinal Surveys begun in 1965 by the Center for Human Resource Research for the Office of Manpower Policy, Evaluation and Research of the U.S. Department of Labor. A National Longitudinal Survey handbook describing the various cohorts, the data

collection schedules, the variables, the documentation, and so forth is available from the Center for Human Resource Research at the following address:

National Longitudinal Survey Users' Office
Center for Human Resource Research
5701 N. High Street
Worthington, OH 43085

Kenneth Wolpin is the principal investigator and project director for the National Longitudinal Surveys at the Center for Human Resource Research. The project director at NORC is Mary C. Burich.

CHAPTER 2

THE DESIGN OF THE SAMPLE

The NLS sample consists of American men and women born in 1957 through 1964. The procedures and methods used to select the sample are specifically designed to yield a database of youth that can be statistically projected (within known confidence levels) to represent the entire population born in 1957 through 1964, and substantively important subgroups within this population. With these objectives in mind, the design specifies the selection of 3 independent probability samples: (1) a cross-section sample designed to represent the noninstitutional civilian segment[1] of American young people aged 14 to 21 as of January 1, 1979, in their proper population proportions; (2) a supplemental sample designed to produce, in the most statistically efficient way, oversamples of civilian Hispanic, black, and economically disadvantaged non-Hispanic, non-black youth;[2] and (3) a military sample designed to represent the population aged 17 to 21 as of January 1, 1979, and serving in the military as of September 30, 1978.[3]

As Chapter 3 describes in detail, these three samples were selected by standard area probability sampling methods.[4] Stratification was introduced at

---

[1] Individuals were included in the noninstitutional civilian segment if they were not on active duty as of September 30, 1978, and not in institutions as of January 1, 1979.

[2] Oversampling was done on a strict probability basis so tha weighting procedures could be used to compensate for this targeted overrepresentation in a statistically appropriate way.

[3] Individuals who separated from the military between the time of their selection in 1978 and the time of the interview in 1979 were, nonetheless, retained in the military sample and were interviewed. The military sample thus represents the military services as of September 30, 1978.

[4] This report assumes basic familiarity with these sampling methods. Readers in need of additional information on area probability sampling should consult a standard survey sampling textbook.

several stages in the selection process and several special procedures were used to minimize losses due to noninclusion of college students or persons living in group quarters and to include persons living in dwelling units missed during the initial listing within segments.

The selection process for the civilian sample[1] involved selection of primary sampling units (PSUs), block groups (BGs) within these, and segments within block groups. Dwelling units were then listed within segments and a probability selection of listing lines were scheduled for screening interviews to locate inscope[2] respondents. With a few exceptions noted in Chapter 3, all inscope respondents located in screening were designated for a baseyear NLS interview. NORC interviewers then attempted to contact the designated individuals and conduct the interview.

The results of these selecting, screening, interviewing, and testing efforts are displayed in Tables 2.1 and 2.2. In the cross-section and supplemental samples, respectively, screening interviews were completed in 91.2 percent and 91.9 percent of the occupied dwelling units selected for screening. This screening completion rate is commensurate with that customarily obtained in U.S. household surveys. No information was obtained about the individuals living in dwelling units where we were unable to conduct a screening interview. Chapter 4 describes the weighting correction that was made for screener nonresponse.

In general, all inscope individuals located in screening were designated for a baseyear interview. However, in the cross-section sample,

---

[1]Selection of the military sample used a different sampling frame but employed strict probability sampling.

[2]For purposes of this study individuals were considered "inscope" (eligible) if they were born in 1957 through 1964, were living within the 50 states, and were not living in an institution on a permanent basis.

TABLE 2.1

CROSS-SECTION SAMPLE YOUTH LOCATED IN SCREENING, SELECTED FOR BASEYEAR INTERVIEW, AND COMPLETED INTERVIEW

Design Cohort	Located in Screening	Out-of-Scope	Selected for Interview	Completed Interview
Males:				
Hispanic.....	249	5	244	216
Non-Hispanic black.....	386	7	379	347
Economically disadvantaged non-Hispanic, non-black...	189	5	184	203[a]
Other.....	2,599	39	2,560	2,238
Total.....	3,423	56	3,367	3,004
Females:				
Hispanic.....	248	--	248	228
Non-Hispanic black.....	451	13	438	404
Economically disadvantaged non-Hispanic, non-black...	180	--	180	198[a]
Other.....	2,620	41	2,579	2,277
Total.....	3,499	54	3,445	3,107
TOTAL.....	6,922[b]	110	6,812	6,111[c]

[a]A number of individuals who had not been identified as economically disadvantaged at the time of screening and selection for interview were reclassified after their interview. This accounts for the apparent discrepancy between the number selected and the number interviewed in this design cohort.

[b]Screening interviews were completed in 91.2 percent of the occupied dwelling units selected for screening.

[c]Interviews were completed with 89.7 percent of those selected for baseyear interviews.

TABLE 2.2

SUPPLEMENTAL SAMPLE YOUTH LOCATED IN SCREENING, SELECTED FOR BASEYEAR INTERVIEW, AND COMPLETED INTERVIEW

Design Cohort	Located in Screening	Out-of-Scope	Deleted by Subsampling	Selected for Interview	Completed Interview
Males:					
Hispanic	1,015	26	161	828	730
Non-Hispanic black	1,318	42	50	1,226	1,096
Economically disadvantaged non-Hispanic, non-black	887	28	1	858	744
Other	--	--	--	--	--
Total	3,220	96	212	2,912	2,570
Females:					
Hispanic	1,060	16	205	839	750
Non-Hispanic black	1,502	31	298	1,173	1,076
Economically disadvantaged non-Hispanic, non-black	1,073	28	--	1,045	899
Other	--	--	--	--	--
Total	3,635	75	503	3,057	2,725
TOTAL	6,855[a]	171	715	5,969	5,295[b]

[a]Screening interviews were completed in 91.9 percent of the occupied dwelling units selected for screening.

[b]Interviews were completed with 88.7 percent of those selected for baseyear interviews.

110 individuals identified as inscope at the time of screening were subsequently found to be out-of-scope and deleted from the sample. In the supplemental sample an unanticipated number of inscope youth were located in screening and it was necessary to delete some of them through subsampling. This was carried out on a strict probability basis and compensated by an appropriate weighting adjustment, described in Chapter 4.

In the cross-section and supplemental samples, respectively, 89.7 percent and 88.7 percent of those designated for a baseyear interview were successfully interviewed in 1979. Those completion rates are typical of rates obtained in similar surveys.

The military sample selected for the NLS was composed of 1,793 men and women serving on active duty in the military as of September 30, 1978 and born in 1957 to 1962. Individuals were selected in a two-stage, stratified selection procedure. Stratification was by Military Service (Army, Navy, Air Force, and Marine Corps) and by five geographical regions. Females were oversampled at a rate approximately six times that of the males in order to produce approximately equal numbers of males and females in the final sample. Table 2.3 shows the distribution of NLS baseyear completed cases by sex and Military Service. Further details on the selection and weighting of the military sample are provided in Chapters 3 and 4.

Those persons who completed interviews in the baseyear survey formed the target sample for the followup surveys conducted in 1980, 1981, and 1982. These followup rounds are referred to as Round II, Round III, and Round IV, respectively The baseyear is occasionally referred to as Round I. There was no subsampling for Rounds II-IV, and the high completion rates are indicated by Tables 2.4 - 2.6.

TABLE 2.3

MILITARY SAMPLE BASEYEAR INTERVIEW COMPLETED CASES

Service	Completed Interview
Males:	
Army	352
Navy	212
Air Force	163
Marine Corps	96
Total	823
Females:	
Army	224
Navy	68
Air Force	131
Marine Corps	34
Total	457
TOTAL	1,280[a]

[a]Interviews were completed with 71.5 percent of those selected for baseyear interviews.

TABLE 2.4

CROSS-SECTION SAMPLE: COMPLETION OF CASES
FOR ROUNDS I, II, III, AND IV

Design Cohort	Round I	Round II	Round III	Round IV
Males:				
Hispanic	216	202	204	204
Non-Hispanic black	347	333	334	331
Economically disadvantaged non-Hispanic, non-black	203	196	196	198
Other	2,238	2,156	2,153	2,152
Total	3,004	2,887	2,887	2,885
Females:				
Hispanic	228	220	219	219
Non-Hispanic black	404	385	390	390
Economically disadvantaged non-Hispanic, non-black	198	190	191	183
Other	2,277	2,190	2,205	2,200
Total	3,107	2,976	3,005	2,942
TOTAL	6,111	5,873	5,892	5,877

TABLE 2.5

SUPPLEMENTAL SAMPLE: COMPLETION OF CASES
FOR ROUNDS I, II, III, AND IV

Design Cohort	Round I	Round II	Round III	Round IV
Males:				
Hispanic	730	684	699	683
Non-Hispanic black	1,096	1,050	1,067	1,043
Economically disadvantaged non-Hispanic, non-black	744	708	711	705
Other	--	--	--	--
Total	2,570	2,442	2,477	2,431
Females:				
Hispanic	750	716	717	707
Non-Hispanic black	1,076	1,045	1,051	1,047
Economically disadvantaged non-Hispanic, non-black	899	872	863	854
Other	--	--	--	--
Total	2,725	2,633	2,631	2,608
TOTAL	5,295	5,075	5,108	5,039

TABLE 2.6

MILITARY SAMPLE: COMPLETION OF CASES
FOR ROUNDS I, II, III, AND IV

Service	Completed Cases for Rounds			
	I	II	III	IV
Males:				
Army	352	330	333	331
Navy	212	187	187	187
Air Force	163	153	152	156
Marine Corps	96	88	88	93
Total	823	758	760	767
Females:				
Army	224	218	216	218
Navy	68	60	65	66
Air Force	131	125	124	128
Marine Corps	34	32	30	32
Total	457	435	435	444
TOTAL	1,280	1,193	1,195	1,211

CHAPTER 3

SELECTION AND IMPLEMENTATION OF THE SAMPLE

This chapter describes in some detail the procedures followed in selecting and implementing (interviewing) the three independent probability samples (cross-section, supplemental, and military) that constitute the NLS sample. It also describes special procedures that were used in the cross-section and supplemental sample to include dwelling units missed in the listing process or constructed after the listing took place, to include college students living in dormitories or other group quarters, and to include other persons living in group quarters.

Selection of the Cross-Section Sample

The cross-section sample was based upon the 102 PSU NORC National Probability Sample developed and initially used in 1973. The sample has been continuously updated since that time. The sampling frame covers the entire United States.

Stage I

The Primary Sampling Units are composed of: Standard Metropolitan Statistical Areas (SMSAs), counties,[1] parts of counties,[2] and independent cities. Stratification criteria used in the first stage of selection include: Census Division, SMSA-nonSMSA, county size, and percentage black. The selection of primary units was carried out with probabilities proportional to 1970 Census population (PPS), using replicated "zone" selection. A total

---

[1]Where necessary, counties were combined so that their aggregated 1970 population exceeded 12,000.

[2]In New England, we defined the portion of a county outside an SMSA as a PSU.

of 204 PSUs was selected. In this survey, we made use of two of the four replicates comprising 102 PSUs.

Stage II

The secondary units of selection are block groups (BGs) in areas for which Census blocks have been designated, and enumeration districts (EDs) in unblocked areas. Prior to selection, the second-stage (within-PSU) frame of EDs and BGs was stratified on the basis of median family income and percentage black.[1] For each primary sampling unit, eighteen secondary selections were made with probability proportional to size from eighteen equal-size zones. A subsample of nine secondary units was used for the cross-section sample.

Stage III

Whenever possible, secondary selections were subdivided[2] into third stage listing units (segments).[3] One listing unit was then selected for each secondary selection with probability proportional to estimated housing. If it was impossible to subdivide a secondary selection into well-defined subunits, this stage of sampling was bypassed (i.e., subsampling at stage III was accomplished with probability one).

NORC interviewers carried out dwelling units (DU) listing within all third-stage segments. Prior to initial use, those listings were updated and

---

[1] In areas that were not tracted, median household income and percentage black were estimated using a regression routine based on MCD or tract information.

[2] For BGs we employed block statistics, for EDs we made field counts.

[3] The minimum size for listing units was 100 dwelling units.

subjected to a number of checks.[1] In order to maintain an accurate record of dwelling units, master sample listings were periodically updated. This updating procedure occured at the end of each field period. During the updating period, and in conjunction with NORC's "missed dwelling unit" procedure, information was gathered regarding changes in the entire segment (e.g., demolition of DUs, new construction). This information was then integrated into the computer-based master listing of NORC PSUs.

Stage IV

The listings of dwelling units and individual quarters[2] from stage III segments were subsampled in order to produce an equal probability sample of households and individual quarters distributed among the 918 segments (102 PSUs x 9 segments per PSU). Selection of these listings was accomplished through the use of ANSPAK (NORC's computerized sampling program package). There was an average of twenty-four selected fourth-stage (listing) units per segment, resulting in an average of 7.4 inscope youths. All inscope youths found in this screening stage were designated for subsequent interview.

In total, screening interviews were scheduled for 22,077 listing units (households or IQs) in the cross-section sample. Of these households, 1,917 were found to be vacant, and 1,038 were found not to be dwelling units. A screening interview was completed with 17,445 or 91.23 percent of the remaining households. There were refusals from 985, or 5.16 percent of the households; 688, or 3.60 percent, were not completed for other reasons.

---

[1]A comparison was made with census estimates and/or field counts. Also, a number of internal consistency checks for sequential listing and procedures were initiated.

[2]Individual Quarters (IQ) is a term used to describe non-dwelling unit, noninstitutional living quarters.

Screening for the cross-section sample located a total of 6,922 eligible youth (those born in 1957 through 1964) distributed across various design cohorts as shown in Table 2.1. Some were subsequently found to be ineligible (out-of-scope) and were deleted, but the remainder were designated for a base year interview.

Selection of the Supplemental Sample

As noted previously, the supplemental sample was designed specifically to yield a highly efficient sample of the three youth cohorts designated for oversampling (i.e., Hispanics, non-Hispanic blacks, and non-Hispanic non-black economically disadvantaged). Thus, for this sample, stratification specifically relevant for these groups was used. In addition, probability proportional to size (PPS) procedures were based on size measures for these cohorts rather than the general population. In multistage samples, PPS procedures typically are used in order to achieve control over the distribution of sample cases among the primary sampling units and within the ultimate clusters that form the primary sampling units. By using size measures based on the three oversampled cohorts, it was possible to more nearly equalize the distribution of these groups among the various sampling units than would have been possible in a cross-sectional design that used PPS procedures based on total population.

Stage I

Primary sampling units consisted of counties and independent cities. First-stage selection of these units was carried out with probabilities proportional to measures of size that reflected the black, Hispanic, and economically disadvantaged non-Hispanic, non-black population within the PSU. These measures of size were constructed from the 1970 Census Fifth Count

(File C), which provided required estimates at the enumeration district-block group level within each county and independent city. Prior to use, 1970 size estimates were updated to 1977 Census estimates on a county basis.

For each primary sampling unit a measure of size was constructed as

$$MOS_i = H_i + .5 \times B_i + ED_i$$

where H_i, B_i and ED_i denote the estimated population sizes for Hispanics, blacks, and economically disadvantaged non-Hispanic non-blacks, respectively, in primary samping unit i. Given that the measures of size need only reflect relative population size, and given the rather uniform ratio of estimated 14-21 cohort to total population, no attempt was made to reapportion size measures to the youth cohort. The factor of .5 was applied to the black population in the construction of PSU measures to reflect the fact that among the three population groups of interest the required oversampling rate for blacks was approximately one-half the rate required for Hispanics and economically disadvantaged non-Hispanic non-blacks.

Prior to sample selection, PSUs were stratified on the basis of the 9 standard U.S. Census Divisions. Within each of these divisions, further stratification was based upon urban-rural location (within or outside the SMSA). Finally, within each of the 18 major strata (9 divisions x 2 urban/rural classes), PSUs were ordered by proportion of PSU population containing target group members. A systematic "zone" selection procedure was used to select 100 primary sampling units with probabilities proportional to the previously discussed target group measures of size.

Stage II

Within selected primary units, the units of second-stage selection were either Census block groups or enumeration districts. These second-stage

sampling units were assigned measures of size by the same procedure that had been used in constructing measures at the first stage of sampling. Since the first-stage measures had been created by aggregating information at the block group and enumeration district level, from the Fifth Count File C Census tape, the process of assigning second-stage measures was simply a disaggregation procedure.

Prior to selection, second-stage units were sort-ordered by estimated proportion of population containing members of the target population. Adjoining units were then linked, when necessary, in order to have a minimum size measure of 25.

Within each selected primary sampling unit, nine secondary units were selected using a systematic zone procedure with probabilities proportional to target group measures of size.

Stage III

Whenever possible, selected secondary selections were subdivided into third-stage listing units (segments). One listing unit was then selected for each secondary selection with probability proportional to estimated housing. If it was impossible to subdivide a secondary selection into well-defined subunits, this stage of sampling was bypassed (i.e. subsampling at stage III was accomplished with probability one). It should be noted that because measures of size used at stages one and two were based upon target population rather than total population, the number of housing units contained within any two third-stage segments with the same measure of size might be quite different. In general, we tried to make use of third-stage segments containing measures of size in the range 25-50 with between 50 and 500 housing units.

NORC interviewers carried out dwelling unit listings within all 900 third-stage segments. Prior to use, these listings were subjected to a number of internal and external checks. Listers were required to seek out reasons for differences between the number of housing units found at the time of listing and the number of housing units reported by the 1970 Census. Within each block, checks were made, where possible, for consistent ordering of street numbering of listed units.

Stage IV

The fourth stage of selection involved sampling of dwelling unit and individual quarters listings within the 900 selected third-stage segments. Screening, which consisted of enumeration of all persons within selected dwelling units (on a family unit basis), was conducted in two waves. In general, selection of third-stage listings was carried out with probabilities designed to equalize the overall probability of selection through the four stages of sampling. However, there was some degree of oversampling (increased probability of selection) among third-stage units that were estimated to contain a higher proportion of individuals in the three population groups designated for overrepresentation (i.e. Hispanics, non-Hispanic blacks, and economically disadvantaged non-Hispanic nonblacks).

Screening interviews were scheduled for 68,861 fourth-stage units in the supplemental sample. Of these, 5,905 were found to be vacant, and 2,275 were found not to be dwelling units. A screening interview was completed in 55,737, or 91.85 percent, of the remaining households and individual quarters. There were refusals from 2,923, or 5.24 percent of the households, and 2,021 or 3.63 percent were not completed for other reasons.

Screening for the supplemental sample located a total of 6,855 eligible youth distributed across various design cohorts, as shown in

Table 2.1. However, the number of individuals in the Hispanic and non-Hispanic black cohorts was somewhat larger than required. Therefore, it was necessary to select a subsample of these individuals for base year interviewing. Table 2.2 shows the number of individuals selected by subsampling in these two cohorts.

Procedures used for the subsampling of individuals in these two cohorts for base year interview were designed to equalize, as much as possible, final overall probabilities of selection for individuals within the same design cohort. Specifically, since some degree of differential oversampling was applied in the fourth-stage selection of dwelling units for screening, individuals located in the screening process had not been selected with the same probabilities. Within the constraints of probability sampling, probabilities associated with the stage five subsampling process were set inversely proportional to the probabilities of selection for prior stages (i.e. product of stages one through four). As a result, the variation in probability of selection among individuals (within a design cohort) retained in the sample after stage five was decreased from the variation in probabilities among all screened individuals within the same design cohort.

Special Procedures Used in Both the Cross-Section and Supplemental Samples

There were several special procedures used in both the cross-section and supplemental samples to accomplish the following goals:

1. Inclusion of dwelling units in the sample that either were missed in the listing process or were constructed after the listing process took place.
2. Inclusion in the sample of noncollege individuals living in non-institutional non-dwelling unit living arrangements.
3. Inclusion in the sample of college students living in non-dwelling unit quarters.

Procedures for Inclusion of Unlisted (Missed) Dwelling Units

As part of its standard field methods, NORC makes use of a procedure to give a proper probability of selection to dwelling units (DU's) that did not exist or were missed at the time of original listing or during segment updating. The method we employ is an application of the half-open interval technique.[1] This procedure explicitly links every nonlisted DU in a segment with exactly one listed DU in that segment. Thus, each listed dwelling unit represents a cluster of dwelling units composed of the listed DU (line) and any missed DUs associated with that line.

Operationally, the procedure is simple. The set of DU listings (lines) for a segment is made up of one or more subsets of lines (blocks). Each block consists of an ordered set of lines. Each of the lines represents either a complete structure (i.e., a single-family dwelling unit) or a subunit within a structure (i.e., an apartment in an apartment building or complex).[2] The half-open interval procedure followed differs slightly depending on whether the selected line is a complete structure or a subunit within a structure.

Whenever a line is selected that is a complete structure, all dwelling units within that structure are included in our sample, as are any dwelling units between[3] the selected structure and the next structure listed in the

---

[1]See Leslie Kish, Survey Sampling, NY: John Wiley & Sons, 1965, p. 56.

[2]Even if a listing contains a within-structure description (e.g., 304 Main, 2nd floor) it is considered a structure listing if there is no other listing that refers to that structure.

[3]If structures have numbered street addresses, "between" is defined in terms of these address numbers. In areas where numbers are not used, "between" is defined in terms of location.

same block.[1] Our instructions to the interviewer are as follows:

(selected line description)
Message 1: Check for missed DUs at the address above.
Check for missued DUs between street address above and street address below.
(next listed line description)

When a line listing within a multiple unit structure is selected, the half-open interval instructions depend on whether the line represents the first subunit in the structure, the last subunit in the structure, or a nonfirst/nonlast subunit.[2]

When we select the first subunit in a multiple structure, we include in our sample all dwelling units that exist within the selected subunit, as well as any dwelling units within the structure that are not already listed. Our instructions to the interviewer are:

(selected line description)
Message 2: Check for missed DUs at this apt. number.
Check for DUs at this street address not listed on the (attached) segment printout.

When the selected line is the last subunit listing of a multiple structure, we include in our sapmle all dwelling units within the selected subunit and all dwelling units between the structure in which the subunit is contained and the next listed structure in the block. Here the instruction to the interviewer is:

(selected line description)
Message 4: Check for missed DUs at this apt. number.
Check for missed DUs between this street address and the street address below.
(next listed line description)

---

[1]The listings within each block are considered circular (i.e., the last listing within a block is followed by the first).

[2]For each listing that identifies a subunit within a structure, there must be at least one other listing within the same structure. (This follows from the definition of a listing as either a complete structure listing or a subunit within a structure.) Our listings are so ordered that for each structure in which subunits are listed there must be a unique first-subunit and a unique last-subunit listing.

If the selected line is a non-first/non-last subunit listing, we inlcude in the sample only dwellings within the selected subunit. In this case, the following instruction is used:

(selected line description)
Message 3: Check for missed DUs at this apt. only.

Procedures to Insure Coverage of the Non-DU Population (College Dormitories and Other Group Quarters)

Since the initial cohort definitions include civilian youth aged 14 to 21 living in all noninstitutional settings, special procedures were used to insure appropriate sample coverage in living units not classified as dwellings. These non-DU living units include college dormitories and other group quarters (GQ's).

In past surveys of the noninstitutional adult population, NORC has used a single procedure to obtain sample coverage of the non-DU, noninstitutional civilian population. Because of the restricted age distribution in the NLS study, NORC made use of two procedures. One of these procedures was used to cover the noncollege portion of this non-DU population; another procedure was used for college students.

The inclusion of the noncollege, noninstitutional, non-DU population aged 14 to 21 was accomplished by the following two-stage procedure. The first stage was carried out prior to the beginning of field interviewing. Each segment in use for the survey was field-enumerated for all group quarters structures, except college dormitories. Within these group quarters structures, a complete listing of individual quarters (IQs: beds and/or rooms with beds) was undertaken. The listing of IQs was then subsampled using the same final-stage selection procedure applied to dwelling units within the segment.

The second stage in the NORC group quarters sampling procedure was

carried out at the time of screening in conjunction with the standard NORC missed-dwelling-unit procedure. All group quarters except college dormitories that were not explicitly listed in the first step of the individual quarters procedure were eligible for selection at this stage. These non-first-stage group quarters are implicitly linked to listed dwelling units by the same linking rules applicable to nonlisted dwelling units. For each selected dwelling unit, a check was made for implicitly linked but unlisted individual quarters units. As is the case with our missed dwelling unit procedure, the instructions for the missed individual quarters procedure were computer-generated for each selected dwelling unit. The interviewer was provided with specific instructions indicating the appropriate DU/IQ checks that must be carried out at each selected dwelling unit.

Special procedures were also used for college students. As of October 1979, approximately one-third of the civilian population between the ages of 18 and 21 was enrolled in college.[1] In many household surveys the coverage of the college population is haphazard and ill-defined. Given the nature of the proposed research, the following special procedures were used to insure complete coverage of this portion of the youth cohort.

Through a set of explicit rules, every full- or part-time college student was "linked" to a unique living unit that had a known probability of entering the sample. These rules "link" college students who live in a non-DU setting (dormitory) away from their parents' home for parts of the year to their parents' home. This alternative was chosen for both sampling and operational reasons. From a sampling standpoint, linkage of college students

---

[1] U.S. Bureau of the Census, "School Enrollment-Social and Economic Characteristics of Students: October 1979 (Advance Report)," Current Population Reports, Series P-20, No. 355, Table 6, August 1980.

living in non-DU settings to parents' DUs tends to minimize the occurrence of small-area "pockets" of inscope population and the resulting large variability in cluster size. From the standpoint of field operations, the parents' home represented a contact location of relative stability. This relative stability is crucial in the yearly follow-up efforts.

The specific linkage rules are as follows:

- College students who live in a specified dwelling unit on a year-round basis are linked to that dwelling unit.
- College students who do not live in dwelling units on a year-round basis are linked to their parents' or guardians' DUs.
- In situations where the application of this condition results in multiple linkages (e.g., divorced or separated parents living in two separate DUs), a unique linkage is established on the basis of maximum financial support.

When this condition does not provide a unique linkage, the following priority scheme is used.

- Living natural or adoptive mother
- Living natural or adoptive father
- Living female guardian
- Living male guardian

In order to implement this procedure, we collected potential linkage information at all sample DUs and GQs (i.e., we asked parents about children that were away at achool). In most situations, unmarried college students in the 14 through 21 cohort were linked to their parents' DU, married couples or cohabiting couples living in DUs on a year-round basis were linked to their own DUs, and married couples or cohabiting couples not living in a DU on a year-round basis were linked to their respective parents' DUs.

The difficulties associated with sample frame coverage of the college student population are well known to survey researchers. Therefore, in

developing the overall NLS weighting procedures (described in Chapter 4) we were prepared to include a post-stratification adjustment for any undercoverage of the college student population. After examining the sample data, however, we found that this separate adjustment was not required. Projected college attendance for the NLS universe (without separate adjustment) was 4,967,000 full-time attendees and 722,000 part-time attendees. The most current U.S. census estimates at the time reported 4,918,000 full-time attendees and 591,000 part-time attendees between the ages of 14 and 21 as of October, 1979.[1]

Selection of the Military Sample

As of September 30, 1978, there were 657,549 members of the active armed forces who would be between the ages of 17 and 21 as of January 1, 1979. Individuals in this group were sampled by a stratified, two-stage selection procedure. The sample design for this portion of the youth cohort was developed in cooperation with the Department of Defense, the Defense Manpower Data Center, the Rand Corporation Department of Defense Survey Group, the Center for Human Resource Rersearch at Ohio State University, and NORC. Actual selection of sample individuals was carried out jointly by the Department of Defense, the Defense Manpower Data Center (DMDC), and NORC.

The basic sample design called for the selection of a sample of approximately 1,300 members of the active armed forces. In order to provide samples of sufficient size for separate estimates with respect to sex, it was decided to sample females at a rate approximately six times that used for males. This would produce approximately 850 males and approximately 450

---

[1] U.S. Bureau of the Census School Enrollment-Social and Economic Characteristics of Students: October, 1979 (Advance Report), Current Population Reports, Series P-20, No. 355, August 1980.

females. Within each group, all individuals were to be sampled with equal probability. Within each sex, the sample was stratified on the basis of Military Service and geographic location. Proportionate allocation was used with respect to these stratification cells. Sample selection was carried out in two stages.

Stage I

Each of the four Military Services (Army, Air Force, Navy and Marine Corps) maintains up-to-date lists of all personnel. Included in these lists is information about age, sex, and assignment UIC (unit identification code). It would have been possible to sample individuals from these lists directly in a single stage of sampling (i.e., simple random element sampling). However, the face-to-face nature of the baseyear interview led us to decide to use cluster sampling.

The primary units of sample selection were composed of individuals within the same unit identification code. This unit code typically defines a group of individuals residing at the same physical location. Over all Services there were a total of 12,488 UIC's containing one or more persons in the 17-21 youth cohort. Because differential sampling rates were to be applied to males and females, these UIC's were first separated into two groups: Group 1 consisted of UIC's with no females in the 17-21 cohort; Group 2 consisted of UIC's with at least one female in the 17-21 cohort.

Each of the two groups of UIC's was divided into 20 basic strata, defined on the basis of Military Service and geographic location as follows:

I. MILITARY SERVICE (4 Services)

A. ARMY

B. NAVY

C. AIR FORCE

D. MARINE CORPS

II. GEOGRAPHIC LOCATION (5 categories)

A. EASTERN UNITED STATES

B. WESTERN UNITED STATES

C. EUROPE

D. FAR EAST

E. OTHER

Within each of these 20 basic strata UIC's were linked together in order to form primary sampling units (PSU's) as follows:

1. UIC's in group 1 (males only) were linked in order to form PSU's with a minimum of 20 males.

2. UIC's in group 2 (at least one female) were linked in order to form PSU's with a minimum of 20 males and 10 females.

In the linkage process, attempts were made to minimize the geographic distance among UIC's within the same PSU. This linkage process resulted in the formation of 3,711 group 1 and 2,256 group 2 PSU's across the 20 basic strata.

First-stage selection of PSU's was carried out within each of the 20 basic design strata separately for males and females. Within each sex the probability of selection for a PSU was proportional to the number of 17-21 youth (of that sex) within the PSU.

Let MOS_{mi} = the number of 17-21 males within the i^{th} PSU

MOS_{fi} = the number of 17-21 females within the i^{th} PSU

For the male sample, the probability of selection for the i^{th} PSU

$$f_{mi} = \frac{150\ MOS_{mi}}{579,508}$$

For the female sample, the probability of selection for the i^{th} PSU was

$$f_{fi} = \frac{110\ MOS_{fi}}{47,305}$$

For both the male and female samples the probability of selection for the i^{th} PSU was constrained to an upper limit of unity. Thus, any PSU whose measure of size for males (MOS_{mi}) exceeded 579,508/150 = 386.38 was selected with certainty. Any PSU whose measure of size for females (MOS_{fi}) exceeded 47,305/110 = 430.05 was selected with certainty.

It should be noted that although separate samples were selected for males and females, a form of the Keyfitz procedure was used in order to maximize the overlap between PSU's selected for the male sample and PSU's selected for the female sample.

In total, 146 PSU's were selected for the male sample and 103 PSU's were selected for the female sample. The overlap among these units was 58.

Stage IIa

Within-PSU selection was caried out by DMDC. On the basis of specifications provided by NORC, selected PSU's were subsampled at the rates of $13.35/MOS_{mi}$ for the male sample and $9.35/MOS_{fi}$ for the female sample. In those instances where Stage I PSU selection had been made with certainty (probability = 1), within-PSU selection was carried out with sampling rates 1/289.3922 for male sample PSU's and 1/45.7495 for female sample PSU's. This sampling produced a list of 3,073 persons.

Stage IIb

The sample produced at Stage IIa was systematically subsampled at a rate of one in two in order to provide 1,537 names. Prior to subsampling the Stage IIa list produced by DMDC was ordered by PSU in order to assure that all PSU's would be included in the subsample. Subsequently, an additional subsample of 256 names was selected by systematic seleciton from the remaining unselected names on the DMDC Stage IIa sample list.

In combination these subsamples produced a uniform stage IIb subsample rate of 1792.5/3073.

The stages of sampling described above produced the following overall sampling rates:

$$f\ (\text{males}) = \frac{150\ MOS_{mi}}{579{,}508} \times \frac{13.35}{MOS_{mi}} \times \frac{1792.5}{3073} = 1/496.124$$

$$f\ (\text{females}) = \frac{110\ MOS_{fi}}{47{,}305} \times \frac{9.35}{MOS_{fi}} \times \frac{1792.5}{3073} = 1/78.851$$

Implementation of the Sample

The selection procedures described above produced a list of individuals selected for inclusion in the NLS sample. During the spring of 1979 NORC interviewers attempted to contact each of these selected individuals and conduct a one-hour personal interview. Tables 2.1 and 2.2 above showed the overall completion rates in the two civilian samples and the number of individuals interviewed in each design cohort. The total number of completed interviews was 12,686.

In probability-based survey sampling completed cases are customarily weighted to reflect differential probabilities of selection and to correct for nonresponse at various stages of the survey process. Chapter 4 describes how this was done for the NLS baseyear completed cases and for the Round II, III, and IV completed cases in order to form a database that can be statistically projected to represent the entire American population born in 1957 through 1964.

CHAPTER 4

WEIGHTING THE SAMPLE

The basic objectives of the weighting procedures were the same for both the civilian and military portions of the total youth sample. In general, these objectives included correction for differential probabilities of selection, correction for differential completion rates among the basic design cohorts, and post-stratification to known total subpopulation size. Given the different nature of the civilian and military sampling frames and selection methods, the weighting steps that were applied to the civilian and military samples were somewhat different.

Weighting the Civilian Sample

In the sections that follow we first describe the steps that were applied in weighting the NLS baseyear interviews. We then desribe modifications of this weighting process that were used in weighting the Round II, III, and IV data.

NLS Baseyear Weights

Weighting for the NLS baseyear civilian sample involved five basic steps. These steps were designed to accomplish the following objectives:

Step 1. Correction for differential probability of selection at the initial stage of household selection.

Step 2. Correction for differential completion rates at the initial "screening phase" of data collection.

Step 3. Correction for differential subsampling rates for Hispanic and black cohort members prior to initial interview. Correction for differential completion rates among all cohort members at the first-year interview stage of data collection.

Step 4. Proper combination of cases obtained in the cross-sectional and supplemental samples.

Step 5. Adjustment of weighted cohort sizes to conform with outside, independent Census estimates projected to January 1, 1979.

Step 1. In the initial step, weights were assigned to each completed case on the basis of the selection probability for the dwelling unit that contained the family unit where the respondent was initially located (i.e., listed). For the ith respondent, this weighting factor was

$$W_{1i} = 1/f_i,$$

where f_i is the probability of selection for the dwelling unit containing the familiy unit where the respondent was initially listed in the screening process.

Step 2. In this step, a cluster-specific adjustment was introduced in order to compensate for differential completion rates in the family unit within dwelling-unit screening process. There were 1,818 selection clusters in the entire sample (918 in the cross-sectional sample and 900 in the supplemental sample). For the ith respondent, this adjustment factor was

$$W_{2i} = \frac{\text{Number of family units selected for screening in the cluster containing the ith respondent}}{\text{Number of family units in the ith respondent's cluster where screening information was obtained}}$$

In those instances where refusals were encountered at the dwelling-unit level (i.e., it was impossible to determine whether or not there was more than one family unit within the dwelling unit), the ratio of family units to dwelling units for the remainder of the cluster was used to estimate the number of family units contained within the dwelling unit. W_{2i} was constrained to an upper limit of 1.5 in order to limit the potential impact of extreme weights.

Step 3. In this step adjustments were made for the additional stage of subsampling applied to blacks and Hispanics screened in the supplemental

sample prior to initial interview. In addition, adjustment factors were applied to all selected respondents to compensate for differential response rates in the first interview. These nonresponse adjustment factors were applied at the PSU level (102 cross-sectional PSU's and 100 supplemental PSU's) for each of the eight basic design cohorts listed below:[1]

1. Hispanic males
2. Hispanic females
3. Non-Hispanic, black males
4. Non-Hispanic, black females
5. Economically disadvantaged, non-Hispanic, non-black males
6. Economically disadvantaged, non-Hispanic, non-black females
7. Other males
8. Other females

Thsu, the step 3 weight factor for the ith respondent was

$$W_{3i} = A_{3i}/s_i,$$

where,

$$A_{3i} = \frac{\text{Number of assigned cases with respondent i's PSU and design cohort}}{\text{Number of completed cases within respondent i's PSU and design cohort}}$$

and

s_i = probability of retention in sample if ith respondent was in black or Hispanic design cohort of supplemental sample,

= 1, otherwise

An upper limit of 1.5 was applied to the factor A_{3i}.

Step 4. The purpose of this step was to rescale the weights developed in steps one, two, and three for cases in design cohorts 1-6 in order to properly combine respondents from the cross-sectional and supplemental samples. Prior to this step, the supplemental and cross-sectional samples

---

[1]All basic design cohorts, except 7 and 8, were sampled in both the cross-sectional and supplemental samples.

were treated as independent units. This rescaling was carried out separately for each of the six design cohorts present in both the cross-sectional and supplemental samples.[1]

Within each of the cohorts a preliminary weight was computed for each respondent within the cohort. For the ith respondent within the cohort, this preliminary weight was the product of weights developed at steps 1, 2, and 3. Specifically,

$$W'_{4i} = W_{1i} \times W_{2i} \times W_{3i}$$

Within each of the cohorts separate means and standard deviations were calculated for these preliminary weights from the cross-sectional and supplemental portions of the cohort. Thus, within a specified cohort

M_c = Mean of weights W'_{4i} from the cross-sectional portion of the cohort.

M_s = Mean of weights W'_{4i} from the supplemental portion of the cohort.

S_c = Standard deviation of weights W'_{4i} from the cross-sectional portion of the cohort.

S_s = Standard deviation of weights W'_{4i} from the supplemental portion of the cohort.

---

[1]Because screener information was not available for approximately 20 percent of the cross-sectional sample, it was not possible to classify all non-Hispanic, non-black individuals on the basis of poverty status (i.e., economically disadvantaged). In the application of procedures that combined portions of the supplemental sample with portions of the cross-sectional sample (step 4), the cross-sectional portion of the economically disadvantaged cohort was restricted to those individuals with income information available from the screening interview who met the conditions for this classification. The resulting weights for this combined group were rescaled, separately by male-female, so that the weighted sum of known economically disadvantaged individuals among the non-black, non-Hispanic cohort would equal the initially weighted sum (after step 3) of known econmically disadvantaged non-Hispanic, non-black males-females in the cross-sectional sample.

These means and standard deviations were used to determine the weighting efficiency factor for the cross-sectional and supplemental portions of the sample for the cohort as follows:

$$WEF_C = \frac{1}{1 + (S_C/M_C)^2}$$ = weighting efficiency factor, cross-sectional portion

$$WEF_S = \frac{1}{1 + (S_S/M_S)^2}$$ = weighting efficiency factor, supplemental portion

These efficiency factors were used in conjunction with the actual number of cases within the cross-sectional and supplemental portions of the cohort to determine the effective sample bases for these portions of the cohort.

Thus,

$$ESB_C = n_C \times WEF_C$$

$$ESB_S = n_S \times WEF_S$$

where

n_C and n_S are defined as the number of sample cases in the cross-sectional and supplemental portions of the cohort, respectively,

and

ESB_C and ESB_S are defined as the effective sample bases for the cross-sectional and supplemental portions of the cohort, respectively.

Using these effective sample bases, adjustment factors were developed for the cross-sectional and supplemental portions of the specified cohort so that the proportion of weighted cases from the cross-sectional and supplemental parts of the cohort would be in the same relationship as the effective sample bases from these two parts of the total cohorts.

Using the preliminary weights W'_{4i}, the total sum of weights from both portions of the cohort is

$$TSW = (n_c \times M_c) + (n_s \times M_s)$$

The adjustment factor for the cross-sectional portion of the cohort was

$$A_{4c} = \frac{P_c \times TSW}{n_c \times M_c}, \text{ where } P_c = \frac{ESB_s}{ESB_c + ESB_s}.$$

The adjustment factor for the supplemental portion of the cohort was

$$A_{4s} = \frac{P_s \times TSW}{n_s \times M_s}, \text{ where } P_s = \frac{ESB_s}{ESB_c + ESB_s}.$$

These adjustment factors were applied to the preliminary step 4 weights W'_{4i} to produce final step 4 weights W_{4i}.

$W_{4i} = A_{4c} \times W'_{4i}$, for i within cross-sectional portion,

$W_{4i} = A_{4s} \times W'_{4i}$, for i within supplemental portion.

Numbers of cases, mean weights, and standard deviations of weights are shown in Tables 4.1 and 4.2.

As an aside, the reader may note that this method attempts to minimize the overall variance of estimation by combining cross-sectional and supplemental sample portions of a specific design cohort in proportion to their effective sample size (equivalent simple random sample size). Estimation of the effective sample base (i.e., effective sample size) for each of the two sample portions (cross-section and supplemental) that comprise a design cohort is based upon an approximation formula that is often used to evaluate the design effect (DEFF) associated with weighted samples. The design effect is proportional to 1 + RV, where RV denotes the relative variance of the case-specific sample weights. The factors A_{4c} and A_{4s} are simple rescaling factors that force the sum of the rescaled weights from the two sample portions to have the same relationship as the effective sample bases.

TABLE 4.1

MEAN WEIGHTS AND STANDARD DEVIATIONS AFTER STEPS 1 THROUGH 3 FOR COHORT GROUPS

Group	Number of Cases	Mean Weight (Mc or Ms) (Hundreds)	Standard Deviation of Weights (Sc or Ss) (Hundreds)
Hispanic males			
Cross-section	216	43.4066	7.6727
Supplement	730	12.0197	5.2937
Hispanic females			
Cross-section	228	43.4185	7.4719
Supplement	750	12.4562	5.3139
Non-Hispanic black males			
Cross-section	347	42.6462	6.5155
Supplement	1,076	15.7230	6.9737
Non-Hispanic black females			
Cross-section	404	42.3624	7.3335
Supplement	1,096	18.4784	5.7822
Economically disadvantaged non-Hispanic, non-black males			
Cross-section	203	40.9998	5.3146
Supplement	744	10.2600	5.7963
Economically disadvantaged non-Hispanic, non-black females			
Cross-section	198	42.1845	5.8513
Supplement	899	10.3569	6.1670
Nondisadvantaged non-Hispanic, non-black males			
Cross-section	2,238	42.7289	5.7063
Nondisadvantaged non-Hispanic, non-black females			
Cross-section	2,277	42.4614	5.6103

TABLE 4.2

MEAN WEIGHTS AND STANDARD DEVIATIONS AFTER STEPS 1 THROUGH 4 FOR COHORT GROUPS

Group	Number of Cases	Mean Weight (Mc or Ms) (Hundreds)	Standard Deviation of Weights (Sc or Ss) (Hundreds)
Hispanic males			
Cross-section	216	12.1711	2.1515
Supplement	730	10.5125	4.6299
Hispanic females			
Cross-section	228	11.8207	2.0342
Supplement	750	10.2968	4.3929
Non-Hispanic black males			
Cross-section	347	17.2401	2.6340
Supplement	1,076	14.7425	6.5388
Non-Hispanic black females			
Cross-section	404	16.3722	2.8243
Supplement	1,096	15.3587	4.8060
Economically disadvantaged non-Hispanic, non-black males			
Cross-section	203	10.7190	1.3895
Supplement	744	8.2621	4.6676
Economically disadvantaged non-Hispanic, non-black females			
Cross-section	198	9.5518	1.3249
Supplement	899	7.1872	4.2796
Nondisadvantaged non-Hispanic, non-black males			
Cross-section	2,238	42.7289	5.7063
Nondisadvantaged non-Hispanic, non-black females			
Cross-section	2,277	42.4614	5.6103

Step 5. In the final step of weighting, the sum of weights from each of 48 post-strata (6 sex-race groups x 8 age groups) was adjusted to estimates of population size derived from U.S. Census estimates. This was accomplished by application of the adjustment factor A_5, within each of the 48 post-strata as follows:

Within each of the 48 post-strata,

NSP = total population estimate developed as above

NSS = total sum of weights W_{4i} for the cohort

A_5 = NSP/NSS

This factor was applied to each of the final step 4 weights to produce a final respondent weight for year one.

$W_i = A_5 \times W_{4i}$ (W_i = final weight for ith respondent)

As noted above, the 48 post strata were defined on the basis of the 6 sex-race groups by 8 age groups, as follows:

6 SEX/RACE GROUPS[1]

Males	- Hispanic
Males	- Black non-Hispanic
Males	- Others
Females	- Hispanic
Females	- Black, non-Hispanic
Females	- All others

8 AGE GROUPS

Single Birth Years 1957, 1958, ..., 1964

Estimates of post-stratum size were derived as follows:

1. Estimates of the civilian population of the U.S. were obtained by sex, single year of age, and race (black, other) as of July 1, 1978 from Table 3, of Current Population Reports, Series P-25, No. 800.

---

[1]Post-stratification weighting did not make use of a separate economically disadvantaged classification because of the lack of classification information.

2. By using the 13 and 21 year cohorts, these population estimates were carried forward 6 months to produce estimates of the 14-17 and 18-21 population by sex as of January 1, 1979.

3. Current Population Reports, Series P-20, No. 339: Persons of Spanish Origin in the United States: March, 1978 was used to estimate the number of non-black Hispanics in each of the single year age cohorts. Current Population Reports, Series P-60, No. 120: Money Income and Poverty Status of Families and Persons in the United States: 1978 was used to estimate the number of economically disadvantaged non-Hispanics, non-blacks in each of the single year age cohorts.

Post-stratification population targets are shown in Table 4.3.

Weights for Rounds II, III, IV

Two basic considertions affect the development of individual case weights to be applied to civilian data collected in Rounds II, III, and IV. The first consideration, which turned out to have minimal impact, stems from the fact that the population cohorts under study decline in numbers from year to year. The second consideration arises because some of the Round I participants either refused to participate in a later-round interview or could not be located.

Shrinkage over time of the cohorts is primarily a result of death or emigration. Given the age range for the cohorts under study, deaths effect a trivial shrinkage (less than 3 per 1,000), which is well within the errors of the target cohort population sizes established by the U.S. Bureau of the Census. Although reliable data pertaining to age-specific emigration data are not available, we believe that outmigration among the age ranges under consideration would also have negligible impact.

While it is impossible to develop a weighting procedure that can fully compensate for every bias that may arise because of the inability to collect Round II, III, or IV data from some baseyear respondents, we can adjust the weights for components of bias attributable to differential Round I, II, III,

TABLE 4.3

CIVILIAN SAMPLE POST-STRATIFICATION POPULATION TARGETS
(in thousands)

Birth Year	Males			Females		
	Hispanic	Non-Hispanic Black	Other	Hispanic	Non-Hispanic Black	Other
1957	112.4	241.0	1,635.6	127.0	280.5	1,689.5
1958	112.1	248.0	1,639.4	135.9	287.0	1,697.6
1959	113.4	258.0	1,641.1	136.1	287.0	1,671.4
1960	131.7	282.0	1,680.3	128.4	290.5	1,665.1
1961	137.9	296.5	1,699.6	125.2	293.0	1,645.3
1962	139.5	298.0	1,679.0	123.5	292.5	1,616.0
1963	139.0	300.0	1,659.0	126.4	294.0	1,591.6
1964	144.3	292.0	1,583.2	140.3	288.0	1,515.7

and IV completion rates among different subgroups. At the same time, we must weigh this potential bias reduction against increases in random sampling error that typically accompany increased weighting.

The same basic steps that were applied in weighting the baseyear NLS cases were applied in weighting the Round II, III, and IV civilian completed cases. The only exception to this occurred in the final step (step 5) post-stratification. For the Round II, III, and IV samples, post-stratification factors A_5 were recomputed on the basis of the completed cases in the sample rather than the completed cases in the NLS baseyear sample. In addition, for some of the 48 post-strata, it was possible to make use of a more refined degree of post-stratification that was felt to provide a method of differential bias correction for the additional nonresponse that occurred between the baseyear interview and the later interviews.

In Rounds II, III, and IV, in order to correct for the various degrees of non-response found in each of the 48 post-strata cells, the post-stratification adjustment factors, A_5, were recalculated for each cell based on the population estimates produced by the completed cases. In each cell we obtained a population estimate by attaching the preliminary baseyear weights, W_{4i}, to the completed cases for Rounds II, II, and IV. These estimates were used to calculate a new set of post-stratification adjustment factors.

$$\text{Round II post-stratification adjustment factor} = A_5^{II} = \frac{\text{Census estimate of population in this cell}}{\text{Estimate derived from sum of the preliminary baseyear weights, } W_{4i}, \text{ attached to Round II completed cases in this cell}}$$

Using these new Round II post-stratification adjustment factors, we calculated new Round II weights for each Round II respondent by taking the product of his or her preliminary baseyear weight, W_{4i}, and the adjustment factor for his or her cell, A_5^{II}. Thus, $W_i^{II} = W_{4i} \times A_5^{II}$, where W_i^{II} equals

the Round II weight for the ith respondent, W_{4i} equals the preliminary baseyear weight for the ith respondent, and A_5^{II} equals the Round II post-stratification adjustment factor for the ith respondent's post-stratum. The weights for Rounds II and IV were developed analogously.

In some of the post-strata it was possible to make use of a modified form of weighting that provides a more satisfactory correction for the small amount of nonresponse that occured between the baseyear interview and the Round II, III, or IV interview. This modified procedure recognizes the fact that the probability that a baseyear respondent will agree to participate in the later round may be related to factors over and above those of age, sex, and race-ethnicity.

After taking into account differences in respondent cooperation rate that were explainable by age, sex, and race-ethnicity, we observed that there remained differences in participation rates depending on whether or not the respondent's household was willing to provide complete income information in the initial screening phase of the study. Thus, this willingness to provide complete income information appears to be an indicator of "general willingness to cooperate over time." In general, respondents for whom complete income information was available in the initial screening phase are more likely to continue to cooperate with this survey over time.

Where sample size allowed, this willingness-to-cooperate factor was added to the final post-stratification weighting by subdividing the target population in a given post-stratum cell into two categories: complete income information provided vs. complete income information not provided. The proportion of the post-stratum total population classified into these two groups was based on the actual distribution of sample respondents who participated in the baseyear of the survey. Weighting factors were then

determined so that the projected number of Round II, III, and IV participants within these two subclassifications would match those that existed as of the baseyear interview.

The adjustment proceeded as follows:

1. Each of the 48 post-stratification demographic subgroups was divided into two categories: (1) respondents for whom income information was obtained in the screener interview, and (2) respondents for whom such information was not obtained. If either category (1) or (2) contained fewer than 20 respondents, the categories were recombined and the subgroup was treated as a whole.

2. For each subdivided subgroup, a new "target population estimate" was calculated as follows:

$$\text{New target} = \text{Original target} \times \frac{\text{The sum of baseyear final weights for respondents in this portion the subgroup}}{\text{The sum of baseyear final weights for all respondents in the entire undivided subgroup}}$$

Thus, a new target was calculated for each portion of every divided subgroup. Since thirty-six subgroups qualified to be subdivided, seventy-two new population targets were obtained. In the twelve subgroups that were not subdivided, the original population targets were retained.

Post-stratification then continued as in Step 5 above. A_5 was calculated using either the original target or the new target as appropriate.

$$A_5 = \frac{\text{Target population for demographic subgroup or subdivided portion of the subgroup}}{\text{The sum of preliminary weights, } W_{4i}\text{, for all Round II (or III or IV) respondents in the subgroup or portion of the subgroup}}$$

Weighting the Military Sample

As in our description of the civilian sample weighting, we first describe the steps that were applied in weighting the NLS baseyear military sample interviews and then describe the modifications of this weighting process that were used in weighting the cases that participated in Rounds II, III, and IV.

NLS Baseyear Weights

Weighting for the NLS baseyear military sample involved three basic steps designed to accomplish the following objectives:

1. Correction for differential probability of selection for males and females.
2. Correction for differential interview completion rates.
3. Adjustment of weighted sample size to conform to known population size by Service, sex, birth year, and race.

Step 1. In the initial step, weights were assigned to each case on the basis of selection probability. For the ith repsondent, this weighting factor was

$W_{1i} = 1/f_i$ is the probability of selection for the ith respondent. For all males, this probability $f_i = 1/496.124$. For females $f_i = 1/78.851$.

Step 2. In the second step a completion rate adjustment factor was calculated on a PSU-by-sex basis as follows:

$$W_{2i} = \frac{\text{Selected individuals of same sex within ith respondent's PSU}}{\text{Number of completed cases of same sex within ith respondent's PSU}}$$

The factor W_{2i} was constrained to an upper limit of 1.5.

Step 3. For each respondent, a preliminary step three weight was calculated by multiplication of the weights from steps one and two

$$W^1_{3i} = W_{1i} \times W_{2i}$$

These preliminary weights were then summed within 80 (4 Service by 2 sex by 5 birth year by 2 race) post-strata. In the third step, final adjustment factors were then determined as the ratio of the actual population within the post-stratum to the sum of step three preliminary weights within the post-stratum.

$$A_{3i} = \frac{\text{(Population size within ith respondent's post-stratum)}}{\text{Sum of step three preliminary weights within ith respondent's post-stratum}}$$

The final weight assigned to the ith respondent was

$$W_i = W_{1i} \times W_{2i} \times A_{3i}$$

It should be noted that population sizes within the 80 post-strata were obtained from the actual sampling frame of all persons in the armed forces as of September 30, 1978 who would be between 14 and 21 as of January 1, 1979.[1] Although some information was available that would have allowed the use of finer post-stratification based upon ethnicity (Hispanic, non-Hispanic), this finer post-stratification was not implemented because it was felt that differences in questions used in the ethnicity classification of all "frame" elements were sufficiently different from the ethnicity classification used for sample respondents to preclude compatibility.

Table 4.4 shows the Population Counts for the 80 Post Strata used.

Weights for Rounds II, III, and IV

Modifications in weighting the portions of the sample selected through the military frame for the Rounds II, III, and IV were similar to those applied to the cases selected from the civilian frame. The final step of baseyear weighting for the military frame sample involved an 80-cell post-

---

[1] In those instances where there were no sample cases within a cell, no factor was applied and no collapsing was used.

TABLE 4.4

MILITARY SAMPLE POST-STRATIFICATION POPULATION TARGETS

Birth Year	Male		Female	
	Black	Other	Black	Other
	Army			
1957.....	21,053	44,131	1,701	4,912
1958.....	22,261	49,121	1,581	4,648
1959.....	19,480	45,364	1,504	4,259
1960.....	8,840	21,832	930	2,223
1961.....	309	2,027	41	61
	Navy			
1957.....	4,904	45,023	269	2,630
1958.....	5,130	45,340	244	2,248
1959.....	4,323	38,502	209	1,874
1960.....	2,359	21,106	104	833
1961.....	209	2,785	5	67
	Marine Corps			
1957.....	4,752	21,735	113	572
1958.....	4,976	22,855	107	679
1959.....	4,637	19,038	112	638
1960.....	2,552	10,389	83	384
1961.....	179	1,021	5	19
	Air Force			
1957.....	4,020	33,120	649	4,742
1958.....	3,707	30,045	552	4,342
1959.....	3,059	24,230	452	3,581
1960.....	1,622	11,528	201	1,516
1961.....	78	743	11	63

stratification adjustment on the basis of birth year, sex, and Service. Weighting of Round II, III, and IV respondents differed from the process used for the baseyear weighting only in the construction of factor A_{3i} which were based on preliminary sums of weights for Round II, III, and IV respondents rather than baseyear respondents.

Because of the explicit list nature of the military sampling frame, the use of a general population eligibility screen prior to baseyeyar interview was not required. Thus, the variable "full income information in screening" was not available as a means of refining the post-stratification process. However, in examining baseyear to later rounds' response rates within the 80 post-stratification cell structure, we observed that individuals who were not on active duty as of the date of scheduled interview tended to have a somewhat higher rate of non-participation. On the basis of this finding we decided to make use of active duty status in much the same fashion as we had used "full screener income" in the civilian frame portion of the sample. When certain minimum cell size conditions were satisfied, we subdivided an original post-stratum into two subclasses (still on active duty versus not on active duty). Target population sizes for the two subclasses that comprised the post-stratum were established by obtaining the active duty status (as of 1980 NLS interview) for all baseyear respondents. (This information was available for all baseyear respondents.) The final stage ratio adjustment factors A_5 were then determined separately for each of the two basic subclasses. Given the relatively small post-stratification cell sizes for the military sample, this refined procedure was applied in only 5 of the 80 post-strata.

CHAPTER 5

RELIABILITY OF RESULTS

Standard errors have been explicitly computed for a number of statistics based upon the entire NLS Youth Sample (Total, Civilian, and Military) and 11 sex and/or race subclasses; see Tables 5.3-5.7. Standard errors for other statistics (defined over the entire sample or the same 11 subclasses) may be approximated with use of the DEFT factors given in the tables below.

Approximate Standard Errors: Percentages

To approximate a standard error of a percentage the following formula is applicable

$$se(P) \triangleq DEFT \sqrt{\frac{P(100-P)}{n}}.$$

where:

se(P) = the approximate standard error for the percentage P

P = the sample percentage (ranging from 0 to 100)

n = the actual unweighted sample size for the demographic subclass from which the percentage was developed

DEFT = the appropriate DEFT factor for the particular demographic subclass and sample type from which the percentage was developed.

For example, for Round II the appropriate DEFT factor for estimating a standard error of the percentage of civilian Hispanic males who were high school dropouts is 1.32241 (see column one, row seven of Table 5.1). The calculated sample estimate (P) equals 22.19 percent and the unweighted sample size is 946. Therefore,

$$se(P) \triangleq 1.322441 \sqrt{\frac{22.19(100-22.19)}{946}}.$$

$$\triangleq 1.7893.$$

TABLE 5.1

DEFT FACTORS[1] FOR ROUND I

	Civilian			Military		Total	
	Proportions[2]	Proportions[3]	Means	Proportions	Means	Proportions	Means
All Youth	1.79472	1.67323	1.65700	1.32891	1.18853	1.72547	1.71282
Males	1.55111	1.43543	1.51606	1.44653	1.01622	1.46605	1.56808
Females	1.60979	1.53060	1.44894	1.12133	1.00976	1.58029	1.49720
Hispanics	1.60075	1.43778	1.44315	1.13513	1.15980	1.44342	1.45699
Blacks	1.44149	1.31829	1.40232	1.31345	1.03720	1.35303	1.43730
Whites	1.62143	1.51794	1.49627	1.32510	1.20744	1.58686	1.56996
Hispanic Males	1.32441	1.24799	1.22836	1.01244	1.01810	1.24321	1.22329
Hispanic Females	1.54178	1.39418	1.24167	1.01152	1.000	1.40353	1.25095
Black Males	1.27884	1.15796	1.18611	1.11046	1.000	1.19457	1.21378
Black Females	1.29725	1.21955	1.22241	1.01847	1.00392	1.24877	1.25243
White Males	1.40640	1.28436	1.38187	1.13528	1.02428	1.33775	1.45962
White Females	1.46954	1.40781	1.34025	1.13794	1.01491	1.46889	1.37581

[1]The statistics used to calculate the DEFT factors are proportion high school dropouts, proportion attending high school, proportion attending college, proportion high school graduates, proportion married, proportion employed, unemployment rate, labor force participation rate, enrollment status, proportion enrolled in government programs, mean years of school completed, mean number of children expected, means years of education expected to be completed.

[2]Averaged over all proportions.

[3]Averaged over proportions relevant to military personnel only.

TABLE 5.2

DEFT FACTORS[1] FOR ROUNDS II-III

	Round II		Round III	
	Proportions[2]	Means	Proportions[2]	Means
All Youth	1.73508	1.70754	1.59287	1.81714
Males	1.44194	1.39016	1.45569	1.50690
Females	1.59083	1.54383	1.45128	1.57655
Hispanics	1.42266	1.31283	1.26369	1.42826
Blacks	1.49075	1.43263	1.38582	1.40854
Whites	1.60776	1.60346	1.45166	1.65078
Hispanic Males	1.22197	1.32761	1.27616	1.39739
Hispanic Females	1.40408	1.29371	1.16159	1.38802
Black Males	1.29502	1.26152	1.24488	1.22916
Black Females	1.34253	1.29028	1.28492	1.24373
White Males	1.34290	1.24036	1.33082	1.33091
White Females	1.47183	1.49030	1.35830	1.48256

[1]The statistics used to calculate the DEFT factors are proportion high school dropouts, proportion attending high school, proportion attending college, proportion high school graduates, proportion married, proportion employed, unemployment rate, labor force participation rate, enrollment status, proportion enrolled in government programs, mean years of school completed, mean number of children expected.

[2]Averaged over all proportions.

TABLE 5.3

STANDARD ERRORS ROUND I TOTAL POPULATION

Characteristic	Demographic Group											
	All	Males	Females	Hispanic	Black	White	Male Hispanic	Female Hispanic	Male Black	Female Black	Male White	Female White
Prop. H.S. Dropouts	.00471	.00627	.00545	.01385	.00835	.00527	.01744	.01814	.01232	.00928	.00710	.00619
Prop. Attending H.S.	.00735	.00893	.01006	.01554	.01151	.00904	.02176	.02146	.01460	.01628	.01085	.01233
Prop. Attending College	.00597	.00729	.00778	.01037	.00784	.00710	.01230	.01460	.00919	.01119	.00862	.00947
Prop. H.S. Graduate	.00658	.00776	.00905	.01277	.01033	.00785	.01440	.01957	.01217	.01448	.00926	.01094
Mean Yrs. of School Completed	.029	.040	.038	.082	.057	.034	.100	.105	.061	.074	.046	.044
Mean Yrs. of School Expected Complete	.046	.059	.047	.108	.064	.055	.125	.117	.079	.079	.071	.055
Prop. Living In South	.02286	.02353	.02324	.05641	.04264	.02544	.04973	.06060	.04555	.04084	.02610	.02601
Mean #'s of Children Expected	.024	.027	.032	.058	.046	.028	.065	.070	.056	.055	.031	.037
Proportion Married	.00454	.00365	.00686	.01023	.00533	.00570	.00923	.01646	.00440	.00884	.00448	.00855

TABLE 5.4

STANDARD ERRORS ROUND I CIVILIAN POPULATION

Characteristic	Demographic Group											
	All	Males	Females	Hispanic	Black	White	Male Hispanic	Female Hispanic	Male Black	Female Black	Male White	Female White
Prop. H.S. Dropouts	.00477	.00645	.00547	.01412	.00839	.00532	.01773	01818	.01248	.00932	.00730	.00621
Prop. Attending H.S.	.00748	.00929	.01008	.01571	.01155	.00919	.02227	.02148	.01471	.01639	.01125	.01240
Prop. Attending College	.00609	.00755	.00781	.01056	.00805	.00724	.01264	.01463	.00959	.01123	.00891	.00950
Prop. H.S. Graduate	.00671	.00809	.00909	.01258	.01028	.00803	.01411	.01958	.01147	.01455	.00966	.01095
Mean Yrs. of School Completed	.030	.041	.038	.082	.058	.034	.102	.105	.063	.074	.048	.044
Mean Yrs. of School Expected Complete	.047	.061	.047	.110	.064	.056	.129	.117	.082	.080	.074	.055
Prop. Living In South	.02330	.02437	.02330	.05760	.04387	.02589	.05132	.06075	.04796	.04097	.02695	.02607
Mean #'s of Children Expected	.025	.028	.032	.059	.047	.029	.067	.070	.059	.055	.032	.037
Proportion Married	.00459	.00372	.00688	.01043	.00527	.00578	.00939	.01649	.00400	.00887	.00458	.00857
Proportion Employed	.00812	.01097	.01027	.01761	.01253	.01028	.01964	.02219	.01726	.01384	.01294	.01210
Unemployment Rate	.00719	.00885	.00958	.01905	.01829	.00743	.02146	.02666	.02411	.02220	.00973	.01016
?LF Participation Rate	.00819	.01000	.00998	.01864	.00922	.00487	.02173	.02342	.01254	.01326	.01195	.1208
Enrollment Status	.00772	.00959	.00987	.01721	.01052	.00927	.02075	.02382	.01396	.01491	.01155	.01189
Prop. Enrolled Since 1/79	.00554	.00691	.00640	.01451	.01345	.00500	.01875	.01714	.01428	.01756	.00753	.00693

TABLE 5.5

STANDARD ERRORS ROUND I, MILITARY POPULATION

Characteristic	Demographic Group											
	All	Males	Females	Hispanic	Black	White	Male Hispanic	Female Hispanic	Male Black	Female Black	Male White	Female White
Prop. H.S. Dropouts	.01228	.01372	.00562	.05297	.01569	.01504	.07077	.00000	.01831	.00000	.01663	.00736
Prop. Attending H.S.	.00186	.00206	.00000	.00000	.00758	.00170	.00000	.00000	.00890	.00000	.00188	.00000
Prop. Attending College	.00927	.01015	.01493	.03368	.02592	.01116	.03955	.06214	.03013	.03510	.01205	.01777
Prop. H.S. Graduate	.01541	.01719	.01539	.05513	.02952	.01804	.07203	.06214	.03436	.03510	.01979	.01829
Mean Yrs. of School Completed	.026	.029	.022	.121	.035	.032	.162	.146	.036	.046	.035	.028
Mean Yrs. of School Expected Complete	.086	.090	.097	.271	.146	.105	.354	.371	.155	.227	.109	.105
Prop. Living In South	.02691	.02623	.03182	.05415	.04883	.02988	.06796	.09063	.05086	.07455	.02900	.03498
Mean #'s of Children Expected	.048	.049	.081	.214	.140	.052	.280	.244	.155	.212	.053	.090
Proportion Married	.01659	.01759	.02577	.05505	.03106	.01991	.06669	.13563	.03225	.03984	.02120	.03044

TABLE 5.6

STANDARD ERRORS ROUND II TOTAL POPULATION

Characteristic	Demographic Group											
	All	Males	Females	Hispanic	Black	White	Male Hispanic	Female Hispanic	Male Black	Female Black	Male White	Female White
Prop. Not on Active Duty	.00135	.00233	.00073	.00189	.00341	.00151	.00375	.00034	.00612	.00245	.00266	.00082
Prop. H.S. Dropouts	.00513	.00612	.00638	.01354	.01010	.00587	.01945	.01808	.01478	.01113	.00694	.00764
Prop. In H.S. or Less	.00700	.00820	.00993	.01776	.01189	.00869	.02489	.02584	.01753	.01399	.00994	.01261
Prop. Attending College	.00635	.00760	.00796	.01194	.01096	.00733	.01164	.01875	.01011	.01524	.00900	.00945
Prop. H.S. Graduate	.00678	.00797	.00996	.01784	.01053	.00833	.02031	.02599	.01522	.01539	.00973	.01249
Prop. Living in South	.01727	.01791	.01801	.05594	.02963	.01953	.05835	.06046	.03354	.03017	.01979	.02091
Prop. Currently Married	.00539	.00504	.00720	.01027	.00774	.00654	.01091	.01573	.00706	.01136	.00626	.00855
Prop. Employed at Present	.00848	.01029	.01067	.01741	.01408	.01012	.02390	.02476	.02035	.01728	.01257	.01259
Prop. Unemployed	.00607	.00840	.00805	.01889	.01669	.00626	.02213	.02746	.01969	.02507	.00957	.00827
Prop. In Labor Force	.00813	.00943	.01022	.01323	.01331	.00990	.01809	.02072	.01892	.01525	.01142	.01220
Prop. Gvt. Training Participant	.00624	.00687	.00770	.01949	.01648	.00683	.02499	.02341	.01917	.02112	.00765	.00809
Average Number of Children	.00754	.00490	.01165	.01997	.01517	.00851	.01521	.03458	.01774	.02432	.00499	.01341
Average Highest Grade	.02943	.04042	.03821	.07213	.05913	.03322	.13401	.10420	.06845	.07529	.04592	.04588
Prop. Currently Enrolled	.00830	.00882	.01169	.02127	.01304	.01026	.02705	.02785	.01730	.01584	.01080	.01463

TABLE 5.7

STANDARD ERRORS ROUND III TOTAL POPULATION

Characteristic	Demographic Group											
	All	Males	Females	Hispanic	Black	White	Male Hispanic	Female Hispanic	Male Black	Female Black	Male White	Female White
Prop. not on Active Duty	.00184	.00332	.00074	.00371	.00459	.00209	.00705	.00213	.00740	.00306	.00389	.00074
Prop. H.S. Dropouts	.00541	.00723	.00564	.01566	.00961	.00609	.02259	.01860	.01374	.01137	.00794	.00678
Prop. In H.S. or Less	.00550	.00764	.00822	.01478	.00978	.00681	.02707	.01788	.01616	.01202	.00923	.01059
Prop. Attending College	.00715	.00984	.00825	.01161	.01018	.00852	.01701	.01705	.01089	.01528	.01157	.00996
Prop. H.S. Graduate	.00657	.00847	.00954	.01619	.01171	.00793	.02270	.02242	.01754	.01601	.01025	.01188
Prop. Living in South	.01699	.01690	.01827	.05483	.02821	.01918	.05436	.06315	.03100	.02908	.01863	.02117
Prop. Currently Married	.00562	.00581	.00839	.01028	.00801	.00714	.01587	.01636	.00913	.01140	.00747	.01054
Prop. Employed at Present	.00672	.00889	.00941	.01516	.01377	.00756	.02034	.02150	.01692	.02024	.01056	.01071
Prop. Unemployed	.00539	.00802	.00700	.01279	.01641	.00544	.01837	.01678	.02015	.02243	.00902	.00751
Prop. In Labor Force	.00654	.00743	.00904	.01358	.01157	.00774	0.01415	0.02447	0.01360	0.01766	0.00901	0.01054
Prop. Gvt. Training Participant	.00304	.00392	.00336	.00796	.00944	.00311	.01181	.00957	.01227	.01159	.00411	.00358
Average Number of Children	.00902	.00669	.01359	.02617	.01590	.00985	.01981	.04582	.01929	.02488	.00719	.01523
Average Highest Grade	.03183	.04259	.03802	.07645	.05611	.03612	.13804	.10275	.06343	.07066	0.04780	0.04521
Prop. Currently Enrolled	.00741	.00926	.01007	.01974	.01031	.00899	.02929	.02015	.01677	.01301	.01104	.01262

To approximate the standard error of the corresponding projected population total (NP/100), one calculates:

$$se(NP/100) \triangleq N[se(P)/100]$$

where:

se(NP/100) = the approximate standard error of the projected population total coresponding to a percentage P within a particular demographic subclass and sample type

N = the appropriate projected total population base for the particular demographic subclass and sample type

For example, the projected total population base for civilian Hispanic males is 1,030,861. The projected number of civilian Hispanic male high school dropouts is equal to NP/100 or 1,030,861 . 22.19/100 = 228,748. Thus, the approximate standard error for the total number of civilian Hispanic male high school dropouts is:

$$se(NP/100) \triangleq (1{,}030{,}861)\ (1.7893/100)$$

$$\triangleq 18{,}445.1959$$

Approximate Standard Errors: Means

One can compute approximate standard errors for means as follows:

$$se(\bar{X}) \triangleq DEFT \sqrt{\frac{s^2}{n}}$$

where:

$se(\bar{X})$ = the approximate standard error of the mean

DEFT = the appropriate DEFT factor for the particular demographic subclass and sample type from which the mean was developed

s^2 = the weighted element variance computed for the demographic subclass and sample type from which the mean was developed

n = the unweighted sample size for the particular mean.

For example, for Round I the DEFT factor for all Hispanics in the military is 1.1598 (see column six, row four of Table 5.1). To approximate the standard error of the mean number of years of education completed by this subclass, where the weighted element variance is .72955 and the sample size is 77, one computes:

$$se(\bar{x}) \triangleq 1.1598 \sqrt{\frac{.72955}{77}}$$

$$\triangleq .1131$$

CHAPTER 6

POTENTIAL IMPACT OF NONRESPONSES

Given the typical rates of completion currently obtained in U.S. household surveys, completion rates for the various stages of the study are quite high. Depending upon the method used to combine the completion rates at screening, and later interview, the overall response rate ranges from 75 to over 80 percent. In spite of this relatively high level of completion we still feel obliged to examine the potential impact on the survey results of noncooperation at the various stages of the survey process.

The impact of nonresponse on the survey results (in this case the NLS interview responses) depends on the product of two factors: (1) the proportion of the sample which is nonresponding, and (2) the magnitude of the difference between responders and nonresponders. A large difference between responders and nonresponders combined with a small percentage of nonresponse will produce minimal impact on the survey results. Similarly, even a high rate of nonresponse will have minimal impact on the survey results so long as the difference between responders and nonresponders is small. Since the impact of nonresponse is a function of the product of the two factors, if either factor is small the impact will be minimal.

The major difficulty in estimating the impact of nonresponse derives from the fact that although the rate of nonresponse (factor 1) is usually well known, the magnitude of the difference between responders and nonresponders (factor 2) can seldom be known with any confidence. Several methods for estimating the characteritics of nonresponders have been suggested in the literature but since non is entirely satisfactory one can assume that the impact of nonresponse is negligible only when the rate of nonresponse (factor 1) is very small.[1]

---

[1]See Leslie Kish, Survey Sampling, N.Y.: John Wiley and Sons, 1965, pp. 557-562; Ronald Anderson, Judith Kasper, Martin R. Frankel, and Associates, Total Survey Error, San Francisco: Jossey-Bass, 1979.

For completed cases in Rounds I-IV of the NLS corrections for nonresponse were introduced in the weighting process. A correction for screener nonresponse was made at Step 2 and for baseyear interview nonresponse at Step 3 in the civilian sample. Screener nonresponse was adjusted at the segment while baseyear interview nonresponse was adjusted at the PSU level separately for each of 8 design cohorts. Thus, in correcting for screener nonresponse we assumed no difference between responding and nonresponding households in the same segment. In the correction for baseyear interview nonresponse we assumed no difference between responders and nonresponders who were in the same PSU and design cohort. Although we are not in a position to test these assumptions, they can be defended on theoretical grounds.

As described in Chapter 4, a correction for nonresponse to Rounds II, III, and IV was made in the calculation of the post-stratification adjustment factors A_5. The use of age, sex and race-ethnicity in the post-stratification adjustment was based on theoretical considerations as well as previous survey experience which suggests that these factors are related to many survey variables of interest. Further subdivision of the post-stratification cells on th basis of whether the respondent's household provided complete income information in the screening interview was based on an empirical relationship observed between this variable and participation rates. Thus, we attempted to provide, where cell size permitted, a somewhat finer correction for differences in nonresponse rates between Round I and later rounds.

In the military sample, weighting correction for nonresponse was made at the post-stratification weighting step by introducing a further cell subdivision based on whether the respondent was on active duty at the time of the baseyear interview.

It is important to point out that all surveys with less than 100 percent response are forced to make assumptions about the characteristics of nonresponders. A unweighted (or self-weighting) survey assumes that all nonresponders taken as a group are like all responders taken as a group. We found it preferable on both theoretical and empirical grounds to make these assumptions at a desaggregated level, i.e., at the segment, PSU, or design cohort level, or based upon such characteristics as whether complete income information was provided in the screening interview or whether the respondent was on active duty at the time of the baseyear interview. We feel that these more exacting corrections treat nonresponse with the care it deserves.[1]

Having made the various adjustments for nonresponse described above, we carried out a simple but fairly standard test to examine differences between the final completed sample and the sample various stages in the overall survey process. Specifically, we compared the sample distribution for several variables among all screener respondents and those respondents who completed the baseyear interview (civilian sample only).[2] Table 6.1 shows the result of this comparison. The rationale for making this comparison is to see whether any significant changes in sample composition appear from one stage to the next as nonresponse attrition reduces the number of survey participants. As Table 6.1 shows, the sample of all screener respondents is almost identical to the sample of all baseyear respondents in terms of the distribution across welfare status, household size, language used in the screener, and type of locality.[3]

---

[1]Needless to say, we also did all we could through field operations to keep the response rate high.

[2]Of necessity, this analysis is restricted to respondents whose households provided answers to screener questions.

[3]Since sample weighting was only applied to respondents who participated in the baseyear interview, the comparison between the screener sample and the baseyear sample is shown on an unweighted basis.

TABLE 6.1

DISTRIBUTIONS (UNWEIGHTED) OF SCREENER VARIABLES[a]
FOR COMPLETED SCREENER RESPONDENTS AND
BASEYEAR RESPONDENTS

(Civilan samples only)

Variable	Distribution among	
	Screener Respondents	Baseyear Respondents
Welfare Status:		
Yes	16.4	16.5
No	83.6	83.5
Persons in Household:		
1-2	15.0	14.3
3	15.5	15.4
4	19.5	19.6
5	19.2	19.4
6 or more	30.7	31.2
Language Used in Screener:		
English	91.2	89.5
Spanish/Other	8.8	10.5
Locality:		
Non-rural	94.3	94.6
Rural	5.7	5.3

[a]Includes only those cases where information was obtained in screening.

Table 6.1 shows that nonresponse attrition did not produce any significant changes in sample composition at later stages of the survey process, at least on the variables measured. Although this test comparison failed to reveal any differences in sample composition due to nonresponse, the limited number of screener variables available for comparison should induce caution in concluding that the distribution of characteristics among nonrespondents would be identical to the distribution among respondents.

In conclusion we would like to point out that the best overall evidence that the impact of nonresponse must be small is the relatively high response rate attained. Whatever difference there may be between the test scores of responders and nonresponders is diluted by the fact that there are four times as many responders as nonresponders. Thus, in the case of proportions a diffference between responders and nonresponders of 5 percentage points would produce only 1 percentage point difference in the overall results. A difference of 10 percentage points would produce an overall change of 2 points.