DATA EVALUATION
AND METHODS RESEARCH

Series 2
Number 76

National Survey of Family Growth, Cycle I:
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 I of the National Survey of Family Growth.

DHEW Publication No. (PHS) 78-1350

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Public Health Service
National Center for Health Statistics
Hyattsville, Md. January 1978

Library of Congress Cataloging in Publication Data

French, Dwight K.
National survey of family growth, Cycle I.
(Vital and health statistics; Series 2, Data evaluation and methods research; no 76) (DHEW publication; (PHS) 78-1350)
Includes bibliographical references.
Supt. of Docs. no.: HE 20.6209:2/76
1. Family size—United States—Statistical methods. 2. Fertility, Human. 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. 76. III. Series, United States Dept. of Health, Education, and Welfare. DHEW publication; (PHS) 78-1350.
RA409.U45 no. 76 [HQ766.5.U5] 312'.07'23s [301.42'1]
ISBN 0-8406-0116-6 77-608168

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

Stock No. 017-022-00604-9

NATIONAL CENTER FOR HEALTH STATISTICS

COOPERATION OF THE NATIONAL OPINION RESEARCH CENTER

In accordance with specifications established by the National Center for Health Statistics, the National Opinion Research Center at the University of Chicago, 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. 76

DHEW Publication No. (PHS) 78-1350
Library of Congress Catalog Card Number 77-608168

PREFACE

This report presents a detailed description of the sample design, estimation procedures, and variance estimation method used in Cycle I of the National Survey of Family Growth. The survey was designed and conducted by the National Opinion Research Center (NORC) of the University of Chicago under a contractual arrangement with the National Center for Health Statistics (NCHS). The sampling plan was developed under the supervision of Martin Frankel and Benjamin King of NORC, in consultation with E. Earl Bryant, Monroe G. Sirken, and William F. Pratt of NCHS.

Much of the report, prepared by Dwight K. French of the Statistical Methods staff, is based upon survey specification documents prepared by NORC and upon internal NCHS memoranda. Dr. Frankel, Dr. King, Mr. Bryant, and Dr. Pratt, along with Drs. Gordon Bonham and Dwight Brock of NCHS, were the primary resource persons for methodological questions.

In addition to 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 of the report but are not involved in producing the report. Mr. Garrie Losee of NCHS carried out the peer review of this report and made many constructive suggestions.

CONTENTS

LIST OF 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 (more than 30 percent relative standard error)	*

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

Dwight K. French, *Statistical Methods Staff*

INTRODUCTION

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

In response to the need for current information on the interrelated topics of fertility, family planning, and their effects on population growth, the National Survey of Family Growth (NSFG) was established as an integral part of the NCHS program in 1972. Since the purpose of the survey was to collect data relating to natality and the process of family formation and dissolution, it was placed within the Division of Vital Statistics. The NSFG is designed as a cyclic survey; that is, data are to be collected every few years by means of a sample survey.

The target population for Cycle I of the NSFG consisted of civilian noninstitutionalized women living in the conterminous United States who were less than 45 years of age and who were currently married, previously married, or single mothers with children living in the household at the time of interview. Data were collected by means of personal interviews with a probability sample of these women. The interviews furnished information for determining trends and differentials in fertility, family planning practices, sources of family planning advice and services, effectiveness and acceptability of various methods of family planning, and aspects of maternal and child health that are closely related to family planning.

The sample design and data collection for the first cycle of the NSFG were contracted to the National Opinion Research Center (NORC) of the University of Chicago. The sample consisted of 10,862 eligible women, of whom 9,800 (90.2 percent) were interviewed during the 8-month period from July 1973 through February 1974. Other reports will discuss the findings of the survey. This report describes in detail the sample design and sample selection procedures used in Cycle I, the techniques used to estimate population parameters, and the procedure used to estimate sampling variances.

DESIGN SPECIFICATIONS

The development of an efficient sample design must take into account the primary survey objectives, amount of funds available, logistical problems, time limitations, organized speculation concerning population parameters and unit operating costs. These requirements dictated a stratified multistage probability sample design for the NSFG, based essentially on the following set of specifications:

1. The target population was defined to be civilian noninstitutionalized women living in the conterminous United States

who were less than 45 years of age and who were either
(a) currently married,
(b) previously married, or
(c) single mothers with one or more of the 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. 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 III). 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 by means of personal interviews lasting an average of 1 hour.

4. All interviewers were required to 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 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). Neither screener nor interview response was supposed to fall below a minimum rate of 90 percent.

8. The contractor would design and implement procedures to measure and control the quality of data collection and data preparation.

9. For the population subgroups shown in table A, the sample design should yield estimated proportions whose standard errors are within the tolerances given in the rightmost column. The tolerances were based on a sample of 3,600 Negro women and 6,400 women of other races, and a design factor of approximately 1.4 (that is, it was assumed that the standard

Table A. Standard error tolerances of estimated proportions for selected population subgroups

Population subgroup	Expected proportion of sample women in the race group who are in the subgroup	Estimated proportion within the subgroup	Standard error tolerance for estimated proportion as specified by NCHS
Negro			
Education:			
Less than high school	.48	.42	.0164
High school and more	.52	.25	.0139
Parity:			
0-2 children	.50	.19	.0128
3 children or more	.50	.51	.0163
Other races			
Education:			
Less than high school	.30	.23	.0133
High school and more	.70	.15	.0074
Parity:			
0-2 children	.59	.10	.0068
3 children or more	.41	.33	.0127

error of statistics based on the NSFG design would be about 40 percent larger than their standard errors based on a simple random sample of the same size).

SAMPLE DESIGN

Summary

The sample design for Cycle I of the National Survey of Family Growth (NSFG) was a five-stage probability design based on the National Opinion Research Center's (NORC) 1972 Master Probability Sample.[1] The counties and independent cities that make up the total land area of the conterminous United States were combined to form a frame of primary sampling units (PSU's). From this frame, the first stage of sampling yielded 203 PSU's which were divided into four replicate groups, three containing 51 PSU's and the fourth having the remaining 50. Two of these replicate groups, containing 101 PSU's, were chosen for the NSFG sample. The next two stages resulted in the selection of several segments (clusters of about 100 dwelling units) from each sample PSU. A member of the NORC field staff listed the dwelling units (DU's) within each segment, and a fourth-stage systematic sample of DU's was selected. At each sample DU, an NORC interviewer attempted to complete the Household Screener questionnaire shown in appendix III and list the names of all eligible respondents. From this list one eligible woman was randomly selected for interviewing.

Determination of Sample Size

After the NSFG contract was awarded to NORC, they agreed to design the sample to produce the race allocation that NCHS suggested in their precision requirements—3,600 Negro women and 6,400 women of other races. Once the allocation was fixed, NORC proceeded to calculate the number of sample dwelling units that were needed to produce the final sample of women. They started by collecting 1970 census information on the number of occupied DU's with Negro heads and the number with heads of other races, as well as population counts of eligible women in the two race classes. The ratio of occupied DU's to eligible women for both race groups was adjusted to account for the following three factors:

1. 1970 census data indicated that 8 percent of all DU's in the United States were vacant.
2. Data from the NCHS Health Interview Survey indicated that 5 percent of all eligible women lived in households containing two or more eligible women.
3. Combined screener and interview nonresponse was expected to be 10 percent.

The adjusted DU-person ratios represented the expected number of DU's that would have to be screened in order to find and interview one sample woman in each race class. By multiplying the final ratios by the desired number of sample persons, NORC calculated the expected minimum number of DU's that would need to be screened to yield 3,600 completed interviews with Negro women and 6,400 completed interviews with women of other races—9,141 Negro DU's and 18,091 DU's of other races, or a total of 27,232 DU's.

These minimum numbers of DU's, however, do not represent the actual number that were required to be screened for the NSFG. In areas where a large proportion of the population was Negro, DU's were oversampled to attain the required number of Negro women. In order to keep the sample essentially self-weighting for women of other races, DU's of other races were subsampled in these areas. However, it was necessary to screen all DU's in order to determine their race. This subsampling procedure increased the number required to be screened from the minimum 27,232 to 31,842.

Stratification and Selection of PSU's

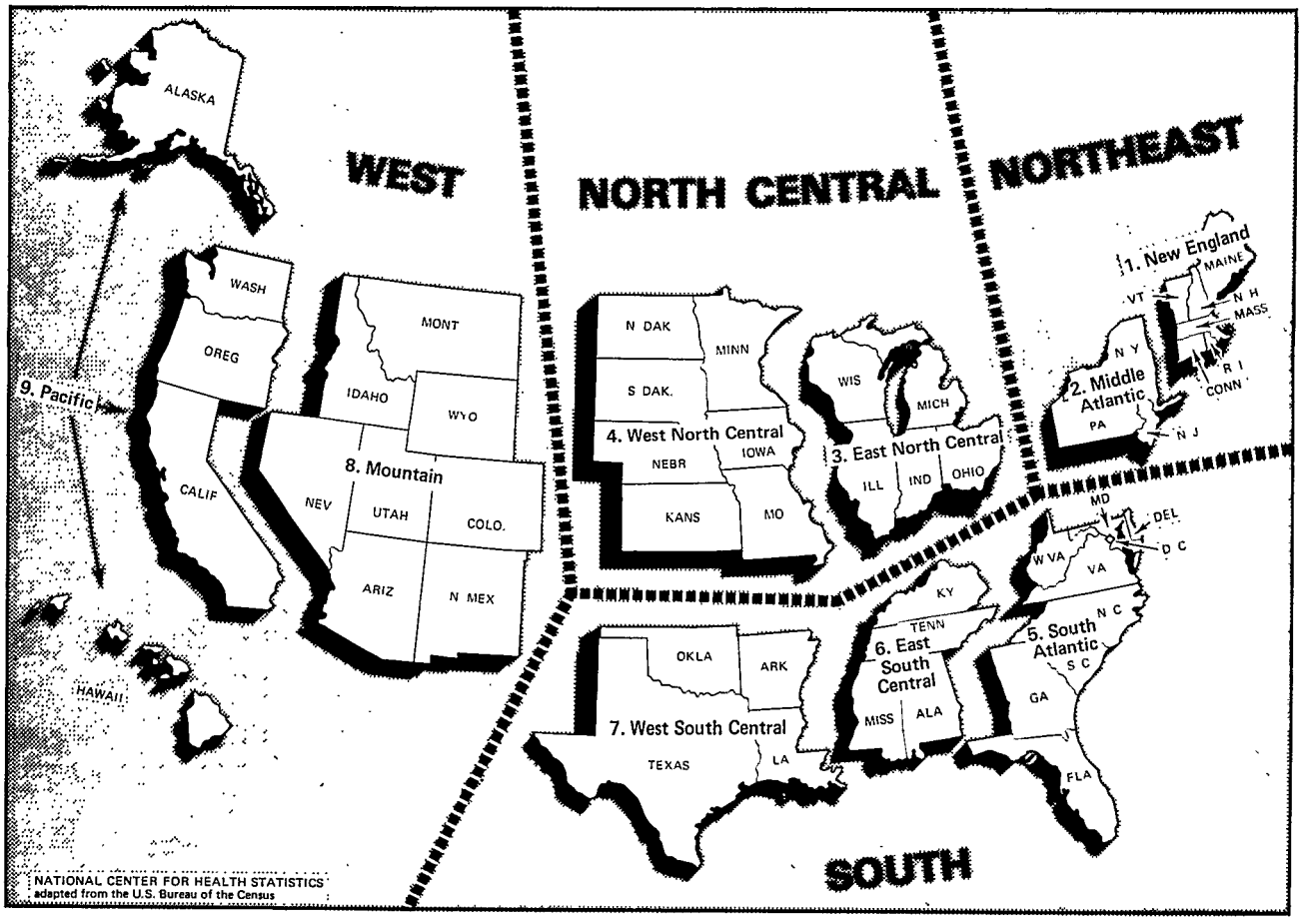
The PSU's in the NORC master sample were selected from separate sampling frames of standard metropolitan statistical areas (SMSA's) and nonmetropolitan areas in the conterminous United States. The SMSA frame consisted of the 246 SMSA's as defined by the U.S. Bureau of the Census in March 1971. The frame was

ordered in the following manner, based on advance 1970 census population data:

1. The SMSA's were first sorted by the nine census geographic divisions as shown in figure 1.
2. Within each geographic division, the SMSA's were sorted into three size classes: 1,000,000 persons or more, 200,000 to 999,999 persons, and less than 200,000 persons.
3. In divisions 1, 2, 3, 4, 8, and 9, the SMSA's in each size class were sorted by State, with the States placed in geographic order from northwest to northeast to southeast to southwest. Within each State the SMSA's were ordered in the same way. In divisions 5, 6, and 7, the SMSA's in each size class were placed in descending order of the number of residents of races other than white, in order to increase the likelihood of selecting an appropriate number of southern SMSA's with large Negro populations.

The total population of the SMSA frame based on preliminary 1970 census data was 138,789,636. After the ordering was completed, the frame was divided into 139 sequential zones, each having 1 million population (the 139th zone contained 789,636 persons and 310,364 "blanks"). In each zone the numbers from 1 to 1 million were assigned in ordered intervals to the SMSA's that were totally or partially included, each SMSA receiving an interval equal to its population within the zone. A "hit number" between 1 and 1 million was randomly and independently selected for each zone, and the SMSA whose assigned interval included that number was the PSU selected to represent the zone. The selection of a hit number was not

Figure 1. Geographic divisions of the United States, and their order for the first-stage sample



necessary for zones that were completely covered by one SMSA. This selection procedure differs somewhat from classical sampling with probability proportionate to size in that SMSA's that overlapped zones could be selected more than once, and very large SMSA's had to be selected several times. For example, the New York City SMSA represented 11 zones, all of which it covered completely. It was not selected to represent two other zones that it did not cover completely.

After the PSU's were selected, the 139 zones were systematically separated into four groups (zones 1, 5, 9, . . ., 137 constituted group 1; zones 2, 6, 10, . . ., 138 constituted group 2; etc.). Groups 1 and 3 were randomly selected, and the SMSA's that represented the zones in those two groups became part of the NSFG sample. The remaining two groups were held in reserve, but were not used in the NSFG. The sample of the SMSA population consisted of 70 PSU's, containing 56 distinct SMSA's (see table B).

All areas in the United States that were not

Table B. Sample PSU's for the National Survey of Family Growth

SMSA's
Akron, Ohio
Allentown-Bethlehem-Easton, Pa.-N.J.
Appleton-Oshkosh, Wis.
Atlanta, Ga.
Baltimore, Md.
Birmingham, Ala.
Boston, Mass.
Buffalo, N.Y.
Chicago, Ill. (represented by 4 PSU's)
Cleveland, Ohio
Columbia, S.C.
Columbus, Ohio
Dallas, Tex.
Denver, Colo.
Des Moines, Iowa
Detroit, Mich. (represented by 2 PSU's)
Flint, Mich.
Fresno, Calif.
Gary-Hammond-E. Chicago, Ind.
Great Falls, Mont.
Hartford, Conn.
Houston, Tex.
Indianapolis, Ind.
Jersey City, N.J.
Knoxville, Tenn.
Little Rock-N. Little Rock, Ark.
Los Angeles-Long Beach, Calif. (represented by 3 PSU's)
Miami, Fla.
Minneapolis-St. Paul, Minn.
Nashville, Tenn.
New Britain, Conn.
New Orleans, La.
New York, N.Y. (represented by 5 PSU's)
Newark, N.J. (represented by 2 PSU's)
Norfolk-Virginia Beach-Portsmouth, Va.
Oklahoma City, Okla.
Orlando, Fla.
Pensacola, Fla.
Philadelphia, Pa.-N.J. (represented by 2 PSU's)
Phoenix, Ariz.
Pittsburgh, Pa.
Portland, Oreg.-Wash.
Racine, Wis.
Reading, Pa.
Rochester, N.Y.
St. Louis, Mo.-Ill. (represented by 2 PSU's)
San Bernardino-Riverside-Ontario, Calif.
San Diego, Calif.
San Francisco-Oakland, Calif. (represented by 2 PSU's)
Santa Rosa, Calif.
Springfield-Chicopee-Holyoke, Mass.-Conn.
Stamford, Conn.
Tacoma, Wash.
Topeka, Kan.
Washington, D.C.
Wichita Falls, Tex.

Non-SMSA PSU's
Auglaize Co., Ohio
Burnett Co.-Washburn Co., Wis.
Carroll Co., Mo.
Columbiana Co., Ohio
Colusa Co., Calif.
Florence Co., S.C.
Gogebic Co., Mich.
Houston Co., Ala.
Iredell Co., N.C.
Knox Co., Ohio
Lincoln Co.-Rock Co., Minn.
Long Branch-Asbury Park, N.J.
Madison Co., S.C.
Manitowoc Co., Wis.
Marion Co., W. Va.
Marquette Co., Mich.
McDowell Co., W. Va.
Meklenburg Co., Va.
Mesa Co., Colo.
Newaygo Co., Mich.
St. Lucie Co., Fla.
Schuylkill Co., Pa.
Smyth Co., Va.
Sumter Co., Ga.
Sussex Co., N.J.
Upshur Co., W. Va.
Victoria Co., Tex.
Washington Co., Ala.
Wilkes Co., N.C.
Willacy Co., Tex.
Yuba Co., Calif.

included in the SMSA frame, can be classified either as counties, independent cities, or, in New England, portions of counties (for convenience, all such areas will hereafter be referred to as "counties"). In certain instances, sparsely populated counties were linked to provide an adequate population base for later stages of sampling. The non-SMSA frame consisted of the conterminous United States minus the 246 SMSA's, with the individual and linked counties ordered as follows:

1. Counties from census divisions 1 and 2 with 50,000 persons or more, arranged in descending order of population (DOP).
2. Counties from divisions 3 and 4 with 60,000 persons or more, arranged in DOP within State. The order of States for divisions 3 and 4 is given in table C.
3. Counties from divisions 3 and 4 with populations between 30,000 and 59,999, where the proportion of the population living in urban areas (as defined by the U.S. Census Bureau) was greater than or equal to 40 percent, arranged in DOP within State.
4. Counties from divisions 3 and 4 with populations less than 30,000, where the urban proportion was greater than or equal to 50 percent, arranged in DOP within State.
5. Counties from divisions 5, 6, and 7 with populations greater than or equal to 30,000, where the urban proportion was 30 percent or greater and the proportion of the population that was Negro was less than 20 percent, arranged in DOP within State. The order of the States is given in table C.
6. Counties from divisions 5, 6, and 7 with populations greater than or equal to 30,000, where the urban proportion was 30 percent or greater and the Negro proportion was 20 percent or greater, arranged in DOP within State.
7. Counties from division 8 arranged in DOP.
8. Counties from division 9 arranged in DOP.
9. Counties from divisions 1 and 2 with populations less than 50,000 arranged in DOP.
10. Counties from divisions 3 and 4 with populations between 30,000 and 59,999, where the urban proportion was less than 40 percent, arranged in DOP within State.
11. Counties from divisions 3 and 4 with populations less than 30,000, where the urban proportion was less than 50 percent, arranged in DOP within State.
12. Counties from divisions 5, 6, and 7 with populations greater than or equal to 30,000, where the urban proportion was less than 30 percent and the Negro proportion was less than 20 percent, arranged in DOP within State.
13. Counties from divisions 5, 6, and 7 with populations less than 30,000, where the Negro proportion was less than 20 percent, arranged in DOP within State.
14. Counties from divisions 5, 6, and 7 with populations greater than or equal to 30,000, where the urban proportion was less than 30 percent and the Negro proportion was 20 percent or greater, arranged in DOP within State.

Table C. Ordering of States for sampling from the National Opinion Research Center non-SMSA frame

Area	Order of States
North Central Region	
Divisions 3 and 4.....	Ohio, Michigan, Wisconsin, Minnesota, Indiana, Illinois, Missouri, Iowa, Kansas, Nebraska, North Dakota, South Dakota
South Region	
Divisions 5, 6, and 7.....	Delaware, Maryland, West Virginia, Virginia, Kentucky, Tennessee, North Carolina, South Carolina, Georgia, Alabama, Mississippi, Louisiana, Arkansas, Florida, Texas, Oklahoma.

15. Counties from divisions 5, 6, and 7 with populations less than 30,000, where the Negro proportion was 20 percent or greater, arranged in DOP within State.

The total population of the non-SMSA frame based on preliminary 1970 census counts was 63,456,729. This frame was partitioned in the same way as the SMSA frame into 64 zones of 1 million persons (zone 64 contained 543,271 "blanks"). The method of selecting a county to represent each zone was exactly the same method used for SMSA's; however, the method of determining the primary and reserve PSU's was different. The 64 zones were divided into 16 sets of 4 (zones 1-4 formed the first set, zones 5-8 the second set, and so forth). Each set of numbers was randomly permuted, and the counties representing the first two zones in the sequence were assigned to the NSFG sample. For example, the permutation of zones 1-4 was 3, 1, 4, 2. Therefore, the counties representing zones 3 and 1 were included in the survey, while the remaining two selections were placed in reserve status. Only 31 PSU's were selected from the non-SMSA frame because a blank was selected as the hit number in zone 64, and zone 64 was listed second in the permutation of set 16. Thus the first-stage non-SMSA sample consisted of 31 counties, none appearing more than once. These 31 counties, when added to the SMSA sample, produced a total first-stage sample of 101 PSU's from 87 distinct localities.

NORC's methods of ordering SMSA's and non-SMSA counties, selecting PSU's from zones of 1 million persons, and subsampling to determine primary and reserve PSU's ensured a reasonable geographic, racial, and urban-rural balance among the PSU's. Even after the preliminary ordering, NORC was concerned that a sample of four PSU's from zones of 4 million persons would result in excessive geographic clustering. They therefore decided on the more detailed method of selection from smaller zones. The systematic group method of selecting primary and reserve SMSA's was changed to random selection within sets for counties because the cyclical ordering of the non-SMSA frame might have caused geographic clustering if systematic sampling had been used.

Selection of Second-Stage Units

The units selected in the second stage of sampling were block groups (BG's) in areas where census blocks were delineated and census enumeration districts (ED's) in other areas. Each of the 87 distinct localities was completely covered by a nonoverlapping frame of BG's, ED's, or a combination of the two.

In order to reduce sampling error and ensure proportionate representation of women by race and income, the second-stage frame within each PSU was ordered by these variables prior to sampling. Because 1970 census income data were not available at the BG and ED level, NORC attached to each unit the income and racial characteristics of the next higher order census unit, which in most areas was the tract. In predominantly rural counties that were not tracted, the minor civil division (MCD) or census county division (CCD) was the next level unit. The National Opinion Research Center purchased from the National Planning Data Corporation of Ithaca, New York, population and housing data for ED's and BG's, and income and racial composition data for tracts, MCD's, and CCD's within each of the 87 distinct localities.

With these data the second-stage frame within each PSU was ordered in the following manner:

1. In SMSA's the ED's and BG's in tracts with less than 10 percent Negro households were placed before units in tracts with at least 10 percent Negro households. (For this purpose, the race of the household was defined as the race of the head.) Within each race group, the units were arranged in ascending order of median tract income. Within each tract the ED's and BG's were arranged in numerical order, with BG's preceding ED's if both types of unit were present.

2. In non-SMSA PSU's the above method of ordering was applied when at least 6 percent of the total population was Negro. In counties where Negroes constituted less than 6 percent of the population, the tracts, MCD's, and CCD's were either placed in ascending order by median income or, in sparsely populated

counties, arranged from northwest to northeast to southeast to southwest. Within each higher order unit, the BG's or ED's (usually ED's) were arranged in numerical order.

After the ED's and BG's were ordered, small units were linked with the unit immediately following to assure a minimum unit size of 100 DU's. The cumulative households in each PSU were then divided into 18 zones of equal size. Within each of these second-stage zones, the numbers between 1 and the zone size were grouped into ordered intervals representing BG's and ED's in the same way that the intervals in first-stage zones represented SMSA's and counties. A random hit number was selected for each zone to determine the ED or BG that would represent the zone. For SMSA's that appeared more than once in the principal sample, this procedure was carried out for each first-stage appearance. Thus the Chicago SMSA, with 4 hits in the sample, was represented by 4 X 18 or 72 second-stage units.

Of the 18 second-stage units associated with each PSU, a subsample of 12 was selected for the NSFG. The 18 second-stage units were put into 6 groups of 3 units apiece (group 1 consisted of ED's and BG's from zones 1, 2, and 3; group 2 included units from zones 4, 5, and 6; etc.). A random number was independently selected for each group to determine which second-stage units were to be eliminated from the NSFG sample. The second-stage sample now contained 1,212 second-stage units within the 101 PSU's.

At this point NORC and NCHS decided to supplement the sample by selecting additional second-stage units in areas where the population was largely Negro, because the design requirements of the NSFG dictated an oversampling of Negro females. Without these additional units later-stage sampling rates in areas with many Negroes would have been much larger than the rates in areas with few Negroes. The resulting large cluster sizes would have greatly increased the variances of survey estimates.

The first step in selecting the supplemental sample was to identify all second-stage units in tracts, MCD's, and CCD's with at least 10 percent Negro households (again, the race of a household was determined by the race of its head). The zones represented by these units were split into two half zones of equal size, say z. If the original hit number h for a zone fell into the first half zone, the number $h + z$ was used to select a tentative supplemental ED or BG. If the original hit number fell into the second half zone, $h - z$ was the supplemental hit number.

The selection procedure had to be modified for Washington County, Alabama, and Sumter County, Georgia. The 36 original ED's representing these rural, predominantly Negro PSU's encompassed virtually all of their population. Therefore, it was necessary to return to the first-stage non-SMSA zones represented by these PSU's and randomly select an additional PSU from each. The two supplemental PSU's were Hale County, Alabama, and Newton County, Georgia. Their BG's and ED's were ordered according to the procedure described earlier, and their cumulative DU's were divided into 12 zones of equal size. The second-stage zone method was used to select 12 tentative supplemental units from each PSU.

To determine which of the tentative supplemental units would be included in the sample, each one was paired with its corresponding original selection and the simple average percent of Negro households was computed for each pair. It was estimated by NORC that in order to produce the required sample of 3,600 Negro women with a sample of 32,000 DU's, the supplemental unit should be included when the average percent of Negro households was 43.4 or larger. Otherwise, only the original unit should be retained. The average exceeded 43.4 percent for 122 of the 1,212 zones. Therefore, the NSFG second stage consisted of $1,090 + 2(122) = 1,334$ ED's and BG's. For purposes of subsequent sampling, the units were divided into two Strata. Stratum I consisted of the 1,090 predominantly white units from the zones where the supplemental unit was not used. Stratum II consisted of the remaining 122 pairs of units.

Selection of Third-Stage Units

The purpose of the third stage of sampling was to select from each ED and BG a subarea called a "segment" containing approximately

100 DU's. When an entire second-stage unit contained only about 100 DU's, no third stage of sampling was required. When a third stage was necessary, however, the logistics of selecting a segment from a BG and an ED were quite different.

The BG is an urban geographic unit introduced during the 1970 census. It consists of a set of city blocks with the same highest order digit in the census block identification number. For the third stage of sampling, NORC subjectively split the BG into groups of blocks, each group containing approximately 100 DU's according to 1970 census counts. One group was randomly selected with probability proportionate to its size relative to the size of the total BG.

The ED is a census divisional unit generally used in nonmetropolitan areas. It is usually the lowest level unit for which decennial census information on households is available. In some instances, however, an area called an "ED" during the second stage was found to be covered by blocks. When that was the case, any necessary third-stage sampling was carried out in the manner described in the preceding paragraph. When blocks were not delineated and the ED had many more than 100 DU's, the ED was split into pseudoblocks which were bounded by roads and other easily recognizable landmarks. Dwelling units within the pseudoblocks were field counted in order to get rough measures of size, and a segment of approximately 100 DU's was selected with probability proportional to its size relative to the size of the entire ED according to the rough count.

The final third-stage NSFG sample consisted of 1,334 segments, one corresponding to each sample BG and ED. Therefore, there were 1,090 segments in Stratum I and 244 in Stratum II.

Selection of Fourth-Stage Dwelling Units

Members of the NORC field staff listed the DU's within each of the 1,334 segments. An independent, systematic sample of DU's was selected from the listing sheet for each segment. The individual fourth-stage sampling rates were chosen so that the sample of DU's was essentially self-weighting within Strata. That is, the overall probabilities of selecting all sample DU's in Stratum I were approximately equal, as were the probabilities associated with sample DU's from Stratum II.

In order for NORC to determine the fourth-stage sampling rate for each segment, they first had to determine P_{I} and P_{II}, the uniform probabilities of selection for the two Strata. To do this, preliminary census data were used to estimate the proportion of households in each Stratum with Negro heads. These proportions, along with the required number of screened DU's for Negroes and persons of other races, were the constraints used to calculate the required number of screened DU's for the two Strata (hereafter denoted by S_{I} and S_{II}). Next, D_T, the number of occupied DU's in the United States at the time of the survey, was projected from census housing totals for 1970 and 1972. The number of occupied DU's in the United States which were in ED's or BG's that met the definition of Stratum I was estimated by

$$D_I = \left(\frac{1{,}090}{1{,}212}\right) D_T$$

where 1,090/1,212 was the proportion of original second-stage sample units that fell into Stratum I. A corresponding estimate of occupied DU's for Stratum II units in the United States was

$$D_{\mathrm{II}} = \left(\frac{122}{1{,}212}\right) D_T.$$

Because 1970 census data indicated that 92 percent of all DU's in the United States were occupied, the selection probabilities for the two Strata were given by

$$P_{\mathrm{I}} = \frac{S_{\mathrm{I}} \cdot 0.92}{D_{\mathrm{I}}} = .000276219 \text{ and}$$

$$P_{\mathrm{II}} = \frac{S_{\mathrm{II}} \cdot 0.92}{D_{\mathrm{II}}} = .002207277$$

Once NORC had calculated P_{I} and P_{II}, the fourth-stage sampling rate for any segment could be obtained by dividing P_{I} or P_{II} by the product

of the known probabilities of selection at the first three stages. The fourth-stage rates yielded an average of 16.7 screened DU's per segment in Stratum I and 56.9 in Stratum II. It is obvious from the difference between these average cluster sizes that oversampling of ED's and BG's at the second stage could not by itself produce the required oversample of Negro women. Further oversampling of DU's was necessary within sample segments.

Because the field listing sheet was the frame for the fourth-stage sample, DU's that were missed during the field listing, or that came into existence between the time of field listing and interviewing, had zero probability of being sampled directly. In order to give these DU's the same probability of selection as listed DU's in their respective segments, NORC developed a set of rules based on the half-open interval procedure.[2] These rules enabled the interviewer to link each missed DU and each new DU with exactly one existing line of the segment listing sheet.

The set of lines for a segment was first divided into subsets that represented listings for individual blocks. In urban areas blocks were usually well-defined units, but in many rural segments "blocks" were of widely varying shapes and sizes. The only requirements for a rural block were clearly defined boundaries and a complete up-to-date listing of DU's by address or location. Each line within a block represented a structure or a subunit within a structure (such as an apartment or room). The listings were considered circular; that is, the last listing within a block was "followed" by the first.

When a sample line represented a complete structure, the interviewer was instructed to complete a screener for any DU's within that structure as well as any DU's between that structure and the structure on the next line. "Between" was defined in terms of address numbers when they were available, and in terms of location when they were not. If a sample line represented a subunit within a structure, the interviewer's instructions depended upon the position of the sample line relative to the lines representing the other subunits. Since the listings were ordered, each multiunit structure had a first- and last-listed subunit and any number (from 0 to n) of other subunits. If the sample line was a first-listed subunit, the interviewer was instructed to complete a screener for all additional DU's within the subunit and all DU's in the structure that were in unlisted subunits. If the sample line was a last-listed subunit, the interviewer was instructed to complete a screener for all additional DU's in that subunit and all DU's between the structure containing the subunit and the next-listed structure. If the sample line represented any other subunit, the interviewer was only responsible for additional DU's in the subunit.

In order to avoid large increases in sample size due to additional DU's, NORC set an arbitrary limit of four DU's per sample line. If more than three additional DU's were associated with the original sample listing, the interviewer was instructed to call headquarters, where a random subsample of exactly four DU's was chosen to receive screeners.

The procedures for listing additional DU's and subsampling excess DU's added 780 unlisted DU's to the original sample, of which 439 were in Stratum I and 341 were in Stratum II. The subsampled DU's were exceptions to the principle of equal probability of selection within Strata.

Additional exceptions to equal probability of selection were made for 10 segments in Stratum I which had grown rapidly from the time of the census DU count to the time of the NORC field listing. The field staff in these "fast-growth" segments returned their listing sheets to NORC's central office. The central office reduced the fourth-stage sampling rates to keep the number of sample DU's from exceeding 50, so that the interviewer's workload would not become overly burdensome.

Fifth-Stage Selection of Sample Persons

To avoid the high correlation of information from eligible women within the same DU, the NSFG design stipulated that no more than one eligible woman from any sample DU would be interviewed. During completion of the Household Screener, the NORC interviewer listed all members of the DU on the second page of the form and relisted the eligible females in order of age in item 13 on the third page (see appendix III). Item 13 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 female she referred to the sampling table on the first page of the questionnaire to determine the person she was supposed to interview. The sampling table consists of five numbers that designate which person to interview when the number of eligible females is two, three, four, five, or six (see table D). National Opinion Research Center personnel filled in the table on every Household Screener by randomly ordering the 720 possible sets of numbers and systematically assigning them to screeners. This method of assigning interviews gave each eligible woman in a given DU the same probability of being selected.

The oversampling in Stratum II at the second and fourth stages of selection led to the desired oversampling of Negro females at the fifth stage. However, since 35 percent of the DU's in Stratum II were expected to be of other races, oversampling would also have produced an unnecessarily large sample of females of other races. NORC avoided that costly problem by subsampling DU's of other races at a rate of 1 out of 7.991. This rate gave the subsampled DU's the same probability of selection as all DU's in Stratum I. Subsampling was accomplished by systematically printing "interview regardless of race" on 1 of every 7.991 screeners used in Stratum II and "interview if Negro only" on the rest. For this purpose, the race of a DU was defined as the race of the person who provided the information for the Household Screener because it seemed likely that the screener respondent would often be an eligible respondent for the survey and because it might have been difficult to determine the race of all household members in some instances.

Table D. 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	2

CHARACTERISTICS OF THE SAMPLE

The first four stages of the design resulted in the identification of 32,818 sample DU's. During the screening process, the interviewers discovered that 3,820 of these either were vacant or did not meet the definition of a DU. Complete Household Screeners were obtained for 26,028 of the remaining 28,998 occupied DU's, for a screener completion rate of 89.8 percent (table E). After 2,674 DU's were removed from the sample by the subsampling procedure in Stratum II, the final sample of 23,354 DU's yielded a fifth-stage sample of 10,879 women, of which 4,362 were Negro and 6,517 were of other races (table F). Complete interviews were obtained for 9,817 women, of which 3,868 were Negro and 5,949 were of other races, for an interview completion rate of 90.2 percent (88.7 percent for Negro women and 91.3 percent for women of other races). Combined screener and interview response rates cannot be computed by race because the race of the majority of nonresponding DU's was unknown. Combined screener and interview response by Stratum is shown in table G.

Seventeen women (10 Negro, 7 of other races) were eliminated from the sample after they were interviewed because it was discovered that they had passed their 45th birthday before the date of interview. In addition, data for three sample women who were less than 15 years of age (two Negro, one of another race) were excluded from all tabulations so that analysis could be conducted for eligible women 15-44 years of age. Therefore, NSFG estimates are based upon data from 9,797 women, of which 3,856 are Negro and 5,941 are of other races.

While the number of interviewed Negro women exceeded the desired sample size of 3,600, the number of interviewed women of other races fell substantially short of the target of 6,400. Most of the additional Negro sample was due to the unexpectedly large number of

Table E. Actual and expected number of sample DU's, number of completed screeners, and screener completion rates, by race and Stratum

Race and Stratum	Total DU's in the sample (1)	Vacant or not a DU (2)	Number of legitimate, occupied DU's (3)	Expected number of legitimate, occupied DU's from presurvey estimates (4)	Number of DU's for which a complete screener was obtained (5)	Screener completion rate[1] (percent) (6)
All races						
Both Strata	32,818	3,820	[2]28,998	29,295	[2]26,028	89.8
Stratum I	18,593	2,038	[2]16,555	16,576	[2]15,134	91.4
Stratum II	14,225	1,782	[2]12,443	12,719	[2]10,894	87.6
Negro						
Both Strata	...	...	9,005	8,410	8,546	[3]...
Stratum I	...	...	773	656	725	[3]...
Stratum II	...	...	8,232	7,754	7,821	[3]...
Other races						
Both Strata	...	...	18,438	20,885	17,328	[3]...
Stratum I	...	...	15,059	15,920	14,278	[3]...
Stratum II	...	...	3,379	4,965	3,050	[3]...

[1][(5) ÷ (3)] X 100.
[2]Includes race unknown.
[3]Appropriate screener completion rates by race cannot be derived because race information was not available for all legitimate, occupied DU's.

Table F. Number of sample women, actual and expected number of completed interviews, and interview completion rates, by race and Stratum

Race and Stratum	Number of sample women (1)	Number of completed interviews (2)	Interview completion rate[1] (percent) (3)	Approximate number of completed interviews expected from presurvey estimates (4)
All races				
Both Strata	10,879	9,817	90.2	10,054
Stratum I	6,758	6,164	91.2	[2]6,451
Stratum II	4,121	3,653	88.6	3,600
Negro				
Both Strata	4,362	3,868	88.7	3,600
Stratum I	394	341	86.5	281
Stratum II	3,968	3,527	88.9	3,319
Other races				
Both Strata	6,517	5,949	91.3	[2]6,454
Stratum I	6,364	5,823	91.5	6,173
Stratum II	153	126	82.4	281

[1][(2) ÷ (1)] X 100.
[2]The target sample size for women of races other than Negro was 6,400. However, the assigned probabilities of selection for DU's of race other than Negro in Stratum I and Stratum II yielded a slightly larger expected sample size than the original target.

Table G. Screener and interview completion rates and combined response rates, by Stratum

Stratum	Screener completion rate (1)	Interview completion rate (2)	Combined response rate[1] (3)
	Percent		
Both Strata	89.8	90.2	81.0
Stratum I	91.4	91.2	83.4
Stratum II	87.6	88.6	77.6

[1] [(1) X (2)] ÷ 100.

Negro DU's identified in Stratum II (table E). This oversample was enough to overcome the higher-than-expected vacancy and nonresponse rates. The number of DU's of race other than Negro in Stratum II was correspondingly much lower than expected. However, because of the subsampling procedure the number of interviewed women of races other than Negro was only 155 less than expected. In Stratum I the number of identified DU's was also somewhat less than the expected value, and the vacancy and interview nonresponse rates were higher than expected. These problems caused the remaining sampling deficit for women of races other than Negro, but had little effect on the size of the Negro sample because less than 5 percent of the DU's in Stratum I were classified as Negro.

Table H shows that the precision of survey estimates was not adversely affected because the target number of women of races other than Negro was not interviewed. Standard errors for the estimated proportions in table A were calculated from NSFG data using the balanced half-sample replication technique described later in this report. For all population subgroups except Negro women of parity 3 or more, these standard error estimates are substantially lower than the corresponding presurvey error tolerance. The estimate for higher parity Negro women is slightly larger than its tolerance, but the difference is unimportant, since the variance estimates themselves are subject to variance.

ESTIMATION

Weighting Procedure

Since the NSFG is designed to produce unbiased estimates for the entire population of

Table H. Comparison of standard error tolerances with the corresponding values obtained from the National Survey of Family Growth (NSFG)

Population subgroup	Proportion of women in the race group who are in the subgroup		Estimated proportion within the subgroup	Standard error tolerance from table A	Estimated standard error based on NSFG data
	From table A	NSFG estimate			
Negro					
Education:					
Less than high school	.48	.48	.42	.0164	.0150
High school and more	.52	.52	.25	.0139	.0121
Parity:					
0-2 children	.50	.58	.19	.0128	.0112
3 children or more	.50	.42	.51	.0163	.0170
Other races					
Education:					
Less than high school	.30	.27	.23	.0133	.0106
High school and more	.70	.73	.15	.0074	.0053
Parity:					
0-2 children	.59	.64	.10	.0068	.0054
3 children or more	.41	.36	.33	.0127	.0099

eligible women in the 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, one or more at each stage of sampling. Three types of adjustments are involved.

Inflation by the reciprocal of the probabilities of selection.—The weight for each woman within a sample PSU is the product of the reciprocals of the probabilities of selecting (1) the ED or BG, (2) the segment, (3) the DU, and (4) the eligible sample person. Because of the possibility (sometimes certainty) that certain SMSA's and counties could be selected as sample PSU's more than once, the first-stage weight is the reciprocal of the expected relative frequency of occurrence of the PSU. The first-stage weight is explained in more detail in the discussion of the estimating equation.

Nonresponse adjustment.—Each sample weight is adjusted for nonresponse to the Household Screener (screener nonresponse) and nonresponse of sample women to the detailed NSFG questionnaire (interview nonresponse). These adjustments are necessary because the phenomenon of nonresponse introduces bias into any probability sample. The respondents to a survey may have a much different distribution of demographic or health characteristics than the nonrespondents. Even if the distribution of demographic and health characteristics is about the same for respondents and nonrespondents, the two groups are by definition different because the respondents participated in the survey while the nonrespondents did not. Nonresponse adjustments minimize the impact of nonresponse bias on final estimates by imputing to nonresponding DU's and women the characteristics of "similar" respondents. Similar respondents were judged to be DU's in the same PSU and Stratum, and women in the same age-race class and PSU. Screener response was 89.8 percent for the entire survey (91.4 percent in Stratum I and 87.6 percent in Stratum II) and ranged from 98.9 percent to 62.5 percent in individual PSU's. Interview response was 90.2 percent for the survey (91.2 percent in Stratum I and 88.6 percent for Stratum II) and ranged from 100 percent to 54.5 percent in individual PSU's.

Table J. Poststratification adjustment factors (ratio of September 1973 population control totals based on Current Population Survey data to National Survey of Family Growth weighted estimates), by race and age

Age	Negro	Other races
14-19 years	0.743	1.036
20-24 years	0.927	1.088
25-29 years	0.975	1.092
30-34 years	0.960	0.991
35-39 years	0.876	0.951
40-44 years	1.082	1.097

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 J make NSFG estimates of ever-married women in each age-race class equal to independent control totals for September of 1973 (the approximate midpoint of data collection). The control totals are based on data from the U.S. Bureau of the Census' Current Population Survey (CPS). No poststratification adjustment is applied to the weights for single mothers because reliable control totals for this population 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. The method used to compute the CPS control totals is discussed in appendix II as part of the evaluation of alternative estimators.

NCHS decided to use a poststratified estimator instead of a simple inflation estimator after conducting research to compare the precision of the two estimators. The methodology and results of the comparison are given in appendix II.

Estimating Equation

The estimate of an aggregate parameter Y is given by[3]

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

where Y'_1 is the estimate for ever-married women and Y'_2 is the estimate for single mothers.

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

is a poststratified estimator. The nonresponse-adjusted estimates Y' for the 12 age-race classes are multiplied by the poststratification adjustment factor $\hat{X}_{\alpha 1}/X'_{\alpha 1}$. $X'_{\alpha 1}$ represents the NSFG estimate of the number of ever-married women in age-race class α and $\hat{X}_{\alpha 1}$ is an estimate of the same population group based upon the CPS.

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

is simply the sum of the nonresponse-adjusted estimates for single mothers in the 12 age-race classes.

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

$$Y'_{\alpha 1} = A_\alpha \sum_{g=1}^{103} W_{1g} \frac{n_{g\alpha}}{\dot{n}_{g\alpha}} \sum_{h=1}^{2} \frac{M_{gh}}{\dot{M}_{gh}} \sum_{i=1}^{S_{gh}} W_{2ghi} W_{3ghij} \sum_{k=1}^{H_{ghij}} \frac{W_{4ghijk} T_{4ghijk}}{F_{4ghijk} R_{4ghijk}} \cdot W_{5ghijk} \cdot \delta_{\alpha ghijk} \cdot Y_{\alpha ghijk}.$$

$Y'_{\alpha 2}$, the corresponding estimator for single mothers, is exactly the same, except that $\delta_{\alpha ghijk}$ is replaced by $1 - \delta_{\alpha ghijk}$.

$Y_{\alpha ghijk}$ = the observed value of characteristic Y for the sample woman selected from DU k, segment j, ED or BG i, Stratum h, PSU g, and age-race class α.

$\delta_{\alpha ghijk}$ = 1 if the sample woman whose observed value is $Y_{\alpha ghijk}$ has ever been married; 0 otherwise.

W_{5ghijk} = the fifth-stage weight applied to the sample woman represented by $Y_{\alpha ghijk}$. The weight is equal to the number of eligible women in her DU.

W_{4ghijk} = the reciprocal of the original fourth-stage sampling rate within segment j.

T_{4ghijk} = 7.991 if segment j is in Stratum II and the race of DU k is other than Negro; otherwise $T_{4ghijk} = 1$.

F_{4ghijk} = the subsampling rate that was applied if segment j was one of the 10 fast-growth segments in Stratum I; otherwise, $F_{4ghijk} = 1$.

R_{4ghijk} = the subsampling rate that was applied if DU k was associated with excess missed DU's; otherwise, $R_{4ghijk} = 1$.

H_{ghij} = the number of sample DU's in segment j.

W_{3ghij} = the reciprocal of the probability of selecting segment j, the segment selected from ED or BG i.

W_{2ghi} = the reciprocal of the probability of selecting ED or BG i from PSU g and Stratum h.

S_{gh} = the number of sample ED's and BG's in PSU g and Stratum h.

$\frac{M_{gh}}{\dot{M}_{gh}}$ = an adjustment for screener nonresponse (DU's for which it was impossible to determine whether or not there were any eligible respondents). The value M_{gh} represents the total number of DU's from PSU g and Stratum h, and $\dot{M}_{gh}$ is the number of DU's that were classified as either including or not including eligible respondents.

$\frac{n_{g\alpha}}{\dot{n}_{g\alpha}}$ = a partial adjustment for person nonresponse, where $n_{g\alpha}$ is the number of sample persons in PSU g, age-race class α, and $\dot{n}_{g\alpha}$ is the number of persons who respond. However, NCHS decided that nonresponse greater than 50 percent within a PSU should be adjusted at the class level rather than the PSU level. Therefore, when $n_{g\alpha}$ is greater than twice $\dot{n}_{g\alpha}$ the nonresponse adjust-

Table K. Example[1] of how to determine A_α from PSU response data

PSU number	Number of sample persons in class α	Number of respondents in class α	Nonresponse adjustment factor	Excess nonrespondents
Totals for class α.....	43	31	...	3
1.....	10	10	1	0
2.....	12	10	1.2	0
3.....	13	6	2	1
4.....	6	5	1.2	0
5.....	2	0	2	2

[1]For this example, $A_\alpha = \frac{43}{43-3} = \frac{43}{40}$.

ment is defined to be 2, and the excess nonresponse is adjusted in factor A_α described below.

W_{1g} = 2,000,000/P_g, where P_g is the 1970 census population of PSU g. This first-stage weight is the reciprocal of the expected frequency of occurrence of PSU g in the NSFG sample. For PSU's that were completely contained in a single zone at the time of selection, W_{1g} is the reciprocal of the probability of selection. However, W_{1g} deviates from the reciprocal of the probability of selection for PSU's that overlap zones. For SMSA's with populations greater than 2 million W_{1g} is less than 1, whereas the reciprocal of the probability of selection for any PSU is always greater than or equal to 1.

A_α = an excess nonresponse adjustment that compensates at the α-class level for nonresponse greater than 50 percent in individual PSU's. Table K gives an example of how to determine A_α.

VARIANCE ESTIMATION

Background

The balanced half-sample replication technique described in detail in other NCHS reports[4,5] is used to estimate NSFG variances. An empirical study by Bean[6] gives evidence that the half-sample technique produces highly reliable, essentially unbiased variance estimates.

There are three important practical reasons why half-sample replication is being used:

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 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.[2]

Let the parameter of interest be denoted by Y, for which an estimate Y' has been obtained from the complete sample. Y' is a linear combination of the sample observations in fully rigorous developments, although 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.[4,5,7,8]

A half-sample replicate is defined as a collection of L PSU's 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 $K > L$, 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 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[9] and Gurney.[10] 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 are 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 satisfy the McCarthy criteria is called a "balanced set," and the procedure is referred to as "balanced half-sample replication."

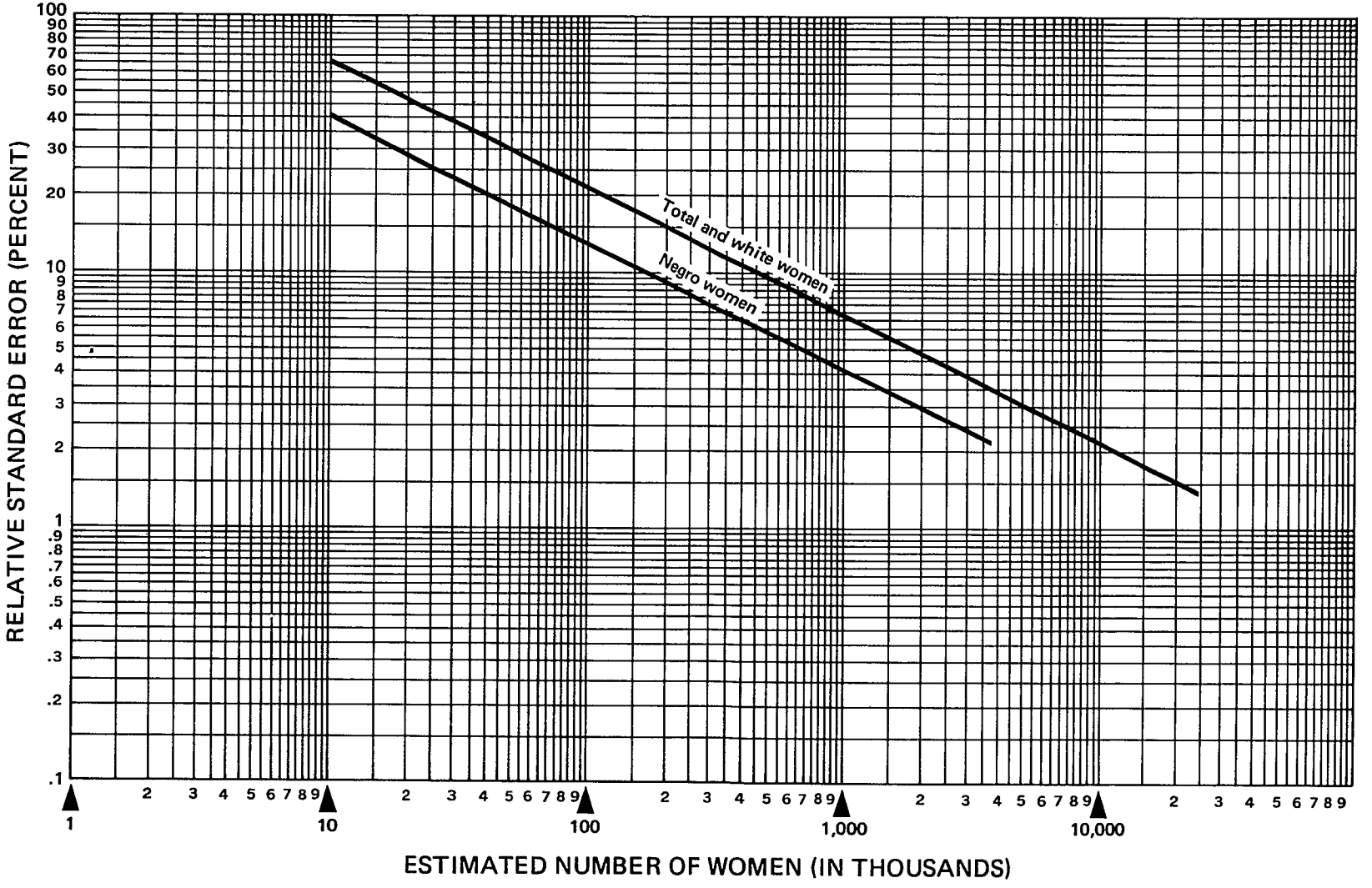
Table L. Example of a half-sample replication pattern

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

Application to the NSFG

In order to apply the balanced half-sample replication technique, NCHS grouped the 103 NSFG PSU's into 48 strata. Seven of the strata were self-representing; that is, they consisted of SMSA's that came into the sample with certainty (except for Boston, which was lumped with 2 of the 5 New York City PSU's). Within each of these strata two pseudo-PSU's were created by listing the PSU's in numerical order and listing the segments in numerical order within each PSU. The first segment and every second segment thereafter was assigned to the first pseudo-PSU, and the remaining segments were assigned to the second pseudo-PSU. Within the other 41 strata, each of the 2 replication PSU's consisted of 1 or more NSFG PSU's (but never more than one locality). The PSU's in each stratum belonged to the same geographic region and

Figure 2. Relative standard errors for aggregates of women, by race



Example of use of chart: An aggregate of 2 million women (on the scale at the bottom of the chart) of all races has a relative standard error of 4.8 percent, or a standard error of 96,000 (4.8 percent of 2 million).

were generally about the same size. The value 1 was assigned to one of the replication PSU's in each stratum, and the value 2 was assigned to the other. Forty-eight half samples were then formed from the entries in successive columns of an orthogonal 48 X 48 matrix of 1's and 2's adapted from Plackett and Burman.[9]

In order to estimate the variance of an NSFG aggregate statistic Y', 48 half-sample analogs of Y' were computed. The formula for the kth half-sample estimate is

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

where

$$Y'_{1k} = \sum_{\alpha=1}^{12} \frac{Y'_{\alpha 1k}}{X'_{\alpha 1k}} \hat{X}_{\alpha 1}$$

and

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

$Y'_{\alpha 1k}$ and $X'_{\alpha 1k}$ were computed in exactly the same way as $Y'_{\alpha 1}$ and $X'_{\alpha 1}$. No additional weighting was necessary because of the poststratification process. On the other hand, the $Y'_{\alpha 2k}$ were computed after multiplying the weight for each single mother by 2 to compensate for the half-sample procedure. The variance of Y' was then estimated by

$$S^2_{Y'} = \frac{1}{48} \sum_{k=1}^{48} (Y'_k - Y')^2$$

Types of aggregate statistics produced from the NSFG include number of currently married women, number of ever-married women, number of pregnancies for ever-married women, and total number of children born to ever-married women. Half-sample variances were not computed for all aggregate statistics because the time and money needed to do so would have been prohibitive. In addition, data reports would be much more cumbersome if a variance estimate were published for each statistic. To avoid these problems, 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 fitted to relative standard error (RSE) estimates for number of currently married women and number of pregnancies, according to the model

$$\text{RSE}\,(Y') = \sqrt{\frac{S^2_{Y'}}{(Y')^2}} = \sqrt{A + \frac{B}{Y'}}$$

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.[11]

Figure 2 shows relative standard error curves for estimates of women by race. The estimates of A and B for the curves are shown in table M. Although the curves for women were fitted to RSE's for estimates of currently married women only, RSE's for estimates of ever-married women fall close to the curves. Therefore, it is appropriate to obtain predicted RSE's for estimates of ever-married women from the aggregate curves for women.

Separate curves were needed for Negro women and women of all other race classifications (women of all races, women of race other than Negro, and white women). Because Negro women were oversampled, an estimate of a given number of Negro women has a smaller RSE than an estimate of the same number of women of all races. The curves in figure 2 clearly show this relationship. For example, an estimate of 200,000 Negro women has an RSE of 9 percent, while an estimate of 200,000 women of all races has an RSE of 15 percent.

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

Curve	A	B
Total and white women	0.000017613	4,493.7916
Negro women	0.000040219	1,600.4393

Variances of aggregate statistics were used to derive variances of percents, which are ratios of two aggregates with the numerator being a subclass of the denominator. Percent estimates usually show the proportion of a population that has a particular characteristic of interest. The RSE of the percent estimate

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

is given by the expression[11]

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

where B is the least squares estimate from the error curve for Y' and Z'.

The RSE of P' is dependent upon the values of both P' and Z'. In order to account for the variation in error due to P' and Z', a set of percent RSE curves was derived from each aggregate RSE curve. Each curve in the set yields RSE's for percent estimates with a fixed denominator.

Figure 3 shows the set of curves for percent of total and white women, along with an example of how to use the error chart. Each curve satisfies the equation

$$\text{RSE}\,(P') = \sqrt{\frac{4493.7916\,(100 - P')}{P'Z'}}$$

where P' is the estimated percent and Z' is the denominator of P'. Linear interpolation yields an acceptable estimate for the RSE of a percent whose denominator is not one of the values shown in figure 3.

In addition to percents, the NSFG provides other types of ratio estimates where the numerator is not a subclass of the denominator, such as the mean number of expected births per woman, the mean number of expected additional births per woman, and the probability that a woman gives birth within a certain number of months following her last previous birth. Variances for these types of estimates could not be derived from variances of aggregate statistics, so they were computed directly by the half-sample technique. The variance of the ratio estimate $R' = Y'/W'$ is given by

$$S_{R'}^2 = \frac{1}{48}\sum_{k=1}^{48}(R_k' - R')^2$$

where R_k' is the kth half-sample analog of R'.

As was the case for aggregate statistics, approximate error curves were fitted to selected variance estimates. The model for these more complicated ratio estimates was

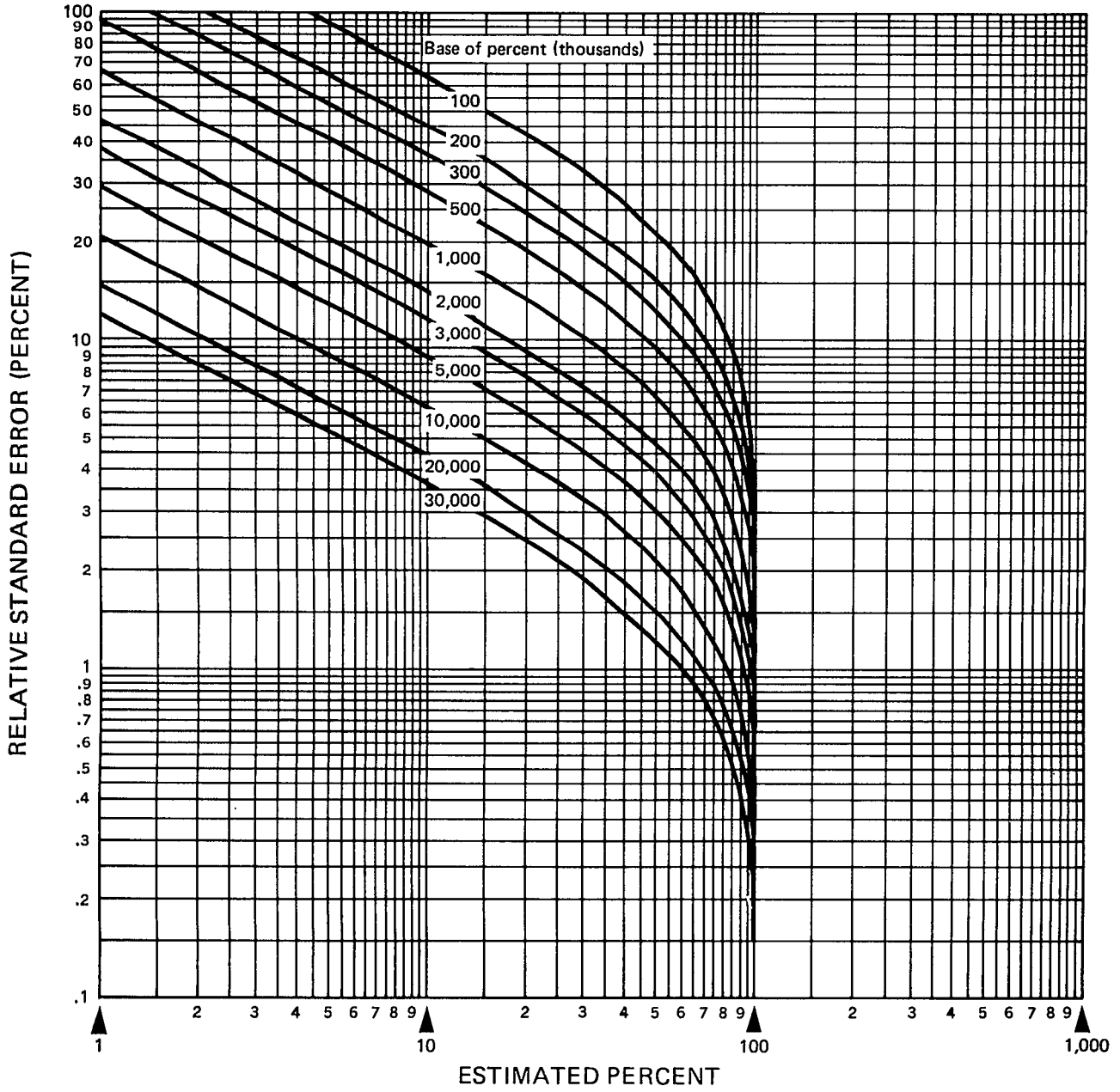
$$\begin{aligned} \text{RSE}\,(R') &= \sqrt{A + \frac{B}{Y'} + \frac{C}{W'}} \\ &= \sqrt{A + \frac{B}{R'W'} + \frac{C}{W'}} \end{aligned}$$

where A, B, and C were the parameters to be estimated. As with percentage estimates, the RSE varies with the size of the estimate R' and its denominator W'. The method of estimating A, B, and C is the same general method used for aggregate estimates.[11] The RSE's for each type of ratio statistic are represented by a set of curves, as was the case for percent estimates.

Figure 4 shows the set of curves for mean number of births expected by women of all races, along with an example of how to use the error chart. The three curves in the chart satisfy the equation

$$\text{RSE}(R') = \sqrt{0.00002488 - 649.05239/R'W' + 1{,}733.8522/W'}$$

Figure 3. Relative standard errors for percent of total and white women (base of percent shown in curve in thousands)



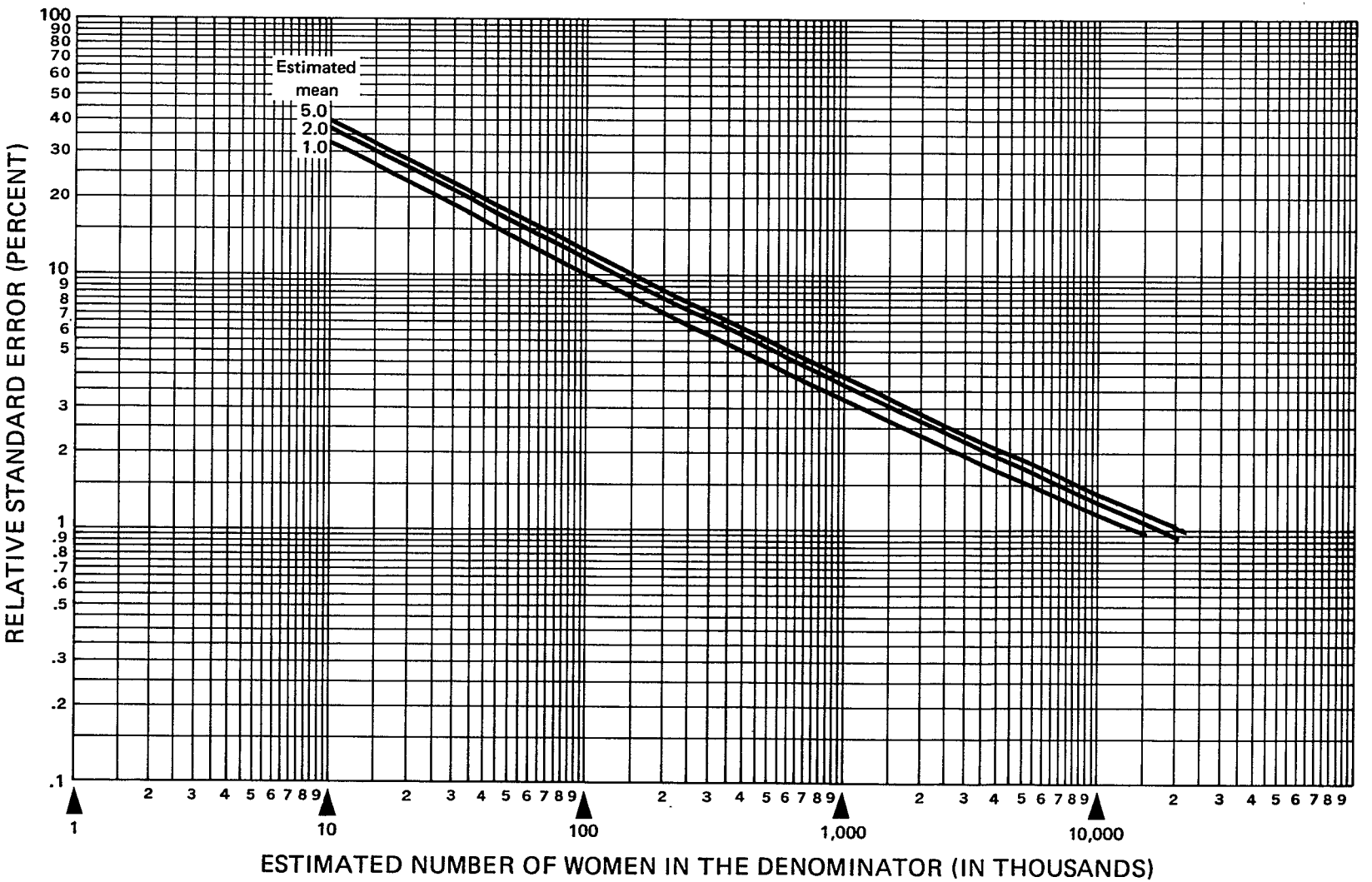
Example of use of chart: An estimate of 10 percent (from the scale at the bottom of the chart) of a population subgroup of 1 million women (fifth curve from the top) has a relative standard error of 20.0 percent, or a standard error of 2.0 percent (20.0 percent of 10 percent).

where R' is the estimated mean and W' is the estimated number of women (the denominator of R'). Linear interpolation yields an acceptable estimate for the RSE of R' when R' is not one of the values shown in figure 4.

Figures 2-4 are examples of relative standard error curves for the various kinds of statistics produced from Cycle I of the NSFG. Each report that discusses findings from Cycle I will include similar charts for all types of statistics presented therein.

An estimate of the standard error of the dif-

Figure 4. Relative standard errors for mean numbers of births expected by women of all races and white women



Example of use of chart: An estimate of 2.0 births per woman (middle line on the chart) for a population subgroup of 900,000 women (read from the scale at the bottom of the chart) has a relative standard error of 4.0 percent, or a standard error of 0.08 (4.0 percent of 2.0).

ference between any two aggregates, percents, or other ratio statistics is given by

$$S_{(Y'_1 - Y'_2)} = \sqrt{S^2_{Y'_1} + S^2_{Y'_2}}$$

$$= \sqrt{(Y'_1)^2 \mathrm{RSE}^2(Y'_1) + (Y'_2)^2 \mathrm{RSE}^2(Y'_2)}$$

This expression provides a good estimate of the standard error for uncorrelated statistics, but can only be considered a rough approximation otherwise. Because NSFG estimates are based upon a large sample of women, the distributions of Y'_1 and Y'_2 (and, therefore, $Y'_1 - Y'_2$) are approximately normal. Frankel[12] 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 strata in the replication design (48 for the NSFG) can reasonably be used as the number of degrees of freedom for the t statistic, although the exact value for the degrees of freedom is unknown. Therefore, individual two-tailed significance tests of differences between NSFG statistics can be performed with an approximate significance level of α by computing t and comparing it to the two-tailed $1 - \alpha$ critical value for the t distribution with 48 degrees of freedom.

Example: Suppose 500,000 currently married Negro women expect an average of 2.46 births per woman and 5 million currently married women of other races expect an average of 2.08 births per woman. To test this race difference at the $\alpha = .05$ level of significance, compute

$$t = \frac{2.46 - 2.08}{\sqrt{(2.46)^2 \mathrm{RSE}^2(2.46) + (2.08)^2 \mathrm{RSE}^2(2.08)}}$$

From figure 4, RSE(2.46) $\approx$ .055 and RSE(2.08) $\approx$ .018 so that

$$t = \frac{2.46 - 2.08}{\sqrt{(2.46)^2 (0.55)^2 + (2.08)^2 (.018)^2}}$$

$$= 2.71$$

The two-tailed .95 critical value $(1 - \alpha)$ for a t statistic with 48 degrees of freedom is 2.01. Therefore, the difference is significant at the .05 level.

REFERENCES

[1] King, B. F., and Frankel, M. R.: Description of the Sample Design for the National Family Growth Survey. Unpublished manuscript.

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

[3] Brock, D. B.: Estimation and Variance Specifications for the National Survey of Family Growth. Unpublished manuscript.

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

[5] National Center for Health Statistics: Pseudoreplication: Further evaluation and application of the balanced half-sample technique. *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. *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] Simmons, W. R., and Baird, J.: Use of pseudoreplication in the NCHS Health Examination Survey. *Proceedings of the Social Statistics Section of ASA.* American Statistical Association, 1968.

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

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

[10] Gurney, M.: The Variance of the Replication Method for Estimating Variances for the CPS Sample Design. Dittoed memorandum, U.S. Bureau of the Census, 1962.

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

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

[13] French, D. K., and Bryant, E. E.: Sample-Dependent Estimation for the National Survey of Family Growth. *Proceedings of the Social Statistics Section of ASA.* American Statistical Association, 1976.

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 popu-

lation of at least 50,000. In addition, other contiguous 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.

APPENDIX II

EVALUATION OF ALTERNATIVE ESTIMATORS

An aggregate parameter Y for all eligible women less than 45 years of age in the United States can be written as $Y = Y_1 + Y_2$, where Y_1 is the total for all ever-married women and Y_2 is the total for all single mothers with one or more of their children currently living in the household. Alternative estimators of Y considered for the NSFG were

$$Y'_{NP} = Y'_1 + Y'_2 \qquad (1)$$

and

$$Y'_P = Y''_1 + Y'_2 \qquad (2)$$

where Y'_1 and Y'_2 are simple inflation estimators and

$$Y''_1 = \sum_{\alpha=1}^{12} Y'_{\alpha 1} \cdot \frac{\hat{X}_{\alpha 1}}{X'_{\alpha 1}} \qquad (3)$$

is a poststratified estimator. The $X'_{\alpha 1}$ represent NSFG estimates of the number of ever-married women in each of 12 age-race classes and the $\hat{X}_{\alpha 1}$ are corresponding estimates for September 1973 (the approximate midpoint of NSFG interviewing) based upon the U.S. Bureau of the Census' Current Population Survey (CPS).

At the time the two estimators were proposed, it was decided that the choice of Y'_P or Y'_{NP} as the official estimator for the NSFG would be based on information from the sample data. Values of certain statistics would be calculated using both estimators, and variances of the two competing estimates would be computed using identical half-sample replication procedures. If the variance of Y'_P was the lower of the two variances for many of the experimental statistics, and if Y'_P did not result in higher variances for more than a few statistics, then it would be chosen; otherwise, Y'_{NP} would be used.

Simple inflation and poststratified estimates of the number of currently married U.S. women and the relative standard errors (RSE's) of both estimates were computed for more than 200 subdomains of the population. The poststratified estimator had a smaller RSE than the inflation estimator for more than 80 percent of the domains. The improvement in precision was 20 percent or better for almost two-fifths of the domains. A more complete discussion of the design and results of the study can be found elsewhere.[13]

Because the RSE of Y'_P was much smaller than the RSE of Y'_{NP} for many aggregate estimates, and was no more than 10 percent larger than the RSE of Y'_{NP} for any parameter in the study, the decision was made to use Y'_P as the official NSFG estimator.

NOTE: The list of references follows the text.

APPENDIX III
HOUSEHOLD SCREENER

NORC-4604
July, 1973

· NATIONAL OPINION RESEARCH CENTER
University of Chicago

OMB No. 68 - S72170
Expires: April 30, 1974

BEGIN DECK 21

Collected for the National Center for Health Statistics

HOUSEHOLD SCREENER

OFFICE USE ONLY	
Final Status ☐	10
Tel. Re-Check:	
Yes 1	11
No 2	

NOTICE: All information which would permit identification of any individual will be held in strict confidence, will be used only by persons engaged in and for the purpose of the survey, and will not be disclosed or released to others for any purpose, as in accordance with Section 305(a) of the Public Health Service Act, Section 1.103(a) of the Public Health Service Regulations [42 CFR 1.103(a)] and under Public Health Delegation of Authority Number 31.

ASSIGNMENT BOX

SAMPLING TABLE

IF NUMBER OF ELIGIBLE FEMALES LISTED IN SUMMARY BOX IS:	THEN INTERVIEW PERSON LISTED ON SUMMARY BOX LINE
Two	
Three	
Four	
Five	
Six or more	

INTERVIEWER: Is this address in a rural area or in some other kind of area? 12

Rural . . . 1 Other . . . 2

INTRODUCTION

Hello, I'm _____ from the National Opinion Research Center. (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.

RECORD OF ALL CONTACTS AT HOUSEHOLD

Day of Week	Date	Time	Type		Results	CODE ONE		Interv'r initials
			Per.	Tel.		Scr.	Quex	
		AM PM	1	2		1	2	
		AM PM	1	2		1	2	
		AM PM	1	2		1	2	
		AM PM	1	2		1	2	
		AM PM	1	2		1	2	
		AM PM	1	2		1	2	
		AM PM	1	2		1	2	
		AM PM	1	2		1	2	
		AM PM	1	2		1	2	
		AM PM	1	2		1	2	

TIME BEGAN: _____ AM PM

HOUSEHOLD ENUMERATION

BEGIN DECK 22

1. To start, how many people live in this household? ☐ NUMBER 10 11

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

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.)

4. Are there other people living here who are not related to (HEAD)?

☐ Yes ☐ No

[redacted]

5. I have listed (READ NAMES IN ORDER). Is there anyone else staying here now, such as friends, relatives or roomers?

☐ Yes ☐ No

[redacted]

AFTER LISTING HOUSEHOLD, ASK QS. 6-9 FOR EACH PERSON AS APPROPRIATE	6. What is (PERSON)'S relationship to (HEAD OF HOUSEHOLD)?	7. CODE SEX (ASK IF NOT OBVIOUS)		8. How old was (HEAD/PERSON) on (his/her) last birthday?	9. IF 13 YEARS OR OLDER, ASK: Is (PERSON) now married, widowed, divorced or annulled, separated or has (he/she) never been married? IF NEVER MARRIED BUT REPORTED AS LIVING TOGETHER, CODE "INFORMAL." IF NEVER MARRIED AND NOT LIVING TOGETHER BUT HAS OWN CHILDREN, CODE "WITH OWN CHILDREN."					
First Name Last Name		M	F		Marr.	In-formal	Wid.	Div. Ann.	Separated	Single with own children	Never Marr.
01	12 HEAD	13 1	2	14 15	1	2	3	4	5	6	7	16
02	17	18 1	2	19 20	1	2	3	4	5	6	7	21
03	22	23 1	2	24 25	1	2	3	4	5	6	7	26
04	27	28 1	2	29 30	1	2	3	4	5	6	7	31
05	32	33 1	2	34 35	1	2	3	4	5	6	7	36
06	37	38 1	2	39 40	1	2	3	4	5	6	7	41
07	42	43 1	2	44 45	1	2	3	4	5	6	7	46
08	47	48 1	2	49 50	1	2	3	4	5	6	7	51
09	52	53 1	2	54 55	1	2	3	4	5	6	7	56
10	57	58 1	2	59 60	1	2	3	4	5	6	7	61

IF MORE THAN 10 PEOPLE IN HOUSEHOLD, GO TO CONTINUATION BOOKLET, PAGE 2.

10. Is there anyone now away from home who usually lives here?

☐ Yes ☐ No

[redacted]

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

☐ Yes ☐ No

[redacted]

12. Are any of the persons in this household now on full-time active duty with the Armed Forces of the United States?

☐ Yes ☐ No

[redacted]

13. SELECTION OF RESPONDENT

(CHECK ONE BOX)

1 ☐ NO ELIGIBLE WOMEN UNDER 45 YRS. OF AGE, NO INTERVIEW REQUIRED; SKIP TO Q. 14. 13

2 ☐ ONE ELIGIBLE WOMAN UNDER 45 YRS. OF AGE, CURRENTLY MARRIED OR INFORMALLY MARRIED; CIRCLE R'S PERSON NUMBER ON P. 2; SKIP TO Q. 14; THEN USE CURRENTLY MARRIED QUESTIONNAIRE

3 ☐ ONE ELIGIBLE WOMAN UNDER 45 YRS. OF AGE, CURRENTLY WIDOWED, SEPARATED, DIVORCED OR ANNULLED, OR SINGLE W/OWN CHILDREN; CIRCLE R'S PERSON NUMBER ON P. 2; SKIP TO Q. 14; THEN USE POST-MARRIED QUESTIONNAIRE

☐ MORE THAN ONE ELIGIBLE WOMEN UNDER 45 YRS. OF AGE, FOLLOW STEPS 1 - 4 TO SELECT CORRECT RESPONDENT

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

SUMMARY BOX		
Line #	Name	Age
1		
2		
3		
4		
5		
6		

14

Step 2: Use Sampling Table on page 1 to determine which eligible female to interview.

Step 3: Circle R's Line # in Summary Box, and write selected R's name here:

_____ (Respondent's Name) 15 16

Step 4: (CHECK ONE BOX)

4 ☐ R is currently: married, or informally married; ask Q. 14 and use CURRENTLY-MARRIED QUEX.

5 ☐ R is currently: widowed, separated, divorced or annulled, single w/own children; ask Q. 14 and use POST-MARRIED QUEX.

14. ASK EVERYONE: May I have your telephone number (in case my office wants to verify this interview)? 17

Telephone no.: Area Code: ___/_____

No phone 2
Refused 3

IF PHONE NO. GIVEN, CODE LOCATION OF PHONE:

In Household 4
In home of neighbor 5
_____ <— Other (SPECIFY) 6

Thank you very much for your help.

TIME SCREENER ENDED: _____ AM PM

INTERVIEWER: FILL OUT A-F BELOW & SAMPLING REPORT & INTERVIEWER SIGNATURE AND NUMBER ON PAGE 4 IMMEDIATELY AFTER YOU LEAVE THE HOUSEHOLD.

A. Race of household (by observation) 18

Black/Negro 1
White 2
Other (SPECIFY) _____
_____ 3
Not able to observe 4

B. Code type of living quarters: 19

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
Apartment house (5 or more units; locked entry, or guarded by doorman, or both) . 6
Other (SPECIFY) _____
_____ 7

C. Length of time for Screener: _____ minutes 20 21

D. Date of Screener: ☐☐ ☐☐ 22 25
MONTH DAY

E. With whom did you conduct the Screener?

_____ 26 27
[Name(s) of screener informant(s)]

F. If screener informant not household member, describe relationship to selected respondent or to household.

_____
(Relationship to Household)

GO TO PAGE 4

CHECK (✓) HERE IF NO MISSED DU OR OTHER CORRECTIONS AT SAMPLE ADDRESS ☐

SAMPLING REPORT ON MISSED DU'S AND OTHER CORRECTIONS

PSU #_____ Seg.#_____ Part #_____ Sample Line #_____ Case #_____

MESSAGE CODE # FROM ASSIGNMENT BOX ☐

OFFICE USE ONLY		Phone Call Col.	Address Corrections, Additions, or Deletions	Description of Corrections, Additions, or Deletions	Correction to Line #	DU's Added			DU's Deleted		OFFICE USE ONLY
DU Code (10)	Line # (11-17)					Additions at Same Address (Check ✓)	Additions at Same Apt # (Check ✓)	Additions between Line # and Line #	Line #	Reason	BLOCK # (77-80)
		1									
		2									
		3									
		4									
		5									
		6									
		7									
		8									
			Total Additional DU's →								

DECK 21

SIGNATURE OF INTERVIEWER: _____

INTERVIEWER ID NUMBER: ☐☐☐☐☐ 28 32

IF 1 TO 4 DU'S:
FILL OUT HOUSEHOLD SCREENER FOR EACH & CONDUCT SCREENER INTERVIEW.

IF 5 OR MORE DU'S:
CALL SAMPLING DEPT. (COLLECT AC 312/684-5600) FOR INSTRUCTIONS BEFORE YOU DO ANY ADDITIONAL SCREENER INTERVIEWS.

AT TIME OF PHONE CALL, CIRCLE CODE IN "PHONE CALL COL" FOR LINE(S) AT WHICH YOU ARE INSTRUCTED TO DO SCREENER INTERVIEW(S); THEN FILL OUT SCREENER FOR EACH OF THOSE & CONDUCT SCREENER INTERVIEW.

☆ U.S. GOVERNMENT PRINTING OFFICE: 1978— 260-937:41

VITAL AND HEALTH STATISTICS PUBLICATIONS SERIES

Formerly Public Health Service Publication No. 1000

Series 1. Programs and Collection Procedures.—Reports which describe the general programs of the National Center for Health Statistics and its offices and divisions, data collection methods used, 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, 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, based on data collected in a continuing national household interview survey.

Series 11. Data from the Health 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

DHEW Publication No. (PHS) 78-1350
Series 2-No. 76

NCHS
U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Public Health Service
National Center for Health Statistics
Center Building
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 HEW
HEW 390

THIRD CLASS

BLK. RATE