

6803

68063

THE NATIONAL YOUTH SURVEY

MH 27552

LEAA 78-JN-AX-0003

Delbert S. Elliott, Principal Investigator (NIMH & LEAA)
Suzanne S. Ageton, Co-Principal Investigator (NIMH)
Brian A. Knowles, Assistant Project Director (LEAA)
Tim Brennan, Investigator (LEAA)
Rachelle J. Canter, Investigator (NIMH)
David Huizinga, Investigator (NIMH)

PROJECT REPORT NO.1

SAMPLE DESIGN OF THE NATIONAL YOUTH SAMPLE

by

D. HUIZINGA

August, 1978

Behavioral Research Institute

Boulder, CO

68063

MAY 23 1980

SAMPLE DESIGN
OF THE
NATIONAL SURVEY OF YOUTH¹

ACQUISITIONS

The 1977 National Survey of Youth, conducted by the Behavioral Research Institute, was based on a multistage area probability sample of households in the continental United States. The sampling units at each stage of selection are (1) primary sampling units (PSU's), which are large geographical areas; (2) secondary sampling units (SSU's), which are smaller geographical areas, within PSU's; (3) segments, which are portions of SSU's; and (4) households within segments. Extensive stratification was used in the first two stages of selection. The probabilities of selection for each stage were established to provide a self-weighting sample (i.e., every household had the same probability of inclusion in the sample).

A primary sampling unit was defined as an entire Standard Metropolitan Statistical Area (SMSA)² or a county or group of contiguous counties containing a minimum of 5000 households. To achieve this minimum, counties with less than 5000 households were combined with neighboring counties to form a PSU meeting this requirement. The approximately 3100 counties of the continental United States were grouped into 2009 PSU's, and accounted for an estimated 70,940,900 households.

¹A technical description of the Sample Design is available from the Behavioral Research Institute, Boulder, Colorado.

²For the purpose of this frame, Standard Metropolitan Statistical Areas or SMSA's are those so designated by the Census Bureau for the 1970 census, with the exception of Census Bureau defined SMSA's in the New England Census Division. In New England, the Census Bureau uses townships and other local boundaries to create SMSA's. To be consistent with the remainder of the country, and because the updated number of households was available by county, SMSA's were redefined in terms of counties. A New England SMSA was taken to be a county or group of contiguous counties, each of which contained one or a portion of a Census Bureau defined SMSA.

To select PSU's for inclusion in the sample, a replicated zone sampling procedure was employed. Using this procedure, stratification of the PSU's is implicit and results from an ordering of the PSU frame. The frame was first divided into an SMSA section and a non-SMSA section. This provided an "urban-rural" split of the frame. Within each of the SMSA and non-SMSA sections, the PSU's were grouped by census division and the census divisions ordered in a serpentine fashion³, thus insuring geographic stratification of the sample. Within the South Atlantic Division, the PSU's were further divided into those whose population was less than 20% black and those whose population was more than 20% black⁴. Finally, within each of the resulting sections of the PSU frame, the PSU's were placed in increasing or decreasing order on the basis of number of households, with the order alternating between adjacent sections. This latter arrangement results in a stratification on the basis of "size of place", where size is measured by number of households.

Following the ordering of the PSU frame, one PSU was selected from each of 76 zones or strata. The selection procedure was designed so that the probability of a particular PSU being selected was proportional to its size, where the measure of PSU size was the number of households contained in the PSU.

Within each selected primary sampling unit, secondary sampling units (SSU's) were taken to be Block Groups (BG's) or Enumeration Districts (ED's) as defined by the Census Bureau for the 1970 Census, with

³The order of Census Divisions was: New England, Mid-Atlantic, East North Central, Mountain, Pacific, West South Central, East South Central, and South Atlantic.

⁴This is the only Census Division in which stratification on the basis of ethnicity proved feasible within the limits of the sampling design.

the requirement that each BG or ED must contain at least 60 households. Any BG or ED not meeting the above minimal requirement was combined with neighboring BG's or ED's to reach the 60 household minimum, and this combined BG/ED was taken as one secondary sampling unit. The number of households in a BG/ED was taken from the "first county data" of the 1970 Census.

A probability proportional to size, systematic sampling procedure was employed to select BG's and ED's from the sample frame constructed for each PSU. The measure of size for each BG or ED was the number of households contained in the BG/ED. Six BG's or ED's were selected from each previously selected PSU.

To take advantage of the systematic draw used in selecting BG/ED's, prior to selection, the BG's and ED's from each PSU were ordered in the following manner. The SSU's were initially arranged with all BG's first, followed by all ED's. Within each of these two divisions, the units were sorted into county groups and the county groups arranged by total county size (number of households), in decreasing order. Within each county, the units were sorted into Minor Civil Division or Census Civil Division (MCD/CCD) groups and these MCD/CCD groups arranged by decreasing MCD/CCD size order. Within each MCD/CCD group, the units were sorted by ascending census tract number and within census tracts by ascending block group number or enumeration district number. For PSU's which contained a sufficiently large black population that an entire interval or intervals of the systematic sampling procedure could be covered by BG's or ED's whose population was more than p% black, the SSU frame was first divided into two segments, those containing p% or less black and those with more than p% black. For rural areas $p=20$ and for highly urban areas

p=50. The above ordering process was then applied independently to each of these two segments.

This ordering process provides an implicit stratification of the SSU frame based on (1) urban/rural characteristics; (2) general size in terms of the number of households in the local area containing a BG or ED; (3) geographical location; and (4) for some PSU's, ethnic distribution.

Within BG's or ED's selected during the second stage, contiguous geographical areas of approximately 100 households were created. One segment was selected from each BG or ED with probability proportional to size. In some instances, the selected segment from each BG or ED contained several hundred households. This resulted from either population growth in the segment since the 1970 census (1970 census data was used for "in-house" segmentation) or because one block, the smallest segment for which published information was available, contained several hundred households. In this case, an additional stage of sampling was employed. These large segments were field counted and sub-segments of approximately 100 households were created. One of these sub-segments was then selected with probability proportional to size.

Segments selected during the third or fourth stage of sampling were completely enumerated. From the resulting listings of households, a systematic sample of households was selected. The sample rates within segments were determined so that the entire sample of households was self-weighting.

The above sampling procedure resulted in the listing of 67,266 households in 456 segments. From this listing, approximately 8000 households were selected for inclusion in the sample. All 11 through 17 year

old youth living in the selected households were the eligible respondents for the study⁵. An attempt was made to interview each youth and one of the youth's parents.

Of the selected households, 379 were vacant. In 59 of the households, no occupant was ever found at home. Among households in which an occupant was located, 6117 households did not contain eligible youth, and in 34 households, respondents refused to participate in the study and would not provide information about household members. In 353 of the households containing youth in the appropriate age range, parents refused to allow their youth to participate in the study. In most cases, these parents did indicate the number of eligible youth living at home and it is estimated that these households contained 610 eligible youth. The remaining 1056 households contained 1765 eligible youth. Of these, 19 refused to participate in the study and 20 were considered ineligible for reasons such as mental retardation. Interview schedules were completed for the remaining 1726 youth.

Parents of youth respondents were also interviewed. Of 1056 potential parent respondents, 17 refused to participate in the study, although allowing their youth to participate.

The completion rates described above are given in tabular form in Table 1.

⁵The approximately 8000 household sample size was determined to provide a sample of approximately 2100 eleven through seventeen year old youth. This number of households was based on assumptions of a 7% vacant household rate, a 75% completion rate of occupied households, and an average of 0.38 11-17 year old youth per household.

TABLE 1
COMPLETION RATES RESULTING
FROM THE NATIONAL SURVEY OF YOUTH

Households

Number of Households in the Sample	7998
Households Not Interviewed	
Vacancies	379
Not at Homes	59
Refusal by Parents at the door	353
Eligible youth live in household.	
Refusal by adult respondent	34
Whether eligible youth live in household is unknown.	
Households Interviewed	
Households with no eligible youth	6117
Households with eligible youth	1056

Eligible Youth Respondents

Estimated number of youth not interviewed because of parent refusal.	610
Number of youth refusing to participate in study.	19
Number of youth considered inappropriate for inclusion in the study.	20
Number of youth that completed interviews.	1726
Total number of eligible youth	2375
Completion rate among eligible youth respondents.	73%

The age, sex and ethnicity characteristics of the youth sample are presented in Table 2. In that table, they are contrasted with recent estimates provided by the Census Bureau for the total 11-17 year old youth

population. The age, sex, and ethnicity of eligible youth not interviewed (for those youth for whom such information is known) is presented in Table 3. As indicated in the table, the loss rate from any particular group appears, in general, to be proportional to that group's representation in the population. Thus, on the basis of demographic characteristics, the sample appears to be representative of the total 11-17 year old youth population.

TABLE 2

DEMOGRAPHIC CHARACTERISTICS OF THE NATIONAL YOUTH
SURVEY SAMPLE AND OF THE TOTAL 11-17 YEAR OLD POPULATION

	<u>SAMPLE</u>	<u>CENSUS BUREAU*</u> <u>POPULATION ESTIMATES</u>
ETHNICITY		
Anglo/Chicano	83%	84%
Black	15%	14%
Other	02%	02%
SEX		
Male	53%	51%
Female	47%	49%
AGE		
11	13%	13%
12	14%	14%
13	16%	14%
14	14%	15%
15	16%	15%
16	14%	14%
17	13%	15%

* Source: Population Estimates and Projections, Series P-25, No. 643.
Bureau of the Census, 1977

TABLE 3

DEMOGRAPHIC CHARACTERISTICS OF ELIGIBLE YOUTH NOT INTERVIEWED

	AGE							SEX		ETHNICITY			
	11	12	13	14	15	16	17	Male	Female	Anglo	Black	Chicano	Other
PARENT REFUSAL to allow youth to participate	67	82	71	65	61	69	70	188	186	271	50	19	16
YOUTH REFUSAL	1	1	0	4	0	6	2	10	5	11	0	0	0
YOUTH INAPPROPRIATE for interviewing	3	6	5	1	4	4	3	13	9	4	2	1	0
TOTAL	71	89	76	70	65	79	75	211	200	286	52	20	16

END