DATA EVALUATION
AND METHODS RESEARCH

Series 2
Number 87

National Survey of Family Growth, Cycle II:

Sample Design, Estimation Procedures, and Variance Estimation

This report describes the procedures used to select the sample, estimate population parameters, and estimate sampling variances for Cycle II of the National Survey of Family Growth.

DHHS Publication No. (PHS) 81-1361

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
Office of Health Research, Statistics, and Technology
National Center for Health Statistics
Hyattsville, Md. February 1981

Library of Congress Cataloging in Publication Data

Grady, William R.
National survey of family growth, cycle II.

(Vital and health statistics : Series 2, Data evaluation and methods research ; no. 87) (DHHS publication ; no. (PHS) 81-1361)
Includes bibliographical references.
1. Family size–United States–Statistical methods. 2. Fertility, Human–United States. 3. Birth control–United States. I. Title. II. Series: United States. National Center for Health Statistics. Vital and health statistics : Series 2, Data evaluation and methods research ; no. 87. III. Series: United States. Dept. of Health and Human Services. DHHS publication ; no. (PHS) 81-1361.
RA409.U45 no. 87 [HQ766.5.U5] 312'.0723s [304.6'3]
ISBN 0-8406-0199-9 80-20552

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402

NATIONAL CENTER FOR HEALTH STATISTICS

STATISTICAL METHODS STAFF, DATA SYSTEMS

DIVISION OF VITAL STATISTICS

COOPERATION OF WESTAT, INCORPORATED

In accordance with specifications established by the National Center for Health Statistics, Westat, Inc. of Rockville, Maryland, under a contractual agreement, participated in the design and selection of the sample and carried out the data collection.

Vital and Health Statistics-Series 2-No. 87

DHHS Publication No. (PHS) 81-1361
Library of Congress Catalog Card Number 80-20552

PREFACE

This report presents a detailed description of the sample design, estimation procedures, and variance estimation method used in Cycle II of the National Survey of Family Growth. The survey was designed and conducted by Westat, Inc. of Rockville, Maryland under a contractual arrangement with the National Center for Health Statistics (NCHS). The sampling plan was developed under the supervision of Joseph Waksberg of Westat, Inc., in consultation with E. Earl Bryant and William F. Pratt of NCHS.

Much of the report is based on survey specification documents and the final report prepared by Westat, Inc., and on internal NCHS memoranda. Parts of the report are also largely based on a previous report, prepared by Dwight K. French, on the Cycle I survey. Mr. French was the primary resource person for methodological questions.

In addition to the usual internal review, NCHS policy stipulates that methodological reports are to be given a peer review for technical merit and readability by one or more persons who are familiar with the subject matter area but who are not involved in producing the report. George A. Schnack and Peter W. Ries, both of NCHS, carried out the peer review of this report and provided many constructive suggestions.

CONTENTS

LIST OF TEXT FIGURES

LIST OF TEXT TABLES

SYMBOLS

Data not available	- - -
Category not applicable	. . .
Quantity zero	-
Quantity more than 0 but less than 0.05	0.0
Figure does not meet standards of reliability or precision	*

NATIONAL SURVEY OF FAMILY GROWTH, CYCLE II: SAMPLE DESIGN, ESTIMATION PROCEDURES, AND VARIANCE ESTIMATION

William R. Grady, Division of Vital Statistics

INTRODUCTION

The primary mission of the National Center for Health Statistics is to collect and publish data relating to the health of the U.S. population. In carrying out this mission the Center data on vital events registered in the United States are collected, inventories of health facilities and manpower are conducted, and probability sample surveys based on household interviews, health examinations, and medical records are conducted. Data collection programs are supplemented by research projects that investigate new techniques of data collection and evaluate currently operating programs.

In response to the need for current information on fertility and family planning and their effects on population growth, the National Survey of Family Growth was established in 1971 as an integral part of the Center's Division of Vital Statistics. The purpose of the survey is to collect data relating to natality and the process of family formation and dissolution. The National Survey of Family Growth was designed as a cyclic survey; that is, data are collected every few years by means of a sample survey. The first cycle of the survey was conducted in 1973, the second in 1976; Cycle III is scheduled to be conducted in 1981.

The target population of Cycles I and II of the National Survey of Family Growth consisted of the civilian household population of women aged 15-44 years, living in the conterminous United States, who were currently married, previously married, or never-married mothers with offspring living in the household at the time of interview. Data were collected from a probability sample by means of personal interviews lasting about an hour. The interviews provided information on fertility trends and differentials, family planning practices, sources of family planning advice and services, use and effectiveness of contraceptives, and aspects of maternal and child health that are closely related to family planning.

The sample design and data collection for Cycle I were contracted to the National Opinion Research Corporation of the University of Chicago. A complete description of that survey can be found in another report.[1] The sample design and data collection for Cycle II were contracted to Westat, Inc. of Rockville, Maryland. The sample consisted of 9,470 eligible women, of whom 8,611 (90.9 percent) were interviewed. All interviews were conducted between January and September 1976 and centered on May 12. The sample design employed to select the women is described in detail in this report. Also described are the techniques used to estimate population parameters and the procedures used to obtain provisional sampling variances.

DESIGN SPECIFICATIONS

The development of an efficient sample design must take into account the primary

survey objectives, the amount of funds available, logistical problems, time limitations, estimates of population characteristics and distribution, and operating costs. These requirements dictated a stratified multistage probability sample design for Cycle II, based essentially on the following set of specifications:

1. The target population was defined to be the civilian household population of women living in the conterminous United States who were 15-44 years of age and either
   (a) currently married,
   (b) previously married, or
   (c) never-married mothers with one or more children born to them currently living in the household.
2. The sample would consist of approximately 10,000 women, selected from an initial probability sample of households. It would include about 4,000 black women and 6,000 women of other races. Trained field staff were to conduct a screening interview with a responsible member of each sample household to determine if there were any eligible women (the screener questionnaire is reproduced in appendix II). When a household contained one eligible woman, she was included in the sample. In households with more than one eligible woman, the staff member would randomly select one woman for the sample.
3. Data were to be collected from the sample women (no proxy respondents were to be accepted) by means of personal interviews lasting an average of 1 hour.
4. All interviewers would be female.
5. The interviewer would collect information on fertility, family planning practices, sources of family planning services, and related maternal and child health practices.
6. The fieldwork would be completed in approximately 6 months.
7. The target interview completion rate for the total sample and both major subsamples by race was 90 percent of the expected number of women from all sample households (i.e., screener and interview nonresponse combined would ideally be no more than 10 percent).
8. The contractor, in cooperation with the National Center for Health Statistics (NCHS), would design and implement procedures to measure and control the quality of data collection and data preparation.

SAMPLE DESIGN

Summary

The sample design for Cycle II of the National Survey of Family Growth (NSFG) was a five-stage probability design that incorporated a supplementary sample of new (post-1969) housing units. This section provides an overview of the design, and it is followed by sections discussing each stage in detail.

The counties and independent cities that comprise the total land area of the conterminous United States were combined to form a frame of primary sampling units (PSU's). During the first stage of the sampling process, which involved extensive stratification, 79 PSU's were chosen from this frame. Census enumeration districts (ED's) were then identified for each of the selected primary sampling units; during the second stage, these enumeration districts were stratified according to the percent of their population that was black, and a systematic sample was drawn. The rate at which enumeration districts were sampled varied by second-stage strata. These differential sampling rates were the first step in producing the desired racial composition of the final sample of women.

The third stage required the identification of area segments (groups of houses) within sample enumeration districts and the random selection of one segment from each district. All sample dwelling units (DU's) built prior to 1970 were selected this way. However, segments of new construction units (post-1969) were drawn from a supplementary sample of building permits

selected from building permit offices in the 79 primary sampling units chosen during the first stage of the sampling process.

The fourth stage resulted in the selection of households within sample segments. In segments from enumeration districts with a 10-percent or greater black population, households of races other than black were selected at a rate that was lower than for black households. These different rates of selection were obtained through a subsampling process (to be described later in this report) and ensured that the desired proportions of black and other women would be included in the final sample. At each sample household an interviewer attempted to complete a Household Screener and identify women eligible for interview. When more than one potential respondent was found during this fifth-stage operation, all eligible women in the household were listed and one was randomly selected.

First-Stage Selection of Primary Sampling Units

Sampling frame.—The counties and independent cities of the conterminous United States have been grouped into about 1,900 primary sampling units (PSU's) by the U.S. Bureau of the Census for its Current Population Survey. These census PSU's were used by Westat, Inc., with minor modifications, as the sampling frame for the National Survey of Family Growth. Each PSU consisted of an individual county or a grouping of contiguous counties. Where standard metropolitan statistical areas (SMSA's) were defined, the counties comprising each SMSA were used as a primary sampling unit.[a] Seventy-nine PSU's were selected in the first stage of the sampling process.

The sample contains 25 self-representing (included with certainty) primary sampling units composed of 18 separate SMSA's. These include the 14 largest SMSA's, with 1970 populations of more than 1,850,000, and 4 slightly smaller ones that were made self-representing because they could not easily be placed in strata with other primary sampling units. Four of the largest SMSA's, which contained several counties each, were subdivided into smaller primary sampling units, each with an average population of about 2,500,000.

Selection of nonself-representing PSU's.—Prior to the selection of the remaining PSU's, all nonself-representing (probability of selection less than unity) PSU's were grouped into 35 strata of approximately equal size. Nineteen strata contained SMSA's and 16 contained non-SMSA areas.

The more than 200 nonself-representing SMSA's were sorted into 19 strata of about 4,000,000 persons each. In defining the appropriate stratum, four characteristics were considered. In order of priority, these characteristics were the following:

1. Region of the country (see table A).
2. Percent change in population between 1960 and 1970.
3. Percent of the population employed in manufacturing.
4. A Socioeconomic Index (developed by Westat, Inc.) that was based on the percent of the population that is white, the percent of households that either lacks plumbing or is overcrowded, and the dependency ratio.[b]

The nearly 1,700 remaining PSU's were then sorted into an additional 16 strata. The criteria in order of priority for this stratification were as follows:

1. Region of the country.
2. Percent of the population living in urban areas.

[a]SMSA definitions used are those of 1970 (see appendix I). No attempt was made to have the sample reflect new SMSA's created since 1970 or changes in SMSA definitions since that date. In New England, where SMSA's are not defined in terms of counties, metropolitan state economic areas (MSEA's), which are comprised of counties, were used as primary sampling units instead of SMSA's. Thus primary sampling units were in all cases defined in terms of complete counties.

[b]The dependency ratio is the ratio of the number of people who are under 18 or over 64 years of age to the number of people who are between the ages of 18-64, inclusively.

Table A. States in conterminous United States, by geographic region

Region	States		
Northeast	Connecticut	New Hampshire	Pennsylvania
	Maine	New Jersey	Rhode Island
	Massachusetts	New York	Vermont
North Central	Illinois	Michigan	North Dakota
	Indiana	Minnesota	Ohio
	Iowa	Missouri	South Dakota
	Kansas	Nebraska	Wisconsin
South	Alabama	Kentucky	South Carolina
	Arkansas	Louisiana	Tennessee
	Delaware	Maryland	Texas
	District of Columbia	Mississippi	Virginia
	Florida	North Carolina	West Virginia
	Georgia	Oklahoma	
West	Arizona	Montana	Utah
	California	Nevada	Washington
	Colorado	New Mexico	Wyoming
	Idaho	Oregon	

3. Percent change in population between 1960 and 1970.
4. The Westat Socioeconomic Index.

This stratification insured the proportionate representation of women by region and socioeconomic status, and may also have reduced sampling error.

Following the stratification process, PSU's were selected in two stages.[c] In the first stage, 1 PSU was selected from each of the 35 strata with a probability proportionate to size. The second-stage selection of the remaining 19 PSU's was then accomplished in 3 steps. First the 35 strata were combined into 19 superstrata. This recombination was done in such a way as to produce strata that were, as far as possible, homogeneous with respect to region, metropolitan composition (SMSA vs. non-SMSA), rate of population change between 1960 and 1970, and the percent of the population that was black. Four of the resultant superstrata contained a single stratum each, 14 contained 2 strata, 1 contained 3 strata. The second step was the selection, with a probability proportionate to size, of a single stratum within each superstratum. The final step was the selection, also with a probability proportionate to size, of 1 PSU from each of the 19 strata. Since the second-stage selection of 19 PSU's was done independently of the first-stage selection of 35 PSU's (i.e., sampling was done with replacement), it was possible for PSU's to be selected twice; this was the case for 1 PSU. All selected PSU's are shown in table B, and a diagram of the first stage of the sampling process is shown in figure 1.

Second-Stage Selection of Enumeration Districts

Sampling rates.—Because the final sample of women for the National Survey of Family Growth was intended to consist of approximately 4,000 black women and 6,000 women of other races, different sampling rates were required for the 2 groups. Given a 100-percent response rate, the approximate sampling fraction needed to produce the required number of black women was 1 in 950, while the corresponding

[c]Two stages were used because a 50-PSU design existed prior to the decision to change to a 79-PSU design. It was thus more efficient to go through a second stage of selecting additional PSU's than to select a new independent sample.

Table B. Sample primary sampling units (PSU's), by type

Self-representing SMSA's	
Anaheim-Santa Ana-Garden Grove, Calif.	Los Angeles-Long Beach, Calif.
Baltimore, Md.	Newark, N.J.
Boston-Lowell-Lawrence, Mass. (MSEA)	New York, N.Y. (represented by 5 PSU's)
Buffalo, N.Y.	Paterson-Clifton-Passaic, N.J.
Chicago, Ill. (represented by 2 PSU's)	Philadelphia, Pa.-N.J. (represented by 2 PSU's)
Cleveland, Ohio	Pittsburgh, Pa.
Detroit, Mich. (represented by 2 PSU's)	San Francisco-Oakland, Calif.
Houston, Tex.	St. Louis, Mo.-Ill.
Kansas City, Mo.-Kans.	Washington, D.C.-Md.-Va.
Nonself-representing SMSA's	
Allentown-Bethlehem-Easton, Pa.-N.J.	Madison, Wis.
Birmingham, Ala.	Miami, Fla.
Bridgeport-Stamford-Norwalk, Conn. (MSEA)	Milwaukee, Wis.
Chattanooga, Tenn.-Ga.	Minneapolis-St. Paul, Minn.
Dallas, Tex.	New Orleans, La.
Denver, Colo.	Oklahoma City, Okla.
Erie, Pa.	Omaha, Nebr.-Iowa
Fayetteville, N.C.	Salinas-Monterey, Calif.
Grand Rapids, Mich.	San Angelo, Tex.
Hartford-New Britain-Bristol, Conn. (MSEA)	Seattle-Everett, Wash.
Indianapolis, Ind.	Spokane, Wash.
Jacksonville, Fla.	Syracuse, N.Y.
Lansing, Mich.	Tacoma, Wash.
Las Vegas, Nev.	Tampa-St. Petersburg, Fla.
Little Rock-North Little Rock, Ark.	Toledo, Ohio-Mich.
Nonself-representing non-SMSA's	
Auglaize Co.-Shelby Co., Ohio	Franklin Co.-Jackson Co.-Williamson Co., Ill.
Benton Co.-Carroll Co., Ind.	Gallatin Co.-Saline Co., Ill.
Calhoun Co.-Clay Co.-Roane Co., W. Va.	Kitsap Co., Wash.
Calvert Co.-Charles Co.-St. Mary's Co., Md.	Lee Co.-Van Buren Co., Iowa
Caroline Co.-Fredericksburg City-King George Co.-Spotsylvania Co.-Stafford Co., Va.	Lincoln Co., Mont.
Chaffee Co., Calif.	Middlesex Co., N.J.
Claiborne Co.-Hamblen Co.-Hancock Co.-Hawkins Co., Tenn.	Montgomery Co.-Otsego Co., N.Y. (represented by 2 PSU's)
Clarendon Co.-Sumter Co., S.C.	Orangeburg Co., S.C.
Danville City-Henry Co.-Martinsville City-Pittsylvania Co., Va.	Reeves Co., Tex.
Darlington Co.-Dillon Co.-Marlboro Co., S.C.	Reno Co., Kans.
Davidson Co.-Rowan Co., N.C.	Sheboygan Co., Wis.
De Soto Co.-Sarasota Co., Fla.	

fraction for women of other races was about 1 in 4,610 (see table C). The first step in producing these disparate rates occurred in the second stage of the sampling process: The selection of 1970 census enumeration districts (ED's) within sample primary sampling units. This was accomplished by stratifying ED's by the percent of the population that was black, and by using a higher rate of selection in strata with a 10-percent or greater black population.

Stratification of enumeration districts.— Before the ED's were stratified, a certain amount of recombination was necessary. Although the block groups (small ED's used by the U.S. Bureau of the Census in densely populated areas) and ED's averaged about 400 housing units in size, there was considerable variation about this mean. Thus the smallest ED's and block groups were combined to form ED's that contained 10 or more housing units each. For

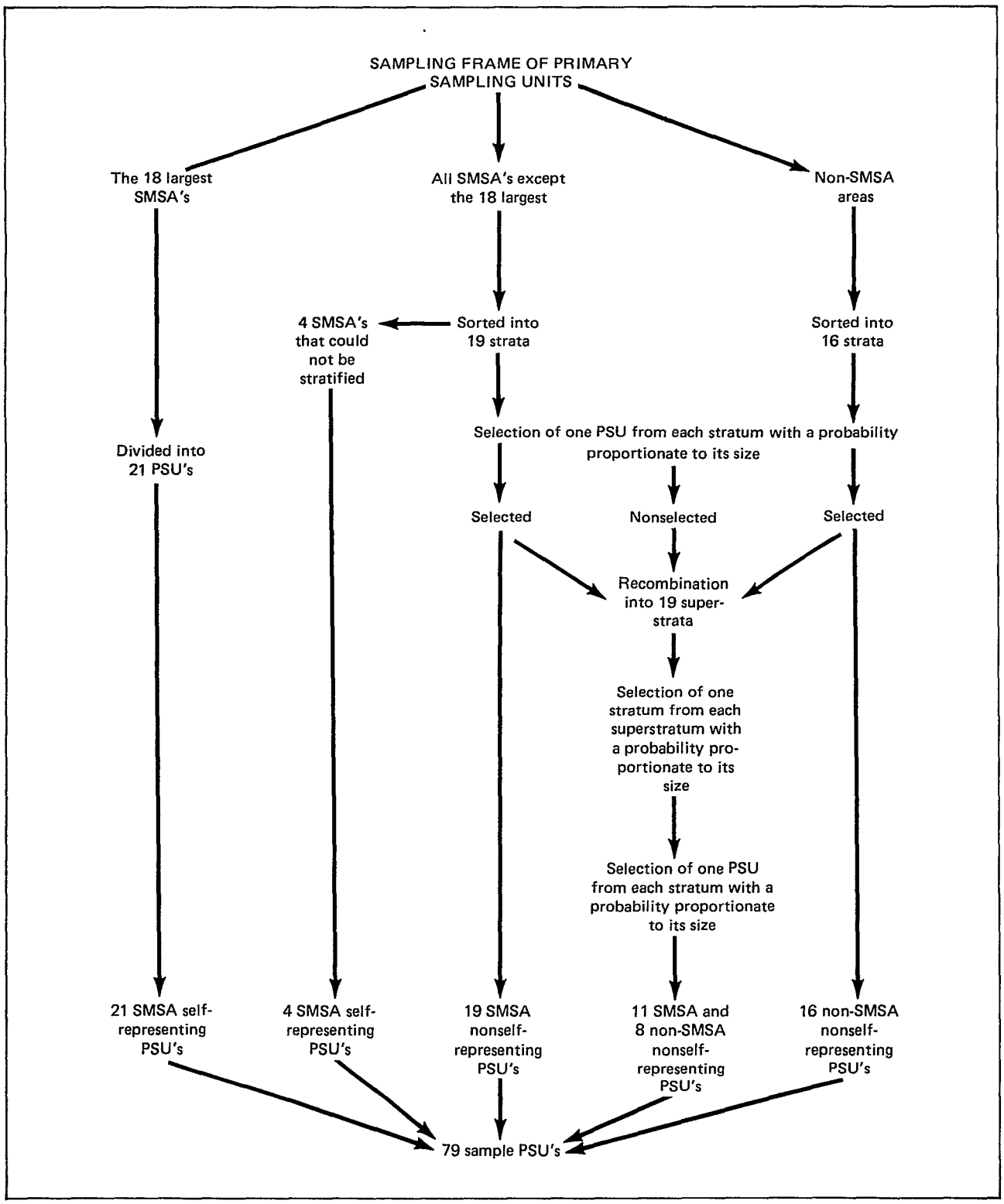


Figure 1. First-stage selection of primary sampling units

Table C. Estimated number of women eligible for inclusion in the sample, expected sample size, and approximate sampling fraction required to produce the expected sample size, by race

Race	Number of eligible women[1]	Expected sample size	Approximate sampling fraction[2]
All races.....	31,444,000	10,000	1 in 3,140
Black	3,795,000	4,000	1 in 950
Other races	27,649,000	6,000	1 in 4,610

[1]Numbers are estimates, produced by the National Center for Health Statistics, of the number of women 15-44 years of age who were either ever-married or never-married and who have children of their own living with them. Estimates are for January 1976.

[2]Sampling fractions represent the overall proportion of eligible women in each race group that would have to be selected in order to produce a sample of the appropriate size and racial composition. The various sampling rates used in the sample design were intended to produce these overall rates after such factors as subsampling in certain strata and nonresponse were taken into account.

ease of discussion, block groups, ED's, and recombined block groups and ED's will be referred to as ED's.

The stratification of ED's was accomplished on the basis of information included in the 1970 Census of Population and Housing.[2] These data were used to determine the proportion of the population of each ED that was black, and thus to which stratum the ED should be assigned. The strata were defined as follows:

Stratum 1: ED's with a black population of 59.5 percent or more.

Stratum 2: ED's with a black population of 29.5 to 59.499 percent.

Stratum 3: ED's with a black population of 9.5 to 29.499 percent.

Stratum 4: ED's with a black population of 0.0 to 9.499 percent.

Determination of size.—After this initial stratification by race, each ED was assigned a measure of size, which was the number of area segments (groups of year-round housing units) that it contained. The size of each area segment, in turn, was a function of the ED stratum. The average number of housing units per segment was 15 in strata 1 and 4, 30 in stratum 2, and 60 in stratum 3. The measure of size for ED's in strata 1 and 4, for example, was thus the number of year-round housing units they contained divided by 15. This ratio was then rounded to the nearest integer for sampling purposes.

If an ED assigned to stratum 2 or 3 did not have a number of housing units equal to at least two-thirds of the average segment size for that stratum, a second combining operation was performed. The newly created ED's were then placed in the proper stratum, and a new measure of size was computed. In both the first and second combining operations two restrictions applied: No more than four ED's were to be grouped together, and combined ED's must all be from the same county.

Sequencing of enumeration districts.—Following the assignment of a measure of size, all enumeration districts were sorted by: (1) primary sampling unit number, (2) stratum code, and (3) the Westat Economic Index.[d] The sorting by primary sampling unit number had the effect of grouping enumeration districts by region and by type of primary sampling unit (self-representing SMSA's, nonself-representing SMSA's, and non-SMSA's) within a region. The Westat Economic Index was sequenced within strata in ascending order in the first and third strata, and in descending order in the second and fourth strata. The sequencing operation, combined with the systematic sampling of ED's that was carried over from one primary sampling unit to the next and from one stratum to the next, produced the effective stratification of ED's by region, residence, and income. This stratification was intended to achieve a small reduction in sample variance, and it ensured a representative sample of women by these variables.

[d]The census ED file contains only information collected on a 100-percent basis and thus does not include income data. Westat, Inc. therefore imputed a median income for each ED by creating a regression of income on such items as value of home and rent paid, and applying the regression model to each ED. These imputed values form the Westat Economic Index.

Selection of enumeration districts.—Prior to the sampling operation, the measure of size for each ED (number of segments) was inflated by the inverse of the probability of selection for the primary sampling unit that it was in. This weighting has two effects: Households from all PSU's have exactly the same overall probability of being included in the final sample, and a representative sample can be achieved while using the same sampling fraction across all PSU's.

A systematic sample of ED's was drawn using a random start and a sampling interval of 760 segments in strata 1 through 3, and 4,256 segments in stratum 4. Thus the segment corresponding to the random start and every 760th or 4,256th segment thereafter was identified, and the ED's in which they fell were selected. A few ED's with an inflated measure of size that was greater than the sampling interval were selected more than once.

Only one random start was needed. The systematic sampling was carried over from one stratum to the next within primary sampling units and from the last ED selected in one primary sampling unit to the first ED in the next.

Third-Stage Selection of Segments

Initial subdivision of segments.—Each enumeration district was divided conceptually into the number of segments that had been determined during the second stage of the sampling process. When the ED was from an area for which census block statistics were available, block maps and census counts were used in the first step of this subdivision. Each block (or occasionally, groups of contiguous blocks) was then assigned one or more segments based on the number of dwelling units it contained and on segment size (determined by the ED stratum).

When block statistics were not available, census county and place maps were used. If the physical features shown contained enough information, the maps were used to accomplish the segmentation. Where these maps were inadequate, one of two alternative procedures was used. Wherever U.S. Department of Agriculture aerial photographs (taken about 1970) were available, a cartographer was employed to do the segmentation. In other areas field interviewers were sent to the sites to prepare maps for this purpose. However, the physical features of these rural areas did not always allow the ED's to be divided into segments of the desired size. Thus, as with block areas, these initial subdivisions sometimes included more than one segment.

Selection, field listing, and subsampling of area segments.—Regardless of the method used to divide each enumeration district into land areas, one area was randomly selected from each district with a probability proportionate to the number of segments it contained. If the area contained only one segment (had a measure of size of 1) it became a sample segment. However some areas had a measure of size that was greater than 1; that is, they included more dwelling units (DU's) than was appropriate for a single segment. In these cases the subsampling of dwelling units was employed to produce sample segments of the proper size.

The first step in this subsampling process was the field listing of area segments. This address listing operation was not limited to segments that were too large, however, since all segments in strata 1-3 were later to be subsampled by race, and also because the listing provided a more accurate count of dwelling units than did census data.

As interviewers were recruited for the survey, they received training as listers and were assigned three segments each. For each segment the lister received a "segment folder" that included a census map, two sketch maps, and address listing sheets. She then proceeded to the segment, verified the segment boundaries, corrected the maps when necessary, and listed the address of each dwelling unit in the segment (or recorded a detailed description of the unit where the house or apartment number was not evident). The listing was also verified through the independent relisting of one out of three segments (and another relisting if large discrepancies were found), and through checks for missed dwelling units during the interview stage of the survey.

For exceptionally large segments, however, an additional operation was introduced prior to the listing of addresses. The maps for these segments were used to prepare sketch maps containing further subdivisions of the areas. The

resultant subsegments were then assigned a measure of size based on approximate DU counts, and the maps were returned to Westat, Inc. for the selection of one subsegment for listing and inclusion in the final sample.

After the listing of addresses, segments that contained more than the required number of DU's were subjected to a systematic sampling procedure. Thus a segment that included twice the number of DU's dictated by the stratum from which it was drawn, for example, had every second DU on the address listing sheets designated for interview.

New construction segments.—Approximately 90 percent of the respondents included in the final sample lived in area segments that were selected this way; however, about 10 percent lived in new construction segments. Although area segments would have provided an unbiased sample of new construction units (dwelling units for which building permits were issued during the period September 1969-July 1975), it was thought that such a procedure would significantly increase sample variances. This would occur since the measure of size used to define area segments was based on 1970 census data and thus would not reflect the true number of DU's in segments containing units built after 1969. Since new construction tends to occur in large clusters, and thus in a small proportion of area segments, a small, highly clustered subset of new DU's would be obtained through area segment selection. Consequently, new construction segments were selected in a separate sampling step whenever possible. However, 15-20 percent of new construction (according to census figures) occurred in areas where no easily accessible file of building permits (the bases of new construction segments) existed, and for these localities new construction was picked up in the area segments along with older units.

The new construction segments were defined to be 15 DU's in size (the same size as area segments from enumeration districts in strata 1 and 4) and were systematically sampled at the rate of 1 in 4,256 (the same as stratum 4 area segments). The sample selection procedures were slightly different for DU's with permits issued during the part-years September-December 1969, and January-July 1975, than for those with permits issued during the period 1970-74.

The U.S. Bureau of the Census produced magnetic data tapes that contain annual totals of the number of housing units authorized for 1970-74 for each permit-issuing place. These tapes were used to draw the sample of new construction DU's with permits issued during that period. The file was first sorted by primary sampling unit (the same 79 PSU's selected in the first stage), by permit-issuing place within each PSU, and by year of permit within each place. The total authorized number of units for each place was then calculated for each year, and this number was divided by 15 (the average segment size) and rounded to the nearest integer. The result was used as the measure of size for each "place-year" on the tapes and is analogous to the measure of size assigned to enumeration districts during their second-stage selection. This measure was next weighted by the inverse of the probability of selection for the PSU in which the place was located, and a systematic sample of place-years was drawn at the rate of 1 in 4,256 segments. Just as during the selection of enumeration districts, this sampling was accomplished using a single random start and continuous sampling across all PSU's.

After the sample of place-years was drawn, a measure of size was assigned to every month within each place-year. This measure of size (number of segments of 15 building permits each) was determined by published census data for larger permit-issuing places,[3] and through contact with building permit personnel for the remainder. One month, or in some cases a group of adjacent months, was then randomly selected from each place-year with a probability proportionate to size, an operation that is analogous to the selection of area segments from enumeration districts.

For 1969 and 1975, census data tapes could not be used for the selection of place-years since only part of each year was included. Monthly census reports[3] and reports from permit-issuing offices, used in the subsampling of place-years for 1970-74, were thus used to directly sample place-months. To this end, a measure of size (number of segments) was assigned to each month for each place, and the measure was weighted by the inverse of the probability of selection for the PSU in which the place was located. The months were then sequenced by

chronological occurrence, PSU, and place, in that order, and a systematic sample of place-months was drawn at the rate of 1 in 4,256 segments.

The new construction segments formed by selected place-months (whether selected directly or from place-years) were next subjected to listing and subsampling operations in order to produce sample segments of the proper size.

Listing and subsampling of new construction segments.—The listing of new construction segments (hereafter referred to as stratum 5 segments) was done in a manner substantially different from that for area segments from strata 1-4. For stratum 5 segments, the listing was accomplished using the file of building permits in each selected building permit office. The lister was required to obtain access to the file of permits for the selected month or months, determine the number of dwelling units represented by each permit, and list all dwelling units on an address listing form similar to that used for area segments. The DU's were then conceptually grouped into segments so that the first 15 encountered formed one segment, the next 15 another, and so on until all DU's had been exhausted. This procedure was used so that the segments thus created formed geographically compact clusters and resulted in reduced interview costs. Following segmentation a random sample of segments was drawn at the rate necessary to produce the final self-weighting sample of 1 in 4,256.

The sample design for new construction segments had a number of desirable characteristics. The most important of these is that all new construction units in the United States had the same overall probability of being included in the final sample. Further, that probability was 1 in 4,256, the same probability established for DU's from stratum 4. The method also yielded segments of 15 dwelling units in size (identical to stratum 4) and produced a maximum spread among places within the 79 sample PSU's. Figure 2 illustrates the selection process for both new construction and area segments.

Integration of area and new construction segments.—The listing and subsampling of area and new construction segments produced approximately 1,700 segments, consisting of about 32,000 households. In order to provide an unbiased sample of all households, it was then necessary to integrate households from the new construction segments with the rest of the sample in such a way that all households had a known, non-zero probability of selection. To this end, reports published by the Bureau of the Census[3] were used to determine which area segments were in localities that did not require building permits. In these segments interviewers were instructed to include all households, both new and old construction, in the sample. Thus the probability of selection for households in any segment in a non-permit locality was not conditional on the number of new dwelling units the segment contained.

In the remaining areas, where new construction segments were sampled separately, all new units were excluded from the area segment sample. This operation precluded the possibility of new units being sampled twice and ensured that the probabilities of selection for old and new units were independent. The exclusion of new units was accomplished by including a question on when the housing unit was built in the Household Screener questionnaire (see appendix II).

Missed dwelling units.—During the interview phase of the survey, interviewers were required to check for dwelling units not discovered during the listing operation. These dwelling units may have been missed as the result of the lister having overlooked either entire structures or individual DU's within listed structures. The procedures for locating and including missed DU's in the sample were designed to give them the same probability of selection as other DU's from segments in the same stratum. The procedure was as follows:

1. At each address that had been listed as a single-family home, the interviewer checked to ensure that all persons living at that address were included in the screening. Any additional dwelling units located at that address were also included in the sample.
2. In multiunit structures, the interviewer looked for missed dwelling units within the structure *only* if the first-listed unit in the building was included in the sample. The check included questioning re-

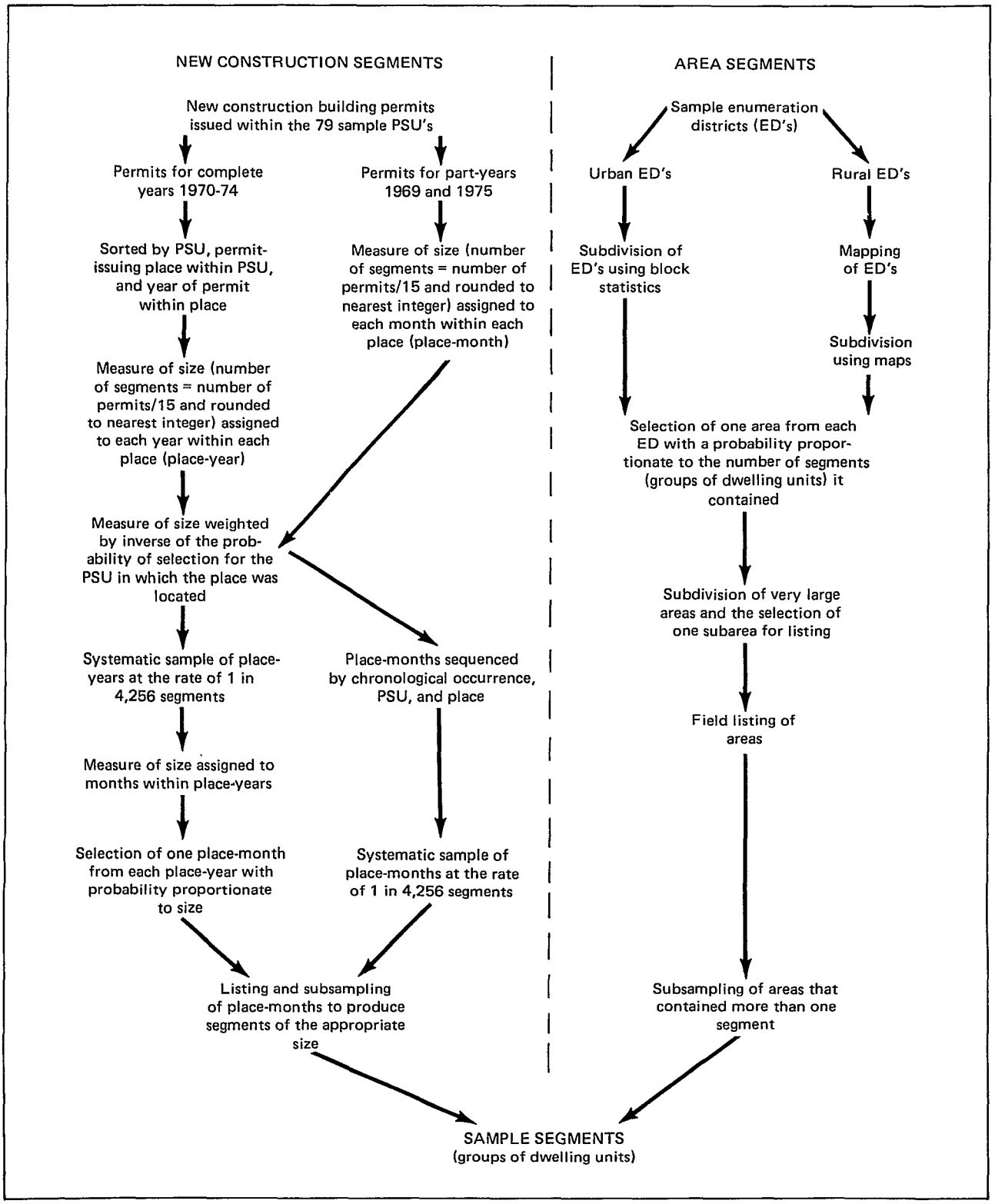


Figure 2. Third-stage selection of sample segments

spondents about possible missed dwelling units and looking at mailboxes or doors that might give indications of a unit not listed. Any additional dwelling units located in this manner were included in the sample.

3. In segments, missed structures were searched for *only* if the first-listed structure in that segment was included in the sample. When missed structure checks were initiated, the procedure included canvassing the segment by foot or by car, using the segment map and listing sheets to look for additional structures, and querying respondents about possible missed structures in the area. Any dwelling units located in missed structures were included in the sample.

The procedure of having the first-listed unit determine whether a missed DU or missed structure check was made ensured that the self-weighting feature of the sample was retained. This was the case since the check was made for all segments and structures in which all the DU's were included in the sample, for half of the segments and structures in which the subsampling rate had been 1 in 2, for one-third of the segments and structures in which the subsampling rate had been 1 in 3, and so on. The only deviation from this self-weighting principle occurred when five or more missed DU's were located in a single missed structure or at a DU listed as a single unit. In these cases the interviewer contacted the sampling department at Westat, Inc., which then selected four of the DU's for interview and inclusion in the sample. About 700 DU's were included in the sample as a result of the missed DU checks.

Fourth-Stage Selection of Households

Subsampling for race.—During the second stage of the sampling process, a higher sampling rate was used for the selection of enumeration districts (ED's) in strata 1-3 than for those in strata 4 and 5. The higher rates in strata 1-3 (ED's where 9.5 percent or more of the population was black) were intended to produce a higher probability of selection for black women than for women of other races and thus yield the approximately 4,000 sample black women (out of a total sample of about 10,000) required by the National Survey of Family Growth design specifications. However, since about 45 percent of eligible women in strata 1-3 were expected to be of races other than black, this procedure alone would also have produced an unnecessarily large sample of these women. Consequently, the fourth-stage selection of sample households involved a subsampling operation designed to reduce the overall sampling rate for these women and produce a sample of the desired size.

This subsampling operation included the use of a computer to designate a portion of sample dwelling units in those strata as subsample units. All households containing eligible black women became sample households, but households containing eligible women of other races were included in the sample only if they were found in designated subsample units. The interviewers received instructions on the Household Screeners as to which households were to include any eligible woman and in which ones only black women were to be included.

The initial sample design called for the subsampling of households in strata 1-3 at the rate of 1 in 5. This rate would have produced approximately equal sampling rates for women of races other than black across all 5 strata: 1 in 3,800 in strata 1-3, and 1 in 4,256 in strata 4 and 5. However, after the fieldwork was well underway and the 1 in 5 subsample already selected, a decision was made to increase the subsampling rate to 13 in 20. This decision was necessary because significantly fewer of these women than expected were being included in the final sample. This undercoverage resulted primarily from two factors: fewer-than-expected dwelling units in strata 1-4 (see table D), and a population shift that produced fewer-than-expected eligible women of races other than black in strata 1-3.

Since the fieldwork was more than half completed under the first regime (a 1 in 5 subsample), the increased subsampling rate was achieved through a supplementary sample selected from households with eligible women that were rejected in the initial subsample. The way in which this supplementary sample was drawn was dictated by the fact that the entire National Survey of Family Growth sample had previously

Table D. Total and expected number of dwelling units (DU's) in the United States, and percent difference, by stratum

Stratum	Total DU's in thousands[1]	Expected DU's in thousands	Percent difference
All strata	72,509	74,725	-3.0
Stratum 1	4,612	5,370	-14.1
Stratum 2	3,096	3,420	-9.5
Stratum 3	4,030	4,940	-18.6
Stratum 4	52,306	52,995	-1.3
Stratum 5	8,465	8,000	+5.5

[1]Estimates are based on the total number of DU's assigned for screening (including those from missed DU checks) and are the product of the number assigned and the rate at which the DU's were sampled.

been divided into four random subsets, which were assigned for interviewing in sequence. At the time the decision was made to increase the subsampling rate the third-period interviewing was almost completed, and the fourth-period interviewing was well underway. Thus a complete list of households with eligible women of races other than black that had been rejected in the initial subsampling during the first three periods was soon to be available. The supplementary sample of households was easily obtained and involved only the removal of the message: "Conduct extended interview if black only" from the label on the screeners for 4 out of 5 of strata 1-3 households. If the household had already been screened and rejected because the respondent was not a black woman, the case was returned to the field. If the case had been returned as a nonresponse, it was included in the followup effort.

To supplement the fourth period would have required either changing methods during ongoing fieldwork or waiting until the fourth-period interviewing was completed under the old regime and then instituting the same procedures used for periods 1-3. Neither alternative was desirable because each involved too much potential for error or delay. Thus the entire supplemental sample effort was restricted to the first three periods of interviewing. Since households had been randomly assigned to each of the four periods, the probability of selection for women of races other than black was constant and not dependent on interview period. Because about three-quarters of the strata 1-3 households was subjected to a 4 in 5 subsample, and about one-quarter of these households was subjected to a 1 in 5 subsample, the overall subsampling rate achieved was the following:

$$\frac{3}{4}\cdot\frac{4}{5}+\frac{1}{4}\cdot\frac{1}{5}=\frac{13}{20}$$

Fifth-Stage Selection of Sample Persons

Subsampling when there was more than one eligible respondent in the household.—To avoid the high correlation of information from eligible women within the same dwelling unit, the sample design stipulated that no more than one eligible woman from any sample DU would be interviewed. During completion of the Household Screener, the interviewer listed all residents of the DU on one page of the form and relisted the eligible females in order of age in item S-14 on another page (see appendix II). Item S-14 provides space for listing up to six persons because six was considered to be a reasonable limit for the number of eligible females to be expected from a single DU. When the interviewer listed more than one eligible woman, she referred to the sampling table at the beginning of the questionnaire to determine which person she was to interview. The sampling table consists of five numbers that designate which person to interview when the number of eligible women is two, three, four, five, or six or more (see table E). Westat, Inc. home office personnel filled in the table on every Household Screener by systematically assigning 12 versions of the table to the screeners. This method of assigning interviews gave each eligible woman in a given household the same probability of being selected.

Nonresponse Followup

Subsampling of nonresponse cases.—An elite corps of traveling interviewers and assistant supervisors was assigned the task of obtaining interviews at sample households included in the followup effort. A household was assigned to this operation after all efforts to complete an

Table E. Example of sampling table on the Household Screener

IF NUMBER OF ELIGIBLE FEMALES LISTED IN SUMMARY BOX IS: ↓	THEN INTERVIEW PERSON LISTED ON SUMMARY BOX LINE ↓
Two	1
Three	3
Four	1
Five	4
Six or More	3

interview (either a screener or an extended interview) were exhausted by local interviewers. Nonresponse cases that appeared to be nonconvertible refusals were excluded, however.

Prior to the followup interview effort, a 50-percent subsample of these cases was selected using systematic cluster sampling procedures. This subsampling, designed to reduce interview costs, was accomplished in two ways. In large-city PSU's, where large numbers of nonresponse cases were expected, the households were grouped by segment, the segments were sequenced in descending order by the number of followup cases they contained, and a systematic sample of half of the segments was drawn. For the remaining PSU's, nonresponse households were grouped by PSU, the PSU's were sequenced in descending order by the number of followup cases they contained, and a 50-percent sample of PSU's was selected systematically. Approximately 4.5 percent of the respondents who were included in the final sample came in as a result of this followup effort. They represent about 9 percent of the population of respondents (since they were subsampled at the rate of 1 in 2) and thus are weighted double relative to other women in the sample (see table F for the percents of women selected at various sampling rates).

Table F. Percent of sample women, by race and sampling rate

Sampling rate[1]	Race	
	Black	Other races
	Percent	
1 in 760	89.5	...
1 in 1,520	4.3	...
1 in 1,169	...	25.1
1 in 2,338	...	0.8
1 in 4,256	5.9	70.5
1 in 8,512	0.3	3.6

[1]The sampling rates exclude the effects of subsampling multieligible households.

CHARACTERISTICS OF THE SAMPLE

Response Rates

The first four stages of the sample design (plus the missed dwelling unit procedure) resulted in the identification of 32,683 sample DU's. However, during the fifth-stage screening operation and prior to the subsampling of nonresponse cases (see preceding discussion), interviewers discovered that 5,306 of these were either vacant or not a DU as defined by the National Survey of Family Growth. Screeners were completed for 23,734 of the remaining 27,377 occupied DU's, and 8,226 extended interviews were completed among the 9,064 DU's found during screening to contain an eligible woman (see table G).

Prior to the intensive followup effort to convert the nonresponse DU's, a 50-percent subsample of these DU's (excluding nonconvertible refusals) was drawn. A total of 3,946 DU's was held for subsampling and 1,958 were selected. The final disposition of these cases is shown in table G.

After the followup effort was completed, 30,695 DU's remained in the sample (the original 32,683 minus the 1,988 held for subsampling but not selected). Screener interviews were completed at 24,188 of the 25,297 DU's found to be legitimately occupied DU's, producing a fifth-stage sample of 9,470 women eligible for an extended interview. Extended interviews were completed for the 8,611 women compris-

Table G. Number of dwelling units (DU's), by nonresponse subsample status and final disposition

Final disposition	Subsample status		
	Before nonresponse subsample drawn	In nonresponse subsample	After nonresponse subsample drawn
All DU's			
Total	32,683	1,958	30,695
Vacant or not a DU	5,306	92	5,398
Legitimately occupied DU	27,377	1,866	25,297
Disposition of legitimately occupied DU's			
Completed screener interview	23,734	1,292	24,188
Eligible for extended interview	9,064	732	9,470
Completed extended interview	8,226	385	8,611

ing the final National Survey of Family Growth sample (see table G).

When calculating screener, interview, and combined response rates, the nonresponse subsampling procedure must be taken into account. For example, to ignore the subsampling procedure entirely would mean that the 1,988 DU's held for subsampling but not selected would enter the denominators of the rates even though they did not receive the complete effort to convert them to response DU's (which enter the numerators of the rates). Thus the procedure would negatively bias the rates. On the other hand, to simply delete the 1,988 nonselected cases from the calculations would *positively* bias the rates. That is, if the unselected DU's had been subjected to the same followup efforts as the selected DU's, and if both groups were included in the calculation of response rates, the rates would probably be lower than if the unselected DU's were simply excluded from the calculations. Response rates calculated from table G provide supporting evidence for this position. For example, the screener response rate (the ratio of completed screeners to occupied DU's) for the nonresponse subsample is 69.2 percent while the comparable rate for the full sample *before* the subsample was drawn (and which thus includes *all* the nonresponse cases) was 86.7 percent.

In order to overcome these problems the response rates presented in this report are weighted response rates. That is, they are based on numbers of DU's in which subsampled DU's are weighted double relative to others. This weighting allows the subsampled DU's to represent not only themselves but also the DU's held for subsampling but not selected. Thus the weighted rates are those that would have occurred if there had been no nonresponse subsampling and the DU's held for subsampling but not selected had been converted to response DU's at the same rate as subsampled DU's.

The weighted numbers of DU's from which the response rates are calculated are found in table H. The weighted response rates are found in table J, which also contains weighted response rates for each of the five first-stage selection strata. The screener response rate is the percent of occupied legitimate DU's for which a screener interview was completed. The interview response rate is the percent of eligible women for whom completed interviews were obtained. The combined response rate is the product of the screener and interview completion rates divided by 100.

Sample Size

The 8,611 women included in the final sample are nearly 14 percent less than the desired sample size of 10,000. Further, this undersampling is greater among black women than among women of other races; the number

Table H. Weighted number of dwelling units (DU's), by final disposition

Final disposition	Weighted number of DU's[1]
All DU's	
Total	32,653
Vacant or not a DU	5,490
Legitimately occupied DU	27,163
Disposition of legitimately occupied DU's	
Completed screener interview	25,480
Eligible for extended interview	10,202
Completed extended interview	8,996

[1]DU's included in the nonresponse subsample are weighted double.

Table J. Weighted screener, interview, and combined response rates,[1] by stratum

Stratum	Screener response rate	Interview response rate	Combined response rate
	Percent		
All strata	93.8	88.2	82.7
Stratum 1	92.3	87.8	81.0
Stratum 2	95.0	87.5	83.1
Stratum 3	94.2	85.2	80.3
Stratum 4	94.4	89.4	84.4
Stratum 5	90.8	89.0	80.8

[1]The results in this table are weighted for the subsampling of nonresponse cases.

of sample black women was about 25 percent below the expected total compared to an undersample of about 7 percent for women of other races. However, much of this racial differential is due to the decision, made after data collection was well underway, to increase the subsampling rate for women of other races in strata 1-3, thus increasing the overall sampling rate for these women and reducing the extent to which they were underrepresented.

The unexpectedly small sample is due to several factors, the most important of which is the subsampling of nonresponse followup cases. Since 50-percent subsampling was used for these cases, each woman brought into the sample as a result of the followup effort actually represents two nonresponse cases. Allowing these women to assume a double weight, the number of women in the sample becomes 8,996 (nearly 90 percent of the expected total), of which 3,161 are black women (about 79 percent of the expected total) and 5,835 are women of other races (about 97 percent of the expected total).

Other important factors are the following:

1. *Higher-than-expected nonresponse rates.* The expected combined response rate was 90 percent compared to an actual rate of 82.7 percent.

2. *Fewer-than-expected dwelling units in some strata.* This factor may have been the result of higher-than-expected rates of abandonment and demolition of structures in some areas (see table D).

3. *Missed eligible women within some dwelling units.*

4. *Population movement across strata.* This problem is especially critical for black women since they were sampled at a much different rate in strata 1-3 than in strata 4 and 5. It is estimated that about 670,000 black women had moved from areas where they would have been sampled at the rate of 1 in 760 to areas where they were sampled at the rate of 1 in 4,256. This factor alone could have reduced the sample of black women by about 725.

ESTIMATION

Weighting Procedures

Overview.—Since the National Survey of Family Growth (NSFG) is designed to produce unbiased estimates for the entire population of eligible women in the conterminous United States, the sample data must be inflated to the level of the population from which the sample was drawn. The inflation factor, or weight, for each woman is the product of several adjustments, including one or more for each stage of sampling. Three types of adjustments are in-

volved: inflation by the reciprocal of the probabilities of selection, nonresponse adjustment, and poststratification adjustment.

Inflation by the reciprocal of the probabilities of selection.—The weight for each woman is the product of the reciprocals of the probabilities of selecting (1) the primary sampling unit, (2) the enumeration district, (3) the segment, (4) the household, and (5) the eligible sample person.

Nonresponse adjustment.—Each sample weight is adjusted for nonresponse to the Household Screener (screener nonresponse) and nonresponse to the detailed NSFG questionnaire (interview nonresponse). These adjustments are necessary because the phenomenon of nonresponse introduces bias into any probability sample; women with certain characteristics are more likely to be nonrespondents than other women; thus the former are likely to be underrepresented in the sample. Nonresponse adjustments minimize the impact of this bias. Adjustment for screener nonresponse was achieved by imputing to women in nonresponding households the characteristics of women in responding households in the same primary sampling unit and stratum. Adjustment for interview nonresponse was achieved by imputing to nonresponding women the characteristics of responding women in the same age-race class and primary sampling unit.

Poststratification by age and race.—The weight for each ever-married respondent is multiplied by a poststratification adjustment factor that is determined by the woman's age and race. The 12 adjustment factors shown in table K make NSFG estimates of ever-married women in each age-race class equal to independent control totals for April 1976, the approximate midpoint of data collection. The control totals are based on data from the Current Population Survey (CPS) conducted by the U.S. Bureau of the Census. No poststratification adjustment is applied to the weights for never-married mothers because reliable control totals are not available. Poststratification achieves much of the improvement in precision that would have been attained if the sample had been drawn from a population stratified by age and race. A discussion of the method used to compute the CPS control totals and a comparison of simple inflation and poststratification estimators can be found in another report.[1]

Table K. Poststratification adjustment factors,[1] by race and age

Age	Race	
	Black	Other races
15-19 years	1.457	0.937
20-24 years	1.078	1.134
25-29 years	0.882	1.079
30-34 years	1.078	1.212
35-39 years	1.000	1.119
40-44 years	1.156	1.106

[1]Ratio of May 1976 population control totals, based on Current Population Survey data, to National Survey of Family Growth weighted estimates.

Estimating Equation

The Cycle II estimator of an aggregate parameter Y for all women in the NSFG target population is

$$Y' = Y'_1 + Y'_2$$

where Y'_1 is the estimated total for ever-married women and Y'_2 is the estimated total for never-married mothers.

$$Y'_1 = \sum_{\alpha=1}^{12} Y'_{\alpha 1} \cdot \frac{X^*_{\alpha 1}}{X'_{\alpha 1}}$$

where

$Y'_{\alpha 1}$ = the weighted, nonresponse-adjusted estimator of Y for ever-married women in age-race class α. The 12 α classes are defined by 2 race groups—black and other races—crossed with 6 age groups: 15-19, 20-24, 25-29, 30-34, 35-39, and 40-44.

$X'_{\alpha 1}$ = the weighted, nonresponse-adjusted estimator of the number of ever-married women in class α.

$X^*_{\alpha 1}$ = an independent estimate of the number of ever-married women in class α based on data from the Current Population Survey.

$$Y'_2 = \sum_{\alpha=1}^{12} Y'_{\alpha 2}$$

is the sum of the corresponding weighted, nonresponse-adjusted estimates for never-married women.

The weighted, nonresponse-adjusted estimator for ever-married women in age-race class α is given by

$$Y'_{\alpha 1} = A_\alpha \cdot \sum_{g=1}^{5} \left\{ B_g \cdot \sum_{h=1}^{79} \left(W_{1g} \cdot I_{\alpha h} \cdot S_{gh} \cdot \sum_{i=1}^{L_{gh}} \left[W_{2ghi} \cdot W_{3ghi} \cdot R_{3ghi} \cdot \sum_{j=1}^{H_{ghi}} (D_g \cdot F_{4ghij} \cdot T_{4ghij} \cdot Y_{\alpha ghij} \cdot \delta_{\alpha ghij}) \right] \right) \right\}$$

where

B_g, S_{gh} = adjustments for nonresponse to the Household Screener.

A_α, $I_{\alpha h}$ = adjustments for nonresponse to the personal interview.

W_{1h} = first-stage weight = reciprocal of the probability of selecting primary sampling unit h.

W_{2ghi} = second-stage weight = reciprocal of the probability of selecting second-stage unit i within stratum g and primary sampling unit h. Strata 1-4 are the race strata, stratum 5 is the new (post-1969) housing stratum.

W_{3ghi} = third-stage weight = the reciprocal of the probability of selecting the sample segment from second-stage unit i.

R_{3ghi} = the reciprocal of the large-segment subsampling rate:
2 for certain segments with large numbers of households;
1 for all other segments.

D_g = the reciprocal of the household-race subsampling rate:
20/13 if g = 1, 2, or 3 and the respondent in the household is not black;
1 otherwise.

F_{4ghij} = the reciprocal of the nonresponse subsampling rate:
2 if household j, or the sample women therein, was in the nonresponse sample;
1 otherwise.

T_{4ghij} = the number of eligible women in household j of the segment in second-stage unit i.

Y_{ghij} = the value of characteristic Y for the sample women from household j.

$\delta_{\alpha ghij}$ = 1 if the respondent has ever been married and belongs to age-race class α;
0 otherwise.

L_{gh} = the number of second-stage units in stratum g and primary sampling unit h.

H_{ghi} = the number of completed interviews from households in the sample segment from second-stage unit i.

$Y'_{\alpha 2}$ is exactly the same as $Y'_{\alpha 1}$, except that $\delta_{\alpha ghij} = 1$ for never-married women in class α.

VARIANCE ESTIMATION

Background

The balanced half-sample replication technique described in detail in other National Center for Health Statistics reports[4,5] is used to estimate National Survey of Family Growth (NSFG) variances. An empirical study by Bean[6] gives evidence that the half-sample technique produces highly reliable, essentially unbiased variance estimates.

Three important practical reasons why half-sample replication is being used are as follows:

1. Programming difficulties are reduced because half-sample variances are computed by taking a simple average of squared deviations of half-sample estimates from the estimate based on the full sample. Instead of having to program an exceedingly difficult variance formula, the programmer must simply adjust the estimation formula to compute estimates from appropriately chosen half samples.

2. The complete algebraic formula for NSFG variances is unknown because of the complexity of the design. Although algebraic expressions can be derived for particular subprocedures—such as the individual stages of sampling and the poststratification and nonresponse adjustments—a single, exact variance equation has not been developed.

3. As stated by McCarthy[4]: "Variance estimates based upon the replicated estimates will mirror the effects of all aspects of sampling and estimation that are permitted to vary randomly from replicate to replicate." Also, replicated half-sample variances include some of the variability due to nonsampling (measurement) error, as well as sampling variability.

Summary of Applicable Theory

The population of interest is classified into L strata, and two sample primary sampling units (PSU's) are drawn from each stratum. Selection of exactly two sample PSU's reflects an essential element of the theory. This requirement may be met for noncertainty PSU's by collapsing two strata having one PSU each, or for certainty PSU's by creating two artificial, or pseudo, PSU's by random methods from a single PSU. The collapsing method produces somewhat positively biased (overstated) variance estimates by introducing a between-stratum component of variance that does not exist.[7]

Let the parameter of interest be denoted by Y, for which an estimate Y' has been obtained from the complete sample. If Y' is a linear combination of the sample observations, it can be shown that Y' is an unbiased estimate of Y. However, several empirical investigations indicate that the bias of half-sample variance estimates for certain ratio estimators and correlation statistics is negligible, if detectable at all.[4,5,8,9]

A half-sample replicate is defined as a collection of L primary sampling units obtained by selecting one of the paired sample PSU's from each stratum. If the PSU's within each stratum are designated by the subscript $i = 1$ or 2 and there are K half samples, where $K \geqslant L$, the pattern may be summarized as in table L. The "+" indicates that a PSU falls into a particular half sample, and the "–" indicates that it does not.

Analogs of Y' corresponding to each half sample are then computed. That is, for the kth half sample, Y' is given by

$$Y'_k = 2 \sum_{h=1}^{L} Y'_{hi}$$

where i = either 1 or 2 depending on which PSU of the stratum is the half-sample k, and Y'_{hi} is, in this example, a total. The estimator Y' is

$$Y' = \sum_{h=1}^{L} (Y_{h1} + Y_{h2})$$

and its variance is estimated by

$$s^2_{Y'} = \frac{1}{K} \sum_{k=1}^{K} (Y'_k - Y')^2$$

Because it is impractical to compute the Y'_k for the entire set of 2^L possible half samples when L is large, a subset of half samples is selected to produce the estimates. A set of side

Table L. Example of a half-sample replication pattern

Half-sample replication	Stratum								
	1		2		3		...	L	
	PSU		PSU		PSU		...	PSU	
	1	2	1	2	1	2		1	2
1	+	–	–	+	–	+	...	+	–
2	–	+	–	+	+	–	...	–	+
3	–	+	+	–	–	+	...	–	+
.	.	.	.	.	.	.	...	.	.
.	.	.	.	.	.	.	...	.	.
.	.	.	.	.	.	.	...	.	.
K	+	–	+	–	+	–	...	+	–

conditions relating to the selection of PSU's for the half samples has been developed by McCarthy,[4,5] based on work by Plackett and Burman[10] and Gurney.[11] These side conditions greatly increase the stability of $s^2_{Y'}$, by eliminating a between-strata component of variance that is otherwise present. The value of $s^2_{Y'}$, obtained from a subset of half samples that is chosen according to the McCarthy criteria is equal to the value that would be obtained using all 2^L half samples. A set of half samples that satisfies the McCarthy criteria is called a "balanced set," and the procedure is referred to as "balanced half-sample replication."

Application to the National Survey of Family Growth

As a first step in applying the balanced half-sample replication technique, the National Center for Health Statistics grouped the 79 Cycle II primary sampling units into 37 replicate strata. Eighteen of the strata were self-representing; each consisted of the PSU or PSU's associated with a single self-representing SMSA (see table B). Within each of these strata two pseudo-PSU's were created by: (1) listing the PSU's in numerical order within each stratum (for multi-PSU strata); (2) listing the sample segments in numerical order within each PSU; and (3) systematically dividing the segments into two groups, with the first segment and every second-listed segment thereafter assigned to the first pseudo-PSU, and the remaining segments assigned to the second pseudo-PSU.

The remaining 19 strata included the noncertainty PSU's. Their composition was dictated by the way in which these PSU's were originally selected for inclusion in the sample. As described previously, this selection was accomplished in two stages: First, the PSU's were stratified and 35 PSU's were systematically selected; second, all PSU's (both selected and nonselected) were recombined into 19 superstrata and 1 supplemental PSU was selected from each. This design produced two independently chosen observations for each superstratum (the original selection or selections and the supplemental selection) and made the superstrata ideal replicate strata. By pairing the original selections from each strata with the supplemental selection, two pseudo-PSU's were created for each. There was no need for a "collapsing" of population strata to form replicate strata; the common problem of extraneous between-stratum variance in the half-sample variance estimates was avoided.

Within each of the 37 replicate strata, a value of 1 was assigned to one of the pseudo-PSU's, and a value of 2 was assigned to the other. Forty half samples were then created by selecting one of the two alternative values within each strata for each half sample; the values chosen were determined by the elements of an orthogonal 40×40 matrix of 1's and 2's adapted from Plackett and Burman.[10]

In order to estimate the variance of an aggregate statistic Y', 40 half-sample analogs of Y' were computed. The formula for the kth half-sample estimate is

$$Y'_k = Y'_{1k} + Y'_{2k}$$

where Y'_{1k} is the half-sample estimate for ever-married women and Y'_{2k} is the half-sample estimate for never-married mothers. The half-sample estimates Y'_{1k} and Y'_{2k} correspond to the full-sample estimates Y'_1 and Y'_2, respectively, and are computed in the same manner (see the preceding section on estimation), but with case weights adjusted to compensate for the half-sample procedure.[e] The variance of Y' was then estimated by

$$s^2_{Y'} = \frac{1}{40} \sum_{k=1}^{40} (Y'_k - Y')$$

Types of aggregate statistics produced from the National Survey of Family Growth include the number of currently married women, number of ever-married women, number of ever-married women and never-married women with offspring, and number of children ever born to

[e] For replicate strata where one of the pseudo-PSU's was composed of more than one sample PSU, case weights were adjusted so that each pseudo-PSU represented half of the stratum population. In addition, case weights were inflated approximately double to compensate for the fact that only one-half of the sample was used to obtain the estimate.

ever-married women. Half-sample variances were not computed for all aggregate statistics, because to do so would have required a prohibitive amount of time and money. In addition, data reports would be cumbersome if a variance estimate was published for each statistic. Thus variances for each type of statistic were computed only for selected population subgroups, which were chosen to represent a wide variety of demographic characteristics and a wide variation in the size of the estimates. Curves were then fitted to the relative standard error (RSE) estimates for each type of statistic according to the model

$$\mathrm{RSE}(Y') = \sqrt{\frac{s^2_{Y'}}{(Y')^2}} = \sqrt{A + \frac{B}{Y'}}$$

A and B are parameters whose estimates determine the shape of the curve. The rationale for the model and the iterative method that was used to estimate A and B are explained elsewhere.[12]

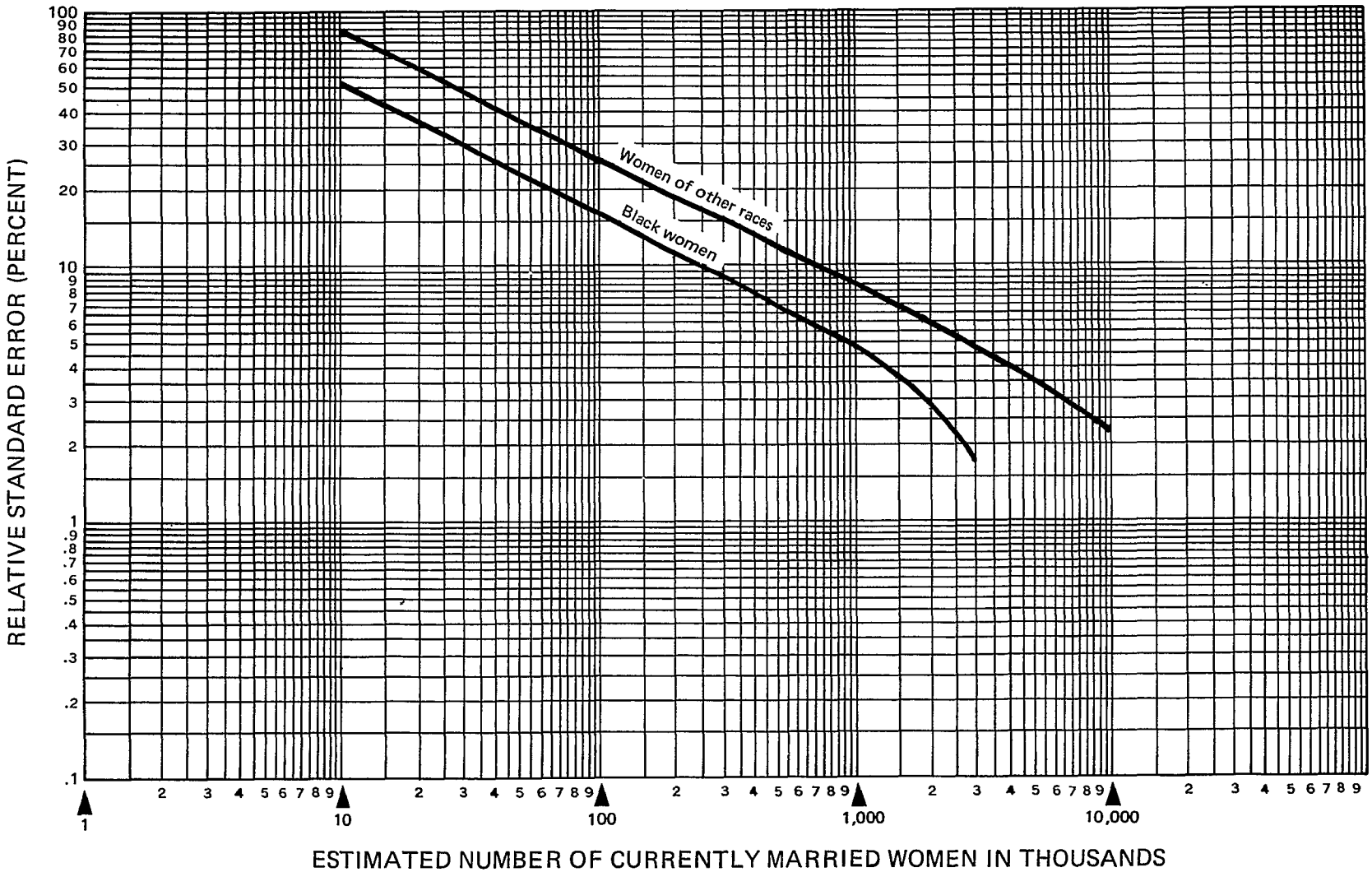
Table M shows estimates of A and B for relative standard error curves by type of statistic and race. For each type of statistic, separate estimates were produced for black women, for women of other races, and for women of all races combined, because black women were sampled at a higher rate than other women. Thus an estimate of a given number of black women has a smaller relative standard error than an estimate of the same number of women of other races. For example, an estimate of 500,000 currently married black women has a relative standard error of 7 percent while an estimate of 500,000 women of other races has a relative standard error of 12 percent. This relationship is also reflected in figure 3, which shows the relative standard error curves, by race, for currently married women.

The relative variances of the aggregate statistics are used to derive the relative variances of percents, which are ratios of two aggregates with the numerator being a subclass of the denominator. The relative standard error (RSE) of a percent estimate

Table M. Estimates of parameters A and B for relative standard error curves, by type of statistic and race

Type of statistic and race	Parameter A	Parameter B
Number of currently married women		
All races	−.0001858989	6751.0619
Black	−.0006310400	2798.6440
Other races	−.0002056235	7021.1665
Number of ever-married women		
All races	.0001700390	6486.5185
Black	−.0004520643	2848.2362
Other races	.0000422037	7111.5185
Number of ever-married women and never-married women with offspring		
All races	−.0001926913	6494.6569
Black	−.0004813358	2698.6043
Other races	−.0002362857	6892.2852
Number of children born to ever-married women		
All races	−.0001015087	18450.3253
Black	−.0003690139	9111.1866
Other races	−.0001792083	19967.9466

Figure 3. Relative standard errors for aggregates of currently married women, by race



Example of use of chart: An aggregate of 1 million black women (on the scale at the bottom of the chart) has a relative standard error of 4.7 percent, or a standard error of 47,000 (4.7 percent of 1 million).

$$P' = \frac{Y'}{Z'} \cdot 100$$

is given by the expression[12]

$$\begin{aligned} \text{RSE}(P') &= \sqrt{[\text{RSE}^2(Y') - \text{RSE}^2(Z')]} \\ &= \sqrt{A + \frac{B}{Y'} - \left(A + \frac{B}{Z'}\right)} \\ &= \sqrt{\frac{BZ' - BY'}{Y'Z'}} \\ &= \sqrt{\frac{BZ' - BY'(P'/Y')}{Y'Z'(P'/Y')}} \\ &= \sqrt{\frac{B(100 - P')}{P'Z'}} \end{aligned}$$

where B is the least squares estimate from the relative error curve for Y' and Z' (table M).

Notice that the relative standard error of P' is a function of the values of both P' and Z'. This relationship is demonstrated in figure 4, which shows separate relative standard error curves for percents based on different numbers of currently married women of all races combined. Each curve satisfies the equation

$$\text{RSE}(P') = \sqrt{\frac{6751.0619\,(100 - P')}{P'Z'}}$$

where P' is the estimated percent and Z' is the denominator of P'.

An estimate of the standard error of the difference between any two aggregates or percents is given by

$$\begin{aligned} s_{(Y_1' - Y_2')} &= \sqrt{s_{Y_1'}^2 + s_{Y_2'}^2} \\ &= \sqrt{(Y_1')^2\,\text{RSE}^2(Y_1') + (Y_2')^2\,\text{RSE}^2(Y_2')} \end{aligned}$$

This expression provides a good estimate of the standard error for uncorrelated statistics, but it can only be considered a rough approximation otherwise. Because estimates from Cycle II of the National Survey of Family Growth are based on a large sample of women, the distributions of Y_1' and Y_2' (and, therefore, $Y_1' - Y_2'$) are approximately normal. Frankel[13] shows empirically that, using balanced half-sample replication estimates of variance, the test statistic

$$t = \frac{Y_1' - Y_2'}{s_{(Y_1' - Y_2')}}$$

approximates the student's t distribution under the null hypothesis of no difference between the parameters estimated by Y_1' and Y_2' against a two-sided alternative. The number of replicates in the replication design (40 for Cycle II) can reasonably be used as the number of degrees of freedom for the t statistic, although the exact value for the degrees of freedom remains unknown. Therefore, individual two-tailed significance tests of differences between statistics from Cycle II data can be performed with an approximate significance level of α by computing t and comparing it to the two-tailed 1-α critical value for the t distribution with 40 degrees of freedom.

Example: In 1976, 29.0 percent of 24,795,000 currently married white women had been surgically sterilized, compared to 21.6 percent of 2,169,000 currently married black women. To test this racial difference at the $\alpha = .05$ level of significance, compute

$$t = \frac{29.0 - 21.6}{\sqrt{(29.0)^2 \cdot \text{RSE}^2(29.0) + (21.6)^2 \cdot \text{RSE}^2(21.6)}}$$

from table M

$$\begin{aligned} \text{RSE}(29.0) &= \sqrt{\frac{7021.1665\,(100 - 29.0)}{(29.0)(24{,}795{,}000)}} \\ &= .026 \end{aligned}$$

and

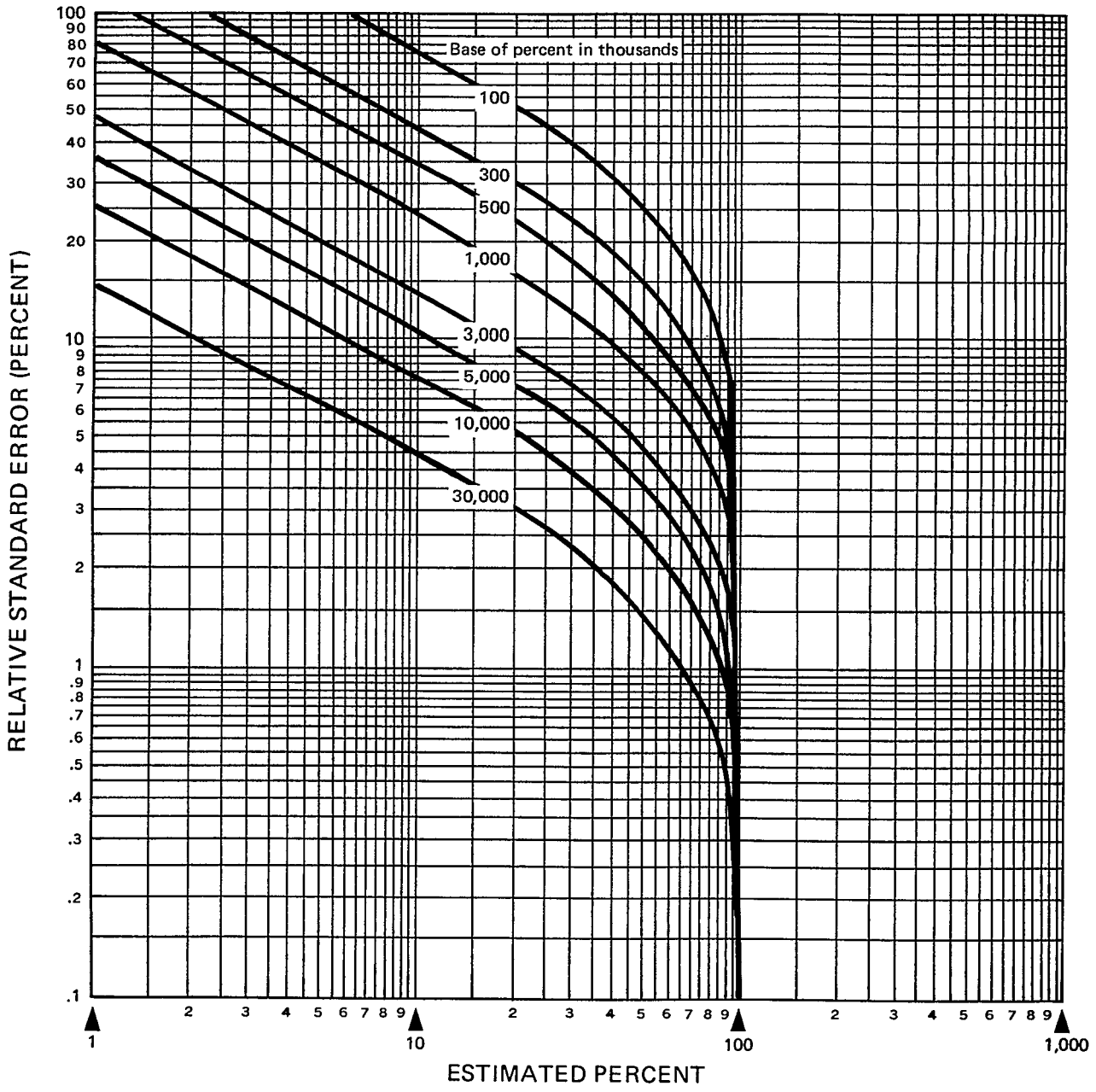
$$\begin{aligned} \text{RSE}(21.6) &= \sqrt{\frac{2798.6440\,(100 - 21.6)}{(21.6)(2{,}169{,}000)}} \\ &= .068 \end{aligned}$$

thus

$$\begin{aligned} t &= \frac{29.0 - 21.6}{\sqrt{(29.0)^2(.026)^2 + (21.6)^2(.068)^2}} \\ &= 4.48 \end{aligned}$$

The two-tailed .95 critical value (1 - α) for a t statistic with 40 degrees of freedom is 2.02. Therefore, the difference is significant at the .05 level.

Figure 4. Relative standard errors for percent of currently married women of all races



Example of use of chart: An estimate of 10 percent (from the scale at the bottom of the chart) of a population of 1 million women (fourth curve from top) has a relative standard error of 24.6 percent, or a standard error of 2.5 percent (24.6 percent of 10 percent).

REFERENCES

[1]National Center for Health Statistics: National Survey of Family Growth, Cycle I: Sample design, estimation procedures, and variance estimation, by D. K. French. *Vital and Health Statistics.* Series 2-No. 76. DHEW Pub. No. (PHS) 78-1350. Public Health Service. Washington. U.S. Government Printing Office, Jan. 1978.

[2]U.S. Bureau of the Census: *Census of Population and Housing, 1970, Census Tracts,* Final Report PHC(1). Washington. U.S. Government Printing Office, 1972. pp. 1-238.

[3]U.S. Bureau of the Census: *Housing Authorized by Building Permits and Public Contracts,* Monthly Summary Report C40, Sept. 1969-July 1975. Washington. U.S. Government Printing Office.

[4]National Center for Health Statistics: Replication: An approach to the analysis of data from complex surveys, by P. J. McCarthy. *Vital and Health Statistics.* PHS Pub. No. 1000-Series 2-No. 14. Public Health Service. Washington. U.S. Government Printing Office, Apr. 1966. p. 12.

[5]National Center for Health Statistics: Pseudoreplication: Further evaluation and application of the balanced half-sample technique, by P. J. McCarthy. *Vital and Health Statistics.* PHS Pub. No. 1000-Series 2-No. 31. Public Health Service. Washington. U.S. Government Printing Office, Jan. 1969.

[6]National Center for Health Statistics: Distribution and properties of variance estimators for complex multistage probability samples: An empirical distribution, by J. A. Bean. *Vital and Health Statistics.* Series 2-No. 65. DHEW Pub. No. (HRA) 75-1339. Health Resources Administration. Washington. U.S. Government Printing Office, Mar. 1975.

[7]Kish, L.: *Survey Sampling.* New York, John Wiley & Sons, 1965.

[8]Simmons, W. R., and Baird, J.: Use of pseudoreplication in the NCHS Health Examination Survey, in *Proceedings of the Social Statistics Section of the ASA.* American Statistical Association, 1968.

[9]Kish, L., and Frankel, M. R.: Balanced repeated replications. *J. Am. Stat. Assoc.* 65(331): 1071-1094, Sept. 1970.

[10]Plackett, R. L., and Burman, J. P.: The design of optimum multifactorial experiments. *Biometrika.* 33(2): 305-325, Aug. 1946.

[11]Gurney, M.: The variance of the replication method for estimating variances for the CPS sample design. U.S. Bureau of the Census, 1962. Unpublished document.

[12]National Center for Health Statistics: Estimation and sampling variance in the Health Interview Survey, by J. A. Bean. *Vital and Health Statistics.* PHS Pub. No. 1000-Series 2-No. 38. Public Health Service. Washington. U.S. Government Printing Office, June 1970.

[13]Frankel, M. R.: *Inference from Survey Samples: An Empirical Investigation.* Ann Arbor. Institute for Social Research, University of Michigan, 1971.

APPENDIXES

CONTENTS

APPENDIX I

GLOSSARY OF TERMS

Conterminous United States.—The land area consisting of the District of Columbia and all States except Alaska and Hawaii.

Dwelling unit (DU).—A single room, or group of rooms, that is intended for separate living quarters. The people who live there must live and eat separately from everyone else in the building (or apartment) *and* the room or group of rooms must have either

1. A separate entrance directly from the outside of the building or through a common hall, or
2. Complete kitchen facilities for the use of this household only. Complete kitchen facilities include *all* of the following:
   a. a range or cooking stove, and
   b. a sink with piped water, and
   c. a mechanical refrigerator.

Education.—The highest grade of school completed.

Geographic region.—For the purpose of classifying the population by geographic area, the U.S. Bureau of the Census has grouped the 50 States and the District of Columbia into four regions, as follows:

Region	*States included*
Northeast.....	Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, Pennsylvania
North Central...	Michigan, Ohio, Indiana, Illinois, Wisconsin, Minnesota, Iowa, Missouri, North Dakota, South Dakota, Kansas, Nebraska
South.....	Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida, Kentucky, Texas, Tennessee, Alabama, Mississippi, Arkansas, Louisiana, Oklahoma
West.....	Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah, Nevada, Washington, Alaska, Oregon, California, Hawaii

Alaska and Hawaii are not included in the NSFG sample design.

Household.—A family living together, or five or fewer unrelated individuals living together in a DU.

Parity.—The number of live births a woman has had.

Screener interview.—A preliminary interview at the household to collect information about the DU and to determine whether or not the household includes one or more women who are eligible for the detailed interview.

Standard metropolitan statistical area (SMSA).—A county or group of contiguous counties (except in New England) which contains at least one central city of 50,000 people or more, or "twin cities" with a combined population of at least 50,000. In addition, other con-

tiguous counties are included in an SMSA if, according to certain criteria, they are socially and economically integrated with the central city.

Urban area.—As defined by the U.S. Bureau of the Census, the urban areas of the United States include all cities or "twin cities" with at least 50,000 population in 1970 together with the surrounding closely settled area and all other incorporated or unincorporated population centers with 2,500 inhabitants or more.

Metropolitan state economic area (MSEA).—In New England, a county with more than half its population in one or more standard metropolitan statistical areas was classified as a metropolitan state economic area if the county or a combination of counties containing the standard metropolitan statistical area or areas had 100,000 inhabitants or more. For further discussion see: U.S. Bureau of the Census: *State Economic Areas*, Washington, D.C., U.S. Government Printing Office, 1951.

———— ooo ————

APPENDIX II

HOUSEHOLD SCREENER

OFFICE USE ONLY
Date Received _____
Recheck: Yes.....1 No.....2
Preliminary Disposition: _____
Final Disposition: _____
ID # [ ][ ][ ][ ][ ] (1 2 3 4 5) [0] (10)

A

BEGIN DECK 25

OMB No. 68-S74071
Expires: Dec. 31, 1976

Collected for the National Center for Health Statistics by Westat, Inc.

HOUSEHOLD SCREENER

ASSURANCE OF CONFIDENTIALITY: In accordance with Section 308(d) of the Public Health Service Act (42 USC 242m) and the Privacy Act of 1974 (5 USC 552a), the National Center for Health Statistics assures each respondent that all information which would permit identification of any individual or family will be held in strict confidence, will be used only by persons engaged in, and for purposes of, this study and will not be disclosed to others for any purposes.

ASSIGNMENT BOX: (13 14 15 16 17 18 19 20 21)

SAMPLING TABLE

IF NUMBER OF ELIGIBLE FEMALES LISTED IN SUMMARY BOX IS: ↓	THEN INTERVIEW PERSON LISTED ON SUMMARY BOX LINE ↓
Two	1
Three	3
Four	1
Five	4
Six or More	3

INTRODUCTION

Hello, I'm _____ from Westat Research, Inc. (SHOW ID BADGE) A letter was sent to you recently explaining the study we are conducting for the U.S. Public Health Service. As you may recall, from the letter, the study is being conducted all over the country and is about family size.

CHECK IF R VOLUNTEERS: Did not receive letter or does not remember letter. . . ◯ [22]

Received but did not read . . . . . ◯

ASK AT SAMPLED HOUSEHOLD AND RECORD VERBATIM. IF A SINGLE YEAR OR RANGE IS GIVEN THAT DEFINITELY FITS ONE OF THE STATED CATEGORIES, CHECK AND FOLLOW INSTRUCTIONS; OTHERWISE CHECK D.K. AND USE PROBE.

The sample of households we visit is scientifically selected to represent all households in our country. In order to be certain our sample is correct, I need to ask

When was this structure originally built? _____ (Year/Range) [23]

◯ 1970 or later. TERMINATE: Thank you very much for your help. My instructions are to interview at homes (apartment houses) built before 1970. Those built in 1970 or later are sampled separately. (COMPLETE NIR)

◯ Before 1970. CONTINUE INTERVIEW.

◯ D.K., No idea, etc. ASK: Was it built before 1970?

(Probably) Yes ◯ CONTINUE.

(Probably) No ◯ TERMINATE.

D.K., No idea, etc. ◯ CONTINUE.

ASSIGNED TO: _____ (Interviewer Signature) _____ (Interviewer No.) [24 25 26 27 28]

REASSIGNED TO: _____ (Interviewer Signature) (1) _____ (Interviewer No.) [29 30 31 32 33]

_____ (Interviewer Signature) (2) _____ (Interviewer No.)

CONTINUE DECK 25

TIME	AM
BEGAN	PM

CONDUCT ONLY WITH A HOUSEHOLD MEMBER AGE 15 OR OLDER

HOUSEHOLD ENUMERATION

34 35

S-1. To start, how many people live in this household? _____ NUMBER

S-2. What is the name of the head of this household? (ENTER NAME ON LINE 01 BELOW.)

S-3. And the other members of this household -- what are their names? Let's begin with everyone related to *(HEAD)*. (BE SURE PERSON INCLUDES [HIMSELF/HERSELF]) (ENTER NAMES IN TABLE BELOW.)

S-4. Are there other people living here who are not related to *(HEAD)*? (IF YES, ENTER NAMES IN TABLE BELOW) ◯ Yes ◯ No

S-5. I have listed (READ NAMES IN ORDER). Is there anyone else living here now, such as friends, relatives or roomers? (IF YES, ENTER NAME BELOW) ◯ Yes ◯ No

PERSON NUMBER	AFTER LISTING HOUSEHOLD, ASK S-6 THROUGH S-9 FOR EACH PERSON AS APPROPRIATE: First Name	Last Name	S-6. What is *(PERSON's)* relationship to *(HEAD OF HOUSEHOLD)*?	S-7. CODE SEX (ASK IF NOT OBVIOUS) M	F	S-8. How old was *(HEAD/PERSON)* on (his/her) last birthday?	S-9. IF 15 YEARS OR OLDER ASK: Is *(PERSON)* now married, widowed, divorced, separated, or has (he/she) never been married? (NEVER MARRIED BUT REPORTED LIVING TOGETHER, CODE (2); IF NEVER MARRIED AND NOT LIVING TOGETHER AND HAS OWN CHILDREN IN HOUSEHOLD, CODE (6). Mar.	Inform. Mar.	Wid.	Div. Ann.	Separated	Sing. with own children	Never Mar.
01			HEAD	1	2		1	2	3	4	5	6	7
02				1	2		1	2	3	4	5	6	7
03				1	2		1	2	3	4	5	6	7
04				1	2		1	2	3	4	5	6	7
05				1	2		1	2	3	4	5	6	7
06				1	2		1	2	3	4	5	6	7
07				1	2		1	2	3	4	5	6	7
08				1	2		1	2	3	4	5	6	7
09				1	2		1	2	3	4	5	6	7
10				1	2		1	2	3	4	5	6	7
11				1	2		1	2	3	4	5	6	7
12				1	2		1	2	3	4	5	6	7

(IF MORE THAN 12 HOUSEHOLD MEMBERS, GO TO CONTINUATION BOOKLET, PAGE 2)

(IN S-10, CONTINUE QUESTION WITH "PROBE" IF CHILD IS LISTED IN HOUSEHOLD BUT PARENT[s] ARE NOT IDENTIFIED.)

S-10. Is there anyone now away from home who usually lives here (PROBE: such as the mother of *[CHILD]*)?

(IF HOUSEHOLD MEMBER, ENTER NAME ABOVE.) ◯ Yes ◯ No

S-11. Do any of the people in this household have a home anywhere else?

(IF YES, PROBE FOR USUAL RESIDENCE. IF NOT HOUSEHOLD MEMBER, DRAW LINE THROUGH NAME ABOVE.) ◯ Yes ◯ No

S-12. Are any of the persons in this household now on full-time active duty with the Armed Forces of the United States? (IF NOT HOUSEHOLD MEMBER, DRAW LINE THROUGH NAME ABOVE) ◯ Yes ◯ No

36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52

SELECTION OF RESPONDENT

CONTINUE DECK 25 [53][54]

S-13. NO ELIGIBLE RESPONDENT by reason of: (CHECK CIRCLE AND SKIP TO S-15)

○ 1. Group Quarters (REQUIRES NIR FORM) ○ 2. Race ○ 3. Sex ○ 4. Age/Marital Status

S-14. ELIGIBLE RESPONDENT between 15 and 44 years old. (CHECK ONLY ONE CIRCLE)

ONE ELIGIBLE WOMAN:

○ 1. Currently or informally married;
- Circle R's Person No. on page 2;
- Skip to S-15; USE
CURRENTLY MARRIED QUEX

○ 2. Post married or SWOC;
- Circle R's Person No. on page 2;
- Skip to S-15; USE
POST MARRIED QUEX

MORE THAN ONE ELIGIBLE WOMAN (FOLLOW STEPS 1-4):

Step 1: List names of eligible women in Summary Box, in order of age, beginning with the oldest on Line #1.

Step 2: Use Sampling Table on Page 1 to determine which eligible woman to interview.

Step 3: Circle R's line number in Summary Box and write selected R's name here:

_____
Respondent's Name

Step 4: Selected Respondent is:

SUMMARY BOX

Line #	Name	Age
1.		
2.		
3.		
4.		
5.		
6.		

○ 3. Currently or informally married
- Circle R's Person No. on page 2;
- Skip to S-15; USE
CURRENTLY MARRIED QUEX

○ 4. Post married or SWOC;
- Circle R's Person No. on page 2;
- Skip to S-15; USE
POST MARRIED QUEX

ASK EVERYONE

S-15. Is there a phone number where you can be reached (in case my office wants to verify this interview)?

Telephone No. (Area Code) _____ (S-16) [55]
No Phone 6 (S-17)
Refused. 7 (S-17)

S-16. Is this phone in your home or elsewhere?

In household 1
In home of neighbor. 2
Other (SPECIFY) _____3

S-17. TIME SCREENER ENDED _____ AM / PM [56][57]

S-18. Does Assignment Box require:

the Missed DU Procedure? ○ No ○ Yes
the Missed Structure Procedure? ○ No ○ Yes

IF NO TO BOTH OF THE ABOVE, CONTINUE.
IF YES TO EITHER OF THE ABOVE →
- AND ELIGIBLE "R" AVAILABLE NOW, COMPLETE REQUIRED PROCEDURES AFTER EXTENDED INTERVIEW.
- AND NO ELIGIBLE "R" OR ELIGIBLE "R" NOT AVAILABLE, COMPLETE MISSED DU PROCEDURE AND FORM - PAGE 4, AND/OR MISSED STRUCTURE PROCEDURE AS OUTLINED ON MISSED STRUCTURE FORM, NOW.

INTERVIEWER: FILL OUT S-19 THROUGH S-23 BELOW IMMEDIATELY AFTER YOU LEAVE THE HOUSEHOLD.

S-19. Code Type of Structure:

Detached single family house. 1
Trailer 2
2-4 Family house/apartment building . 3
Row house (3 or more attached units). 4
Apartment house (5 or more units; free access to housing units). . . . 5 [58]
Apartment house (5 or more units; locked entry, or guarded by doorman, or both). 6
Other (SPECIFY) _____
_____7

S-20. Race of Household (BY OBSERVATION):

Black/Negro 1
White 2
Other (SPECIFY) _____ [59]
_____3
Not able to observe 4

S-21. Date of Screener Interview

___/___/___
MONTH DAY YEAR
[60][61][62][63][64][65]

S-22. With whom did you conduct the screener?

Person No. from pg. 2 _____
[66][67] [68][69]

S-23. This screener interview was conducted in:

English 1
Spanish . . . . . 2 [70]
Other (SPECIFY) _____ 3

CONTINUE DECK 25

71 | 72 73 74 75

MISSED D.U. PROCEDURE

1. Before leaving household, say: "We want to be sure that every household in this area has been given a chance to participate in this important survey. At this address we listed _____ households (in your structure/on this floor). Are there any other living quarters here that we may have missed?"

2. Also, check in the lobby and around the outside of this (house/building) for additional units or entrances in this structure.

3. Record discovered D.U.'s on form below. If no additional D.U.'s, check circle in the box in the upper left-hand corner of the form.

4. If 1 to 4 Missed D.U.'s are discovered, fill out an assignment box on a blank Screener for each (instructions for how to do this are in the Interviewer's Manual on page 10-54) and conduct screener interview. Add the discovered D.U.'s to the Listing Sheet, Selected DU List, and all copies of the Interviewer Weekly Status Report.

5. If 5 or more Missed D.U.'s are discovered, call supervisor for instructions before you do any additional screener interviews. Add all of the discovered D.U.'s to the Listing Sheet and the selected sample D.U.'s to the Selected DU List and all copies of the Interviewer Weekly Status Report. Then fill out an assignment box on a blank Screener for each selected sample D.U. and conduct screener interview.

MISSED D.U. FORM

CHECK (✓) IF NO MISSED D.U. AT SAMPLED STRUCTURE ○

PSU #_____ SEG #_____

D.U. # ASSIGNED	ADDRESS OF DISCOVERED D.U.

Number discovered D.U.'s sequentially within segments beginning with D.U. 800 or 801 as instructed on Selected DU List. Each number must be assigned only once within a segment.

TOTAL ADDITIONAL D.U.'s ☐

NOTE: Be sure to thank respondent and complete page 3, S-19 through S-23.

NON-INTERVIEW REPORT FORM

BEGIN DECK 2E

N-1. Why were you unable to complete (screener/extended) interview?

- Vacant or Not a Dwelling Unit . 1 *(N-2)*
- Household Screener NIR. . . . . 2 *(N-4)*
- Extended Interview NIR. . . . . 3 *(N-5)*

N-2. Why is the listed address not an occupied dwelling unit for our sample?

- Condemned 01
- Demolished. 02
- Place of Business 03
- No such address/No such DU. . .04
- Group quarters. 05
- Vacation Cabin. 06
- Not usable as permanent residence. . . . . .07 *(N-3)*
- Transient use (less than 1 month). 08
- Not a DU for other reason . . .09
- Still under construction. . . .10
- Post 1970 structure 11
- Improperly listed, out of segment 12
- Vacant. . . . . .13 *(N-13)*

N-3. Is there any additional information regarding this unit?

GO TO N-14

N-4. Whom did you contact in the household?

- No one. 1
- Head or spouse of head. 2
- Relative of HH. 3
- Non-related adult in HH . . . . 4 *(N-6)* 13
- Child under 15 years of age . . 5
- Other (SPECIFY) _____
- _____ 6

N-5. Unable to complete extended interview with respondent who is:

- Currently married 1
- Post married or SWOC. 2

N-6. What was the problem in obtaining information?

- Unable to enter structure . . . 22 *(N-8)*
- No respondent after 4 calls . . 23 *(N-8)*
- Language Problem (SPECIFY) ____
- _____ 24 *(N-10)*
- Unavailable during field period . . . . . 25 *(N-8)*
- Too ill . . . . . 26 *(N-8)*
- Breakoff. . . . . 27 *(N-7)*
- Refusal . . . . . 28 *(N-7)*
- Other (SPECIFY) _____
- _____ 29 *(N-7)*

N-7. What was the reason you could not complete this (screener/extended) interview? (RECORD ANY EXPLANATIONS "R" GAVE AND YOUR OWN IMPRESSIONS. THEN CODE THE REASON YOU BELIEVE IS THE MOST IMPORTANT.)

_____
_____
_____
_____
_____
_____

- Did not want to answer questions, did not believe in surveys. 1
- Did not have time, didn't want to be bothered 2
- Afraid to let interviewer in, afraid to answer, told not to answer questions. . 3
- Objected to this particular survey . . . 4
- Claimed this survey did not apply to HH. 5
- Other (SPECIFY)_____
- _____ 6
- Could not determine any reason 7

CONTINUE DECK 26

N-8. Name and phone number of sample household, if available.

NAME _____ Phone ( ) -

N-9. Race of household?

Black 1
White 2
Other (SPECIFY) _____ 3
Could not find out. 8

14

N-10. What information could you find out as to the best time and/or circumstances at which the (screener/extended) interview could be obtained?

IF EXTENDED INTERVIEW NIR, SKIP TO N-13.

N-11. Have you learned anything about the age and marital status of the women in the assigned household? (EXPLAIN)

N-12. Based on the information you have obtained, do you think the assigned household is:

Definitely eligible 1
Probably eligible 2
Probably not eligible 3
Definitely not eligible 4
Don't know. 8

15

N-13. Code the type of structure?

Detached single family house. . 1
Trailer or trailer space. . . . 2
2-4 family house/apartment building 3
Row house (3 or more attached units). 4
Apartment house (5 or more units; free access to housing unit). 5
Apartment house (5 or more units; locked entry or guarded by doorman or both). . 6
Other (SPECIFY) _____

_____ 7

16

N-14. Date of NIR

MONTH / DAY / YEAR

17 18 19 20 21 22

NAME OF SUPERVISOR WHO APPROVED NIR: _____

DATE: _____

N. I. R.

SUPERVISOR ONLY: SUMMARIZE PROBLEMS ENCOUNTERED, ACTIONS TAKEN AND YOUR RECOMMENDATIONS FOR MAIN OFFICE FOLLOWUP:

NAME: _____

I.D. Number ___/___/___ ___/___/___
MONTH/ DAY/ YEAR

MAIN OFFICE USE ONLY:

ACTION(S):
- Telephone Call . . . . . ○
- Letter . . . . . ○

} ENTER ON RECORD OF HH ACTION

- Returned to field for additional followup . . . . ○
- Determined to be Final NIR . ○

NAME: _____

DATE: _____

RECORD OF HOUSEHOLD ACTION

Date	Day of Week	Time	TYPE: Personal	TYPE: Telephone	TYPE: Mail	RESULT: Screener (USE CODES)	RESULT: Extended Interview (USE CODES)	NON-ELIGIBLE INFORMANT CONTACT: Non-HU Member (CHECK BOX)	NON-ELIGIBLE INFORMANT CONTACT: Underage HU Member (CHECK BOX)	NON-ELIGIBLE INFORMANT CONTACT: Neighbor (CHECK BOX)	REMARKS	FIELD Intv. No. / Main Office NAME	Is this contact a call? Yes (CHECK BOX)	Is this contact a call? No (CHECK BOX)

IF VACANT OR NOT A D.U., COMPLETE NIR

RESULT CODES

SCREENER		EXTENDED INTERVIEW (IF SCREENER COMPLETE USE CODES, OTHERWISE LEAVE BLANK)	
Vacant/not DU	0	No eligible R	0
Screener completed	1	Interview completed	1
Unable to enter structure	2	Unable to enter structure	2
Eligible screener R not home	3	Eligible R not home	3
Language problems	4	Language problems	4
Unavailable during field period	5	Unavailable during field period	5
Too ill	6	Too ill	6
Breakoff	7	Breakoff	7
Refusal	8	Refusal	8
Other (Specify under Remarks)	9	Other (Specify under Remarks)	9

VITAL AND HEALTH STATISTICS Series

Series 1. Programs and Collection Procedures.—Reports which describe the general programs of the National Center for Health Statistics and its offices and divisions and data collection methods used and include definitions and other material necessary for understanding the data.

Series 2. Data Evaluation and Methods Research.—Studies of new statistical methodology including experimental tests of new survey methods, studies of vital statistics collection methods, new analytical techniques, objective evaluations of reliability of collected data, and contributions to statistical theory.

Series 3. Analytical Studies.—Reports presenting analytical or interpretive studies based on vital and health statistics, carrying the analysis further than the expository types of reports in the other series.

Series 4. Documents and Committee Reports.—Final reports of major committees concerned with vital and health statistics and documents such as recommended model vital registration laws and revised birth and death certificates.

Series 10. Data From the Health Interview Survey.—Statistics on illness, accidental injuries, disability, use of hospital, medical, dental, and other services, and other health-related topics, all based on data collected in a continuing national household interview survey.

Series 11. Data From the Health Examination Survey and the Health and Nutrition Examination Survey.—Data from direct examination, testing, and measurement of national samples of the civilian noninstitutionalized population provide the basis for two types of reports: (1) estimates of the medically defined prevalence of specific diseases in the United States and the distributions of the population with respect to physical, physiological, and psychological characteristics and (2) analysis of relationships among the various measurements without reference to an explicit finite universe of persons.

Series 12. Data From the Institutionalized Population Surveys.—Discontinued effective 1975. Future reports from these surveys will be in Series 13.

Series 13. Data on Health Resources Utilization.—Statistics on the utilization of health manpower and facilities providing long-term care, ambulatory care, hospital care, and family planning services.

Series 14. Data on Health Resources: Manpower and Facilities.—Statistics on the numbers, geographic distribution, and characteristics of health resources including physicians, dentists, nurses, other health occupations, hospitals, nursing homes, and outpatient facilities.

Series 20. Data on Mortality.—Various statistics on mortality other than as included in regular annual or monthly reports. Special analyses by cause of death, age, and other demographic variables; geographic and time series analyses; and statistics on characteristics of deaths not available from the vital records based on sample surveys of those records.

Series 21. Data on Natality, Marriage, and Divorce.—Various statistics on natality, marriage, and divorce other than as included in regular annual or monthly reports. Special analyses by demographic variables; geographic and time series analyses; studies of fertility; and statistics on characteristics of births not available from the vital records based on sample surveys of those records.

Series 22. Data From the National Mortality and Natality Surveys.—Discontinued effective 1975. Future reports from these sample surveys based on vital records will be included in Series 20 and 21, respectively.

Series 23. Data From the National Survey of Family Growth.—Statistics on fertility, family formation and dissolution, family planning, and related maternal and infant health topics derived from a biennial survey of a nationwide probability sample of ever-married women 15-44 years of age.

For a list of titles of reports published in these series, write to:

Scientific and Technical Information Branch
National Center for Health Statistics
Public Health Service
Hyattsville, Md. 20782

DHHS Publication No. (PHS) 81 1361
Series 2-No. 87

NCHS

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
Office of Health Research, Statistics, and Technology
National Center for Health Statistics
3700 East-West Highway
Hyattsville, Maryland 20782

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

For publications in the *Vital and Health Statistics* Series call 301-436-NCHS.

POSTAGE AND FEES PAID
U.S. DEPARTMENT OF H.H.S.
HHS 396

THIRD CLASS

U.S.MAIL