Vital and Health Statistics

National Survey of Family Growth, Cycle IV, Evaluation of Linked Design

Series 2:
Data Evaluation and Methods Research
No. 117

Research was undertaken to quantify the effects of costs of alternative methods for selecting sample women for the National Survey of Family Growth (NSFG) from the National Health Interview Survey (NHIS). This report presents estimates of the effects of alternative design options, obtained by statistical modeling techniques, for linking the NSFG with the NHIS; the cost data and the statistical precision of estimates were based on data from the NSFG, Cycle IV. The estimated survey costs and projected response rates for alternative linked design options and for the unlinked design are compared for fixed precision. The findings confirm that substantial gains in the NSFG design efficiency were obtained by linking the NSFG sample design to that of the NHIS.

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
Centers for Disease Control and Prevention
National Center for Health Statistics

Hyattsville, Maryland
July 1993
DHHS Publication No. (PHS) 93-1391

Suggested Citation

Waksberg J, Sperry S, Judkins D, Smith V. National Survey of Family Growth, Cycle IV, evaluation of linked design. National Center for Health Statistics. Vital Health Stat 2(117). 1993.

Library of Congress Cataloging-in-Publication Data

National Survey of Family Growth, cycle IV : evaluation of linked design.
p. cm. – (Vital and health statistics. Series 2, Data evaluation and methods research ; no. 117) (DHHS publication; no. (PHS) 93-1391)
Prepared by J. Waksberg and others.
July 1993.
Includes bibliographical references.
ISBN 0-8406-0482-3
1. National Survey of Family Growth (U.S) 2. National Health Interview Survey (U.S.) 3. Family size – United States – Statistical methods. 4. Fertility. Human – United States – Statistical methods. 5. Family life surveys – United States. 6. Health surveys – United States. 7. Health Surveys – United States. 8. Sampling Studies. I. Waksberg, Joseph. II. National Center for Health Statistics (U.S.) III. Series. IV. Series: DHHS publication ; no. (PHS) 93-1391.
[DNLM: 1. National Survey of Family Growth (U.S.)
2. National Health Interview Suirvey (U.S.) 3. Family Characteristics.
4. Research Design. 5. Population Growth – United States.
W2 A N148vb no. 117 1993]
RA409.U45 no. 117
[HQ762]
362.1'0723 s – dc20
[304.6'2'0723]
DNLM/DLC
for Library of Congress 93-24687
CIP

For sale by the U.S. Government Printing Office
Superintendent of Documents
Mail Stop: SSOP
Washington, DC 20402-9328

National Center for Health Statistics

Manning Feinleib, M.D., Dr.P.H., *Director*

Jack R. Anderson, *Acting Deputy Director*

Jacob J. Feldman, Ph.D., *Associate Director for Analysis and Epidemiology*

Gail F. Fisher, Ph.D., *Associate Director for Planning and Extramural Programs*

Peter L. Hurley, *Associate Director for Vital and Health Statistics Systems*

Robert A. Israel, *Associate Director for International Statistics*

Stephen E. Nieberding, *Associate Director for Management*

Charles J. Rothwell, *Associate Director for Data Processing and Services*

Monroe G. Sirken, Ph.D., *Associate Director for Research and Methodology*

David L. Larson, *Assistant Director, Atlanta*

Office of Research and Methodology

Monroe G. Sirken, Ph.D., *Associate Director*

Kenneth W. Harris, *Special Assistant for Program Coordination and Statistical Standards*

Lester R. Curtin, Ph.D., *Chief, Statistical Methods Staff*

James T. Massey, Ph.D., *Chief, Survey Design Staff*

Andrew A. White, Ph.D., *Chief, Statistical Technology Staff*

Preface

This report is the third in the NCHS Vital and Health Statistics Series that evaluates the error and cost effects of linking the sample design of the National Survey of Family Growth (NSFG) to the design of the National Health Interview Survey (NHIS). NSFG Cycles I–III had been designed as stand-alone area household sample surveys. The sampling paradigm of the linked NSFG survey design is that the file of names and addresses of the NHIS sample households, including the information collected about them in NHIS, serves as the NSFG sampling frame.

The first report, Integration of Sample Design for the National Survey of Family Growth, Cycle IV, With the National Health Interview Survey (Series 2, No. 96), estimated the NSFG design effects resulting from linking the NSFG to the NHIS design instead of designing the NSFG independently. The design effects based on statistical modeling investigations were quite encouraging and indicated that linkage was likely to produce substantial gains in the NSFG sample design efficiency.

The second report, Linking the National Survey of Family Growth With the National Interview Survey (Series 2, No. 103), presented results that were based on survey experiments in which alternative design options for linking the NSFG to the NHIS were tested. The experiments demonstrated conclusively the feasibility as well as the efficiency of linking the NHIS and NSFG sample designs. Although the findings did not lead to a definitive determination of the optimal design for integrating the NHIS and NSFG designs, the findings did provide the basis for making informed choices for integrating the NSFG Cycle IV design with the NHIS.

This, the third report, evaluates the linked design that was executed in conducting NSFG Cycle IV. For fixed precision, it compares the actual NSFG Cycle IV costs with those expected for an unlinked design and for two somewhat more efficient linked designs that might have been executed under more ideal conditions.

After reading an initial draft of this report, I asked Westat, Inc, to add a section that compared the linkage cost reductions actually realized in NSFG Cycle IV with those conjectured on the basis of the earlier experimentation and to reconcile any differences. The reader is referred to the section "Reconciliation with earlier projections" for these comparisons. Suffice it to say that the agreement is really quite close—the linked design was expected to and actually did reduce NSFG Cycle IV costs by about 25 percent. Plans are under way to link the NSFG Cycle V to the 1993 NHIS in much the same way that NSFG Cycle IV was linked to the NHIS.

It is not often that researchers have the opportunity to develop and test innovative survey designs to proceed systematically in the manner that was done here—from modeling, to field experiments, and finally to the main survey. Hence, it is noteworthy that this in-depth research effort has been more than justified by the improvements that were achieved in design efficiency and by the analytic enhancement possibilities provided by the merged NSFG and NHIS microdata sets.

Congratulations are owed to the Westat staff who prepared these three reports and the NCHS staff who worked with them. The NCHS staff included Andrew White, who provided technical oversight in the preparation of the first report; Deborah Trunzo (nee Bercini), who provided oversight during the entire second phase of the research project; and Steve Botman, who provided oversight during the final phase of the project including the preparation of this report. As noted in the preface to an earlier report, this project would not have been possible without the support and cooperation of Robert Fuchsberg and Owen Thornberry, the former and current directors, respectively, of the Division of Health Interview Statistics, and William Pratt, former chief of the Family Growth Survey Branch, Division of Vital Statistics.

Monroe G. Sirken, Ph.D.
Associate Director Research and Methodology

Contents

Appendixes

Text tables

Symbols

---	Data not available
. . .	Category not applicable
-	Quantity zero
0.0	Quantity more than zero but less than 0.05
Z	Quantity more than zero but less than 500 where numbers are rounded to thousands
*	Figure does not meet standard of reliability or precision

The National Survey of Family Growth, Cycle IV, Evaluation of Linked Design

by Joseph Waksberg, Sandy Sperry, David Judkins, M.A., and Valerija Smith, M.A., Westat, Inc.

Introduction

The National Survey of Family Growth (NSFG), Cycle IV, conducted in 1988, provides information on childbearing, contraception, and related aspects of maternal and child health for women who were of childbearing age at that time. The NSFG sampled women were selected from women who had previously been sampled for the National Health Interview Survey (NHIS). Complete NHIS interviews included women who responded directly to the interviewer (self-respondents) and those who responded indirectly through proxies. In this report, both groups are referred to simply as "respondents" or as "the interviewed sample." The NHIS sample women had been interviewed between the fourth quarter of 1985 and the first quarter of 1987. The purpose of the linkage was to reduce NSFG costs while keeping sampling error constant. A report by Waksberg and Northrup (1) had projected that the linked design would be more efficient than the approach of drawing fresh area-based samples, which was the approach used for the first three cycles of the NSFG.

Indeed, the achievement of cost or variance efficiencies in surveys such as the NSFG was a major goal in the redesign of the NHIS for 1985–94. The 1985 design did not use the decennial Census lists in sample selection that all the other current surveys of the Bureau of the Census use (e.g., the Current Population Survey, the National Crime Survey, the Survey of Income and Program Participation, and the American Housing Survey). When these lists are used, the sample addresses cannot be released for 70 years to anyone who is not a sworn employee of the Bureau of the Census. By avoiding decennial list-based sampling, the NHIS sample addresses may be made available to non-Census employees under strict privacy safeguards. A drawback for NHIS of using mixed area- and permit-based samples is that the area-based component is much more expensive than a comparable list-based component. Another drawback of the mixed area- and permit-based samples is that it is more difficult to avoid double coverage of some housing units with this dual approach than it is to avoid double coverage between decennial list-based and permit-based samples. The obvious question is whether the efficiencies for follow-on surveys to the NHIS pay for the higher cost of the NHIS itself.

The purpose of this study was to answer part of this question and to quantify the cost savings of linkage for the NSFG. Although considerable research on this same topic (1,2) preceded the decision to use NHIS as the sampling frame for Cycle IV, such prior research can never quite operate under the actual conditions that prevailed during the data-collection operations. Studies similar to this study will need to be carried out on other follow-on surveys to the NHIS to answer the overall question of whether the area sample of old construction for NHIS better suits the overall program of the National Center for Health Statistics (NCHS) than would a list sample of old construction.

Strategy

The basic strategy was to project the cost for an unlinked NSFG design that would provide the same sampling precision and compare it with the cost for the linked design. The first conceptual problem encountered was determining which linked design would be considered. These considerations were brought on by the very special circumstances that surrounded the Cycle IV linked design. The survey was originally intended to be conducted in mid-1987 with women interviewed for NHIS in 1986. However, there were major revisions in the questionnaire late in the survey planning process, so the interviewing was conducted around March 1988 instead of July 1987. These delays raised costs and hurt response rates by requiring more extensive tracing. The other special circumstance was that, as a result of budgetary difficulties, the sample size for the 1986 NHIS was only one-half the intended size. NSFG requirements for black sample sizes could not be met from this reduced NHIS sample. As a result, black women who had been interviewed for NHIS in late 1985 and in early 1987 were added to the sampling frame.

This dilemma was resolved by estimating the cost of three linked designs:

- Design B, which assumes that NSFG IV data collection was conducted around July 1987 and not March 1988;
- Design C, which was the actual 1988 design, was conducted around March 1988; and
- Design D, which assumes that the 1986 NHIS was at full strength and that NSFG IV data collection was conducted around July 1987 and not March 1988.

These statements do not specify unique designs. Additional assumptions and estimates of components of variance are required. The same is true for Design A, the unlinked design. Before explaining how the components of variance were estimated, it is helpful to review the designs for NHIS and Cycle IV of NSFG.

Chapter 1
Design of Cycle IV

Summary

The sample for Cycle IV of the NSFG is a subsample of women whose households had participated in the NHIS, a continuous survey of the civilian noninstitutionalized population of the United States. When the full NHIS sample can be used, interviews are obtained at 47,600 housing units each year in a fixed set of 198 metropolitan areas and clusters of nonmetropolitan counties. Data are collected on each household member about disabilities, health conditions, doctor visits, hospitalizations, and other health-related topics. A new set of households is interviewed each year.

NCHS provided computer files to Westat of households that participated in the NHIS together with address information, rosters, and some basic demographic data on household members. Households were included that had been interviewed for NHIS any time between the fourth quarter of 1985 and the first quarter of 1987, inclusively. From these, Westat selected the NSFG sample. Households were drawn from 156 of the 198 primary sampling units (PSU's) in the NHIS design. In comparison, Cycle III was confined to 79 PSU's. Spreading the sample across more PSU's resulted in smaller sampling errors.

No more than one woman was selected for the NSFG sample per household. Interviewers attempted to locate these women, following them to new addresses if necessary. After locating a sampled woman, the interviewer conducted a brief screener to confirm that she was indeed eligible (between the ages of 15 and 44, inclusively).

Design of the National Health Interview Survey

The form of the NHIS sample redesign of 1985 (3) made it possible for NCHS to transmit data on NHIS sample households to private contractors for use in conducting follow-on surveys, which are then said to be linked to the NHIS. The confidentiality of the transmitted data is protected under section 308(d) of the Public Health Service Act.

The NHIS sample for the years 1985 through 1994 is restricted to 198 PSU's. These sample PSU's were selected from a much larger set of PSU's that covers the United States. The sample selection method was based on a stratified probability design. This means that the PSU's were grouped prior to selection to ensure that the selected PSU's would be broadly representative of the total U.S. population in terms of several demographic and economic characteristics. Some of these PSU's are so populous that they were included in the sample with certainty. These are called self-representing (SR) PSU's. There are 52 SR PSU's in the full NHIS design. The remaining 146 PSU's had a chance of not being selected. These PSU's would thus represent both themselves and other PSU's that were not selected. Hence, they are called non-self-representing (NSR) PSU's.

To allow flexibility to conduct the NHIS with any of several different sample sizes, the PSU's are divided into four panels, each of which can be used to represent the Nation, if need be. The largest SR PSU's are in all four panels. Medium-sized SR PSU's are in two panels. There are 62 PSU's in a single panel sample, 112 PSU's in a two-panel sample, 156 PSU's in a three-panel sample, and 198 PSU's in the full design.

Within each sample PSU, a sample of blocks (or small groups of blocks) was selected. In PSU's in which black persons constituted 5–50 percent of the population, blocks in enumeration districts with a higher percent of black persons were selected with a higher probability than other blocks. Within each block, a cluster of eight housing units was selected. These housing units were spread as evenly throughout the block as possible.

To gain better control over the size of the sample, housing units constructed since the 1980 census were selected through a sample of building permits rather than through area sampling. These units were selected in clusters of four instead of eight.

To provide continuous coverage of the population throughout the year, the sample of households was spread over the 52 weeks of the year, with each week's sample a random subsample of the total sample. Each year, a totally new sample of households is selected. However, they tend to be neighbors of the households interviewed the previous year.

Subsampling of NHIS completed interviews

The procedure for selecting the NSFG sample from the NHIS sample was complex. In this section, the factors motivating the design are described in tandem with the design features themselves. For readers more interested in any effects of the design than in motivating factors, suffice it to note (a) that in some PSU's, only black women were selected, (b) that neighborhood clusters of black women tended to be larger than clusters of women of other races, and (c) that households containing more than one eligible woman of a race other than black were selected at a higher rate than households containing just one such woman. This last point considerably reduces the variability in the sampling weights of women of races other than black. The weights of black women tend to vary much more strongly than the weights of other women.

The NSFG sample was drawn from women whose households had participated in the NHIS in the fourth quarter of 1985, any time during 1986, or in the first quarter of 1987. Because of a lack of adequate funding, the 1985 NHIS sample was restricted to three panels (156 PSU's), and the 1986 NHIS sample to just two panels (112 PSU's). Funding was augmented for 1987; thus the 1987 NHIS sample is found in all 198 PSU's of the full NHIS design. Unfortunately, even combining all six of the available quarters together did not provide as many black women as were selected for Cycle III of the NSFG. The decision was thus made to select as many of these women as possible, subject to operational constraints and the restraint of selecting just one woman per household. The only black women who were not selected were those who resided in the 42 PSU's that were used by NCHS only in 1987. It was judged that the travel costs per completed interview would have been too high for the women in these PSU's.

Combining all six of the available quarters together provided many more women of races other than black than were required for the NSFG. In deciding how to subsample, the general preference was to take the most recently interviewed because they would be the least likely to have moved since the NHIS interview. (Such a procedure does not introduce bias because each week's sample is a random subsample of the total sample.) It appeared, however, that the household information from the first quarter of 1987 might not be available for sampling in time; therefore, an initial decision was made to restrict the sample to 1986. Subsequently, timing ceased to be as tight and more funding was made available, so the sample of women other than black was expanded to include some from the first quarter of 1987.

The first step was to select households. The second step was the selection of persons from households. Rules for selection of households are summarized in table A and listed below:

- All NHIS sample households in the 156 PSU's used in the fourth quarter of 1985 containing one or more eligible black women were selected.
- All NHIS sample households from 1986 containing one or more eligible black women were selected.
- All NHIS sample households from the first quarter of 1987 containing one or more eligible black women were selected if they lived in one of the same 156 PSU's used in the fourth quarter of 1985.
- All NHIS sample households from 1986 containing more than one eligible woman of a race other than black were selected.
- The NHIS sample is split into 52 subsamples corresponding to weeks of the year. Households from 30 of these 52 subsamples from 1986 containing exactly one eligible woman who was not black were selected. Drawing fewer of these households than of households containing more than one eligible woman who was not black makes up for the fact that only one of the women in each multieligible household could be interviewed. If the sample had not been selected in this manner, women who are not black from the following sorts of households would have been underrepresented: mother and daughter both between 15 and 44 years of age, sisters both between 15 and 44 years of age, and unrelated women both between 15 and 44 years of age. Note that black women in multieligible households are underrepresented because there were not enough eligible black women in the total NHIS sample to allow the subsampling of black women in single-eligible households. For this and other reasons, it is thus crucial for microdata users to use the provided sampling weights.
- A few NHIS sample households (1 week out of 13) were selected in two of the available panels from the first quarter of 1987 if they met a specific criterion. The criterion was that the household contain exactly one eligible woman who was not black and no eligible black women were selected.
- Households assigned to 12 out of the 13 weekly NHIS subsamples in two of the available panels from the first quarter of 1987 were selected if they met a different specific criterion. That criterion was that the household contain exactly two eligible women who were not black and no eligible black women.
- All NHIS sample households in two of the available panels from the first quarter of 1987 containing three or more eligible women other than black women but no eligible black women were selected.

Within a given household, all eligible women had the same probability of selection. The probability of selection was one divided by the number of eligible women. Eligibility was defined in terms of exact age on March 15, 1988: A woman had to be 15–44 years of age on that date. There was one minor exception to this rule. Within multiracial households selected from the first quarter of 1987, only black women had a chance of selection. Each of the black women in such a household had the same probability of selection.

Field adjustments

There were rare instances in which the sampled woman was under age 15, over age 44, or not a woman at all. (NHIS age and sex information were imputed if missing, causing some errors. Even where the data had not been imputed, other errors were found.) In these cases, the interviewer selected among other eligible women then residing in the household. If there were no eligible women, the case was dropped.

Subsampling for nonresponse followup

After all efforts to complete an interview were exhausted by local interviewers, a 50-percent subsample of nonresponse cases was selected for intensive followup. This subsampling, designed to reduce interview costs, was accomplished in two ways. In PSUS for the six largest metropolitan areas, where there were large numbers of nonresponse cases, the nonresponse cases were sequenced by an identification number, and a systematic sample of one-half of them was drawn. The remaining PSU's were sequenced in descending order by the number of nonresponse cases they contained. A 50-percent sample of these PSU's was selected systematically. Among the selected cases, those that appeared to be convertible were assigned to a corps of traveling interviewers and assistant supervisors who had previously demonstrated superior ability in refusal conversion.

Prior to the followup, the response rate was 77.9 percent. Of the 8,450 final respondents, 220 were obtained as a result of the nonresponse followup. Counting each of these 220 interviews twice, because each woman represents herself and one other woman, boosts the response rate from an unadjusted 80.0 percent (8,450/10,562) to an effective response rate of 82.1 percent ((8,450+220)/10,562).

Table A. Rules for selecting households from the National Health Interview Survey sample for Cycle IV of the National Survey of Family Growth, by year and quarter interviewed in the National Health Interview Survey and race and number of eligible women living in the household

Race and number of eligible women living in the household	*Year and quarter interviewed*		
	1985 fourth quarter	*1986 all quarters*	*1987 first quarter*
		Households selected for the NSFG	
Black			
At least 1	All in 156 PSU's	All in 112 PSU's	All in 156 PSU's
All others			
Exactly 1	None	30 weeks of every 52 in 112 PSU's[1]	1 week of 13 in 112 PSU's[2]
Exactly 2	None	All in 112 PSU's	12 weeks of every 13 in 112 PSU's[3]
At least 3	None	All in 112 PSU's	All in 112 PSU's

[1] The last 30 weeks of 1986.
[2] The last week of the first quarter of 1987.
[3] The last 12 weeks of the first quarter of 1987.
NOTES: NSFG is National Survey of Family Growth. PSU is primary sampling unit.

Chapter 2
Components of variance

To estimate the components of variance, the technique of balanced repeated replications (BRR) was used. General information on this technique may be found in Wolter (4). The actual NSFG sampling strata were not set up in the way required for BRR. To fit the sample into the BRR model, the sample women were regrouped into variance strata and variance units, corresponding to the sampling strata and half-samples of classical statistical theory. Different sets of variance strata and variance units were formed to estimate the various components of variance.

Total variance

Two types of variance strata were established to estimate total variance. The first set consists of groups of NHIS second-stage units (segments) within SR PSU's (self-representing primary sampling units). A total of 58 variance strata of this type were formed. Each such variance stratum is a systematic subsample of groups of four consecutive NHIS segments from SR PSU's. The subsamples are exhaustive and mutually exclusive. Prior to the grouping and systematic subsampling, the segments were sorted by region, PSU, and order of selection (by the Bureau of the Census) for NHIS. The second set of variance strata consists of groups of NSR PSU's. There were 42 of these strata. Each variance stratum of the second type consisted of between two and five NSR PSU's from similar sampling strata.

The two variance units for each of the first type of variance stratum consisted of a systematic subsample of the NHIS segments in the variance stratum and the complement of that subsample. The sort was the same as for the identification of the variance strata. The pattern of variance-stratum and variance-unit assignment in SR PSU's can thus be illustrated as: 1A, 1B, 1A, 1B, 2A, 2B, 2A, 2B, ..., 58A, 58B, 58A, 58B, 1A, 1B, 1A, 1B, ... (Consecutive sets of four segments were grouped together into the same variance stratum rather than the more traditional consecutive sets of two segments because, at the time, it was believed that this might improve the stability of the variance estimator. Subsequent discussions have thrown this thinking into doubt. For future work, consecutive pairs are recommended.)

The two variance units for each of the second type of variance stratum were formed by dividing the PSU's in the variance stratum into two groups. Care was taken to ensure that the exclusively black PSU's were always in opposite variance units.

The perturbation factors were the standard 0 and 2 of BRR for the variance units in the first type of variance stratum. The perturbation factors were 0 and roughly 2 for the variance units in the second type of variance stratum. The deviations from the standard 2 were made to allow for odd numbers of PSU's and variation in the sizes of the sampling strata represented by the PSU's in a variance stratum. More detail on the variance stratum-unit formation and perturbation may be found in Judkins, Mosher, and Botman (5).

Within-unit variance

The variance strata for within-PSU variance were much simpler than those for total variance. All NHIS segments with any women designated for the NSFG IV sample were sorted by region, PSU, and order of selection, grouped into consecutive preliminary variance strata of four segments each, and then systematically subsampled into 100 samples. Each sample of preliminary variance strata constituted a final within-PSU variance stratum. Each variance stratum was systematically split into two variance units. The pattern of variance-stratum and variance-unit assignment within PSU's can thus be illustrated as: 1A, 1B, 1A, 1B, 2A, 2B, 2A, 2B, ..., 100A, 100B, 100A, 100B, 1A, 1B, 1A, 1B, ...

Within-segment variance

The variance strata for within-segment variance were formed in a manner similar to the within-PSU variance strata. All designated women were sorted by region, PSU, segment order of selection, and household identification number within segment, grouped into consecutive preliminary variance strata of four women each, and then systematically subsampled into 100 samples. Each sample of preliminary variance strata constituted a final within-PSU variance stratum. Each variance stratum was systematically split into two variance units. (As at the segment level, it is now believed that consecutive pairs might have yielded better results than consecutive sets of four women. Also, special modifications for segments with one or three sample women would have improved the within-segment variance estimates.)

Replication

All stages of adjustment were repeated on each of the 100 sets of replicated weights for each component of variance. This means that nonresponse adjustment and the iterative raking to Current Population Survey estimates and demographic control totals was repeated 300 times, i.e., 100 times for total variances, 100 times for within-PSU variances, and 100 times for within-segment variances.

Smoothing

Direct estimates of between-PSU variance and within-PSU-between-segment variance could have been computed for each characteristic of interest. Past experience has demonstrated, however, that these individual estimates of the "between" components are highly unstable. In fact, negative estimates of variance can occur. To counteract this, separate generalized variance functions (GVF's) were fit for total, within-PSU and within-segment variances, and components of variance were estimated by subtracting *b* parameters rather than individual estimates. This smoothing of the variances prior to subtraction improves stability and usually results in better overall estimates of all components (6). As input to the fitting of GVF's, variances were estimated directly for a large number of characteristics.

The characteristics included cross-tabulations of age by marital status; education by type of current contraceptive; usual source of family planning services by metropolitan status; religious affiliation by ever had intercourse, by expectations for additional children; parity by fecundity; age by relative order of marriage and first birth; Hispanic origin by age at first marriage; age by education by age at first intercourse for never-married women; Hispanic origin by living arrangement at age 14 by current contraceptive use for never-married women; religion by metropolitan status by ever-usage of family planning services for never-married women; and education of mother by ever had intercourse for never-married women. There were other tables as well. All the tables were repeated for black women by themselves.

The curves were fit to the formula

$$V^2 = b(1/X - 1/T) \qquad (1)$$

where V^2 is the directly estimated relative variance for a statistic, X is that statistic, and T is the upper bound on the possible value for X. The results are shown in table B.

Table B. *b* parameters for women, by race and components of variance

Components of variance	*Black*	*All races*
Total	4,906	10,192
Within-PSU[1]	4,350	9,500
Within-segment	4,258	9,453

[1] PSU is primary sampling unit.

NOTES: These *b* parameters are slightly different from those in Judkins, Mosher, and Botman (5). A different model was used here with a different set of items to improve comparability between total, within-PSU, and within-segment variances.

Caveats

These estimates imply a 7-percent between-PSU variance for all races combined and an 11-percent between-PSU variance for black women. They also imply that between-segment variance was a trivial component of variance for most estimates. Although we believe that the general direction of these estimates is correct, several caveats are in order.

A number of approximations were made in the variance computations, almost all of which tended to exaggerate the between-PSU component. The estimates of between-PSU variances are thus upper bounds. First, there are some very large sampling fractions at the first stage (some as large as 0.75). These have a strong diminishing effect on between-PSU variance that was not reflected in the variance estimates. Second, even though some small adjustment was made to compensate for the bias in the collapsed-stratum variance estimator, such an estimator is normally positively biased. This tends to exaggerate between-PSU variance. Third, the PSU's of only black women lead to even worse positive biases because there is only one such PSU in every other stratum. Fourth, the variance strata for within-PSU variance probably should have been assigned on the basis of all NHIS segments, not just those with a woman designated for an attempted NSFG interview. This has the effect of exaggerating between-PSU variance and understating between-segment variance. (However, because poststratification by race and age essentially makes all inference conditional on the achieved national sample sizes, this is probably a very minor caveat.) Fifth, the within-PSU variance strata pair segments in different PSU's when there are PSU's with odd numbers of segments. This fifth caveat has the opposite effect of the fourth caveat but is probably smaller. Sixth, the within-segment variance strata pair women in different segments or even PSU's when there are segments with odd numbers of designated women. This has the effect of exaggerating within-segment variance, thereby understating between-segment variance.

Chapter 3
Model for components of variance

Theoretical form

A fairly good model for the relative variance of a ratio-adjusted estimator from a three-stage design such as NSFG is

$$V^2 = V^2_{\text{srs}}[1 + P\rho_1\lambda_1 + \rho_2(\lambda_2-1)\xi_2 + \xi_2-1 + a^2 + b^2 + c^2] \quad (2)$$

where V^2_{srs} is the relative variance that would be obtained from a simple random sample of women of the same size, P is the percent of women in non-self-representing PSU's, ρ_1 is the intraclass correlation at the PSU level, λ_1 is the average number of sample women per PSU, $P\rho_1\lambda_1$ is the relative increase in variance due to sampling at the first stage, ρ_2 is the intraclass correlation at the segment level, λ_2 is the average number of sample women per segment, ξ_2 is a complex term due to variation in the number of eligible women per segment that is usually greater than one and that persists even with poststratification but is difficult to estimate, $\rho_2(\lambda_2-1)\xi_2 + \xi_2-1$ is the relative increase in the variance due to sampling at the second stage, a^2 is the relative variance in within-household inverse probabilities of selection, b^2 is the relative variance of inverse probabilities of selection across second-stage strata, and c^2 is the increase in relative variance of the final weights due to subsampling of nonrespondents, adjustment for nonresponse, and other adjustments.

Following common practice, the term ξ_2 was not directly estimated for this study. This type of factor is discussed at greater length in Hansen, Hurwitz, and Madow (8) volume 1, chapter 6, section 8; volume 1, chapter 8, sections 1 and 11; and volume 2, chapter 8, section 4. There it is written as $\hat{V}^2/V^2$. There is also more discussion in an appendix to this report.

The formula used by Waksberg and Northrup was slightly different. They used

$$V^2 = V^2_{srs}[1 + P\rho_1\lambda_1 + \rho_2(\lambda_2-1) + V^2_{\lambda 2} + a^2 + b^2] \quad (3)$$

where $V^2_{\lambda 2}$ is the relative variance in the number of eligible women per segment, also not directly estimated. It might be argued that the omission of c^2 was an oversight, but other studies have indicated that poststratification can have variance-reducing properties that directly counteract some of this effect. Empirical studies in the last decade have indicated that $V^2_{\lambda 2}$ is usually on the order of 0.1 or 0.2 for area samples of persons through their housing units. Such values are too large to be reasonable. A substitution was made for the ξ_2-1 term, which is expected to be considerably smaller than $V^2_{\lambda 2}$. Note that Hansen, Hurwitz, and Madow also has a formula using $V^2_{\lambda 2}$, but that the form is slightly different. Adapting from that source to the problem at hand, the approximation from Hansen, Hurwitz, and Madow would appear to be

$$V^2 = V^2_{srs}[1 + P\rho_1\lambda_1\xi_1 + \rho_2(\lambda_2-1)\xi_2 + \xi_2-1 + a^2 + b^2 + c^2] + V^2_{\lambda 2}/m_2 \quad (4)$$

where m_2 is the number of segments. It is also important to note that the right bracket has moved, leaving the term involving variation in segment size as an absolute rather than relative term and that the term declines in importance as the number of segments increases. Based on the discussion in Hansen, Hurwitz, and Madow, this formula is more appropriate for modeling the relative variance of an estimated total that was obtained with simple inverse-probability sampling weights than for modeling the relative variance of an estimated total that was obtained by poststratified sampling weights. Poststratification largely eliminates the additive ${V_{\lambda 2}}^2$ effect, leaving only the multiplicative factor ξ_2.

To maintain consistency with Waksberg and Northrup, a value was picked for ξ_2 that had the same effect on the design effect as the $V^2_{\lambda 2}$ that was assumed. A value of $\xi_2 = 0.042$ is comparable to the $V^2_{\lambda 2} = 0.05$ that they assumed. A strategy for estimating ξ_2 in future studies is given in Appendix II.

Fitting the model for components of variance

The intraclass correlations implied by these variance components are not very consistent with those found by Waksberg and Northrup when examining Cycle II. This was true even though there was considerable overlap in the table structures that served as the basis for inference about components of variance. (For example, both studies estimated components of variance for the estimated number of black women who used contraception.) Of course, in addition to all the caveats mentioned previously, there are problems in translating the b parameters into components of variance.

First-stage components

In this section, values for P, λ_1, and ρ_1 are obtained. There were a total of 156 PSU's with one or more designated sample women for Cycle IV. Of these, 130 were NSR PSU's. However, no attempt was made to interview in several of these PSU's because the sample sizes simply did not warrant the tremendous travel expenses associated with the PSU's. The number of NSR PSU's with at least one completed interview of a black woman is 88. The number of NSR PSU's for all-race estimates is 112, and the number for other-than-black estimates is 88. Table C shows, by racial grouping, the numbers of NSR PSU's, the complete sample sizes in NSR PSU's, the proportions of the eligible populations in NSR PSU's, and the resulting values of λ_1. Substitution of these parameters into the formula

$$\rho_1 = \frac{b_T - b_{WP}}{b_{WP} P \lambda_1} \tag{5}$$

where b_T is the b parameter for total variance and b_{WP} for within-PSU variance, gives the estimates of intraclass correlation, ρ_1, at the PSU level shown in table C.

These do not line up very well with the Waksberg and Northrup estimates also shown in table C. Intraclass correlation should be about the same for all races as for all races other than black by themselves. It is thus particularly troubling that the estimated correlation for black women in Cycle II was much lower than that for persons other than black but that this relationship is reversed in Cycle IV with the estimated correlation for black women being considerably larger than that for all races. The difference in smoothing techniques is part of the reason for the disparity. The Cycle IV smoothing technique was applied to the Cycle II variances rather than averaging direct-difference estimators of between-PSU variance. This reduced the discrepancy somewhat in favor of the Cycle IV estimates, particularly for black women; hence the lean toward the Cycle IV estimate for black women. Finally, the Cycle IV estimates of between-PSU variance were positively biased for the reasons given earlier in this section (whereas the Cycle II estimates were not) and that the Cycle IV estimates were prepared with more PSU's than the Cycle II estimates and should thus have better stability. However, there was concern that there was still a large variance in our variance-component estimate, and thus it was assumed that the single intraclass correlation of 0.005 shown in the "composite" column of table C for all races combined, races other than black by themselves, and black women by themselves.

Second-stage components

Similar problems and inconsistencies beset the determination of the intraclass correlation at the second stage (within segments). Table D gives, by racial grouping, the numbers of NHIS segments with one or more completed interviews, repeats the completed sample sizes, and gives the average segment sizes (λ_2). The standard formula for intraclass correlation at the second stage is

$$\rho_2 = (b_{WP}/\, b_{WS} - 1 - \xi_2)/\, [(\lambda_2 - 1)\xi_2] \tag{6}$$

Substitution gives the estimates shown in table D. The estimates obtained by Waksberg and Northrup are also shown. It is apparent that the estimate of ξ_2 is too large for this study or that b_{WP}/b_{WS} has not been well estimated; it is simply not reasonable to assume a negative intraclass correlation. The very small number of completed interviews per segment probably led to substantial contamination of b_{WS} by between-segment differences. Partially counteracting this was the fact that women other than black selected from the 1985 and 1987 NHIS samples were in segments adjacent to those selected from the 1986 NHIS sample. This was reflected in the direct estimates of variance components and smoothing but was not reflected in the computation of the number of nonzero segments. If allowance had been made for the neighboring segments, this would have reduced the effective number of segments, thereby boosting the average cluster size and bringing the estimated intraclass correlation closer to zero. In trying to reconcile these two sets of estimates, it must be kept in mind that the average number of completed interviews per segment for Cycle II was much larger than for Cycle IV, even though the typical land area was similar. The larger number of completed interviews means that segments consisting of just a single woman were rare. That rareness allows for more accurate estimation of the intraclass correlation. Thus, the Cycle II estimates were weighted more heavily in coming up with a composite intraclass correlation.

Table C. First-stage intraclass correlation for Design C

Race	*Non-self-representing primary sampling unit*			Average non-self-representing workload (λ_1)	Cycle IV (ρ_1)	Cycle II (ρ_1)	Composite (ρ_1)
	Number with nonzero sample	Number of interviewed women	Proportion of population[1]				
All races	112	4,693	0.69	41.9	.0025	---	.0050
Black	88	1,362	0.58	15.5	.0142	.0007	.0050
Other than black	88	3,331	0.71	37.9	---	.0066	.0050

[1] Not actually estimated from the National Survey of Family Growth. Estimated from 1991 Westat Master Sample.

Table D. Second-stage intraclass correlation for Design C

	Number		Segment size				
Race	Effective segments	Interviewed women	Average (λ_2)	Effect of variation (ξ_2)	Cycle IV (ρ_2)	Cycle II (ρ_2)	Composite (ρ_2)
All races	3,143	8,450	2.69	1.042	–0.02	---	0.03
Black	1,056	2,811	2.66	1.042	–0.01	0.042	0.03
Other than black	2,382	5,639	2.37	1.042	---	0.046	0.03

Table E. Components of within-segment design effects for Cycle IV, by race

Component	All races	Black	Other than black
Within-segment b	9,453	4,258	---
Sampling interval.	6,852	2,732	---
Within-segment design effect	1.38	1.56	---
a^2+b^2	0.27	0.55	0.07
c^2	0.12	0.00	0.09
Relative variance in weights			
Before nonresponse sampling . .	0.27	0.55	0.07
After nonresponse sampling . . .	0.39	0.64	0.17
Final weights.	---	0.51	0.16

Effects of variation in weights

The b parameters for within-segment variance in table B are still larger than what would be expected from a simple random sample. Using the sampling intervals shown in table E gives the estimated within-segment design effects also shown in table E ($\text{deff}=b/SI$). (The design effect [deff] is the ratio of the actual variance of a sample to the variance of a simple random sample of the same number of elements.) These within-segment design effects are explained by the variation in weights as is theoretically to be expected. As explained above, the effect of the variation in weights can be decomposed into three terms: a^2 is for the subsampling of women within households containing multiple eligible women; b^2 is for the oversampling in the NHIS of neighborhoods with a high percent of black persons; and c^2 is the subsampling of nonrespondents at the close of regular interviewing for the special followup effort and the adjustments for nonresponse and poststratification. The terms a^2 and b^2 were not separately estimated for Cycle IV. The term c^2 is equal to zero for black women because the relative variance in the weights of black women was actually smaller after subsampling for nonresponse, adjusting for nonresponse, trimming, and poststratification than it was prior to these steps. This is shown in more detail in the final rows of the table.

Comparison of fitted model with empirical design effects

The fitted model was used to recompute the achieved precision for Cycle IV. The purpose of this was not to provide an improved estimate of the precision because the direct empirical measures of precision are probably better. Instead, the purpose was to provide a more even playing field for the comparison of linked and unlinked designs. Table F shows, by racial grouping, all the parameters of the fitted model for design effects along with the direct design effects. It also shows the effective sample sizes (nominal sample sizes divided by their respective design effects).

Table F. Comparison of modeled and observed design effects for Cycle IV, by parameter and race

Parameter	All races	Black	Other than black
P.	0.69	0.58	0.71
Number of NSR PSU's	100	85	88
λ_1	58.3	19.2	45.5
ρ_1	0.005	0.005	0.005
$P\lambda_1\rho_1$	0.20	0.06	0.16
Number of nonempty segments . .	3,143	1,056	2,382
λ_2	2.69	2.66	2.37
ρ_2	0.03	0.03	0.03
ξ_2	1.042	1.042	1.042
$\rho_2(\lambda_2-1)\xi_2+\xi_2-1$	0.09	0.09	0.08
a^2+b^2.	0.27	0.55	0.07
c^2.	0.12	0.00	0.09
Modeled design effect	1.68	1.70	1.40
Observed design effect.	1.49	1.80	---
Number of interviewed women . . .	8,450	2,811	5,639
Modeled effective sample size . . .	5,029	1,654	4,027

NOTE: NSR PSU's are non-self-representing primary sampling units.

Chapter 4
Design specification

Various aspects of required sample sizes are worked out for each design in this section. The goal was to keep precision at the level of the modeled effective sample sizes shown in table F and then to determine the cost penalties or savings associated with the alternate designs. To determine the cost of an alternate design, it was necessary to know the number of PSU's, the number of segments, the number of designated households, the number of designated women from those households, and the number of interviewed women. These critical statistics are shown in table O for each of the designs.

Design C—Most of the parameters for Design C (the actual Cycle IV design) were already set in table F. The response rate (among those who responded to the NHIS) was 82 percent. Even though original design specifications called for a sample mix nearer to parity between black women and women other than black, this will serve as the baseline reliability for the evaluation. (Specifications for future designs may repeat this, establish requirements for women of Hispanic origin, expand the eligible universe to women 54 years of age, or utilize other criteria.) The other designs have been specified in a manner to provide approximately the same reliability.

Design B—This design is a slight variation on design C. Going to the field within 7 months of the close of NHIS interviewing for 1986 would have improved response rates a little and would have lowered costs. Projections of these improvements are given in the final section. From the point of view of statistical design, the only difference between Designs B and C is that B requires a slightly smaller designated sample size than C. There is no reason to think that design effects would have been affected at all.

Design D—This design is similar to Design B, but some reduction in design effects could be expected. First, the sample could be spread out across 146 NSR PSU's instead of just 100 or so and across 5,000 segments (of the total 8,200 in the full NHIS), instead of just 3,200. Second, a greater proportion of the population would be covered by SR PSU's. Third, because there is a larger sample frame available, it would be possible to oversample multi-eligible households even more sharply than in Cycle IV. It should be easy to reduce the total design effect for statistics for women other than black women from the modeled 1.37 to 1.3. A design effect of 1.15 is probably a lower limit, given subsampling of nonrespondents, but 1.3 is conservative. Similarly, the increase in the numbers of sample PSU's and segments and the reduction in the need to take all households containing just one eligible woman will serve to reduce the design effect for statistics for black women from the modeled 1.67 to 1.6. A further consideration making these reductions in design effects plausible is that there is variation in NHIS baseweights resulting from differential sampling by minority density stratum. Much of this variation for single-eligible households could be removed during the subsampling of NSFG from a full NHIS. These design-effect reductions imply that the sample sizes in terms of completed interviews could be reduced from 5,639 and 2,811 to 5,351 and 2,693 for women other than black and black women, respectively. A reasonable rounded total is thus 8,000 completed interviews instead of 8,450.

Design A—The greatest uncertainties of design specification pertain to the unlinked design. It is not appropriate to simply compare effective sample sizes from Cycle II or Cycle III with those from Cycle IV because the requirements for sample allocation were sharply different. Nor can the estimates of design effects for the unlinked design in Waksberg and Northrup be used because there have been changes in reliability requirements. Rather, it is necessary to start with the components of variance derived above to build a new design. The same models are used as described in the section "Theoretical form," but new parameters are obtained. The parameter derivation follows the same structure as described under "Fitting the model for components of variance."

First-stage components

The Cycle III model had 80 PSU's, 60 of which were non-self-representing. Some of those contained no or very few interviewed black women. A rough projection of the effective number of PSU's for black women is 30. The number of interviewed women was obtained as the result of a directed-iteration approach that gave the desired level of precision taking into account all contributions to design effects. Derivation of the effect is shown in table G.

"Between-stratum" is an approximation of the size of a component that would exist in an unlinked design but did not exist in the Cycle IV linked design. In Westat's typical unlinked design for NSFG, there are 60 NSR PSU's selected from 40 NSR strata. This 1.5 PSU's per stratum design arises from Westat's desire to have a

Table G. First-stage contribution to design effect for Design A

Race	Effective number of NSR PSU's	Number of interviewed women	Proportion in NSR PSU's[1]	Average NSR workload (λ)	Composite (ρ)	Contribution to design effect $P(\lambda_1-1)\rho_1$	Between-stratum component
All races	60	11,763	0.73	143.1	0.005	.52	.02
Black	30	3,180	0.61	64.7	0.005	.20	.02
Other than black	60	8,582	0.75	107.3	0.005	.40	.02

[1] From Waksberg and Northrup, table 13 (1).

NOTE: NSR is non-self-representing. PSU is primary sampling unit.

flexible design that can be fielded with either 40, 60, or 80 NSR PSU's in addition to a constant 20 SR PSU's. However, this feature introduces an additional component of variance known as between-stratum variance. It has never been measured for NSFG characteristics but is assumed to be small. For purposes of this evaluation, Westat's estimate is that it adds 0.02 to the typical design effect.

Second-stage components

Black women were oversampled in Cycle III by oversampling block groups that had a fairly high percent of black persons in the prior decennial census and then by screening the listed households to identify those with black female occupants. To determine the within-PSU design effects, it is first necessary to work out the distribution of the sample across the second-stage strata. The within-PSU design starts with the assumption that four strata would have been used, defined by density of the black population within the block. These strata are referred to as "old construction strata" because the area sample would have been used mainly for residents of buildings constructed prior to the previous census. Additionally, because the survey took place near the end of the decade, a separate stratum would have been established for new construction built in localities that issue building permits. The housing units in this stratum would have been sampled by sampling the permits. Also, because the interviewing took place mainly in the spring, it would have been more efficient to interview college women in their dormitories rather than at home. (NHIS also has a dormitory sample.) The proportions of black, other than black, and all race women in each stratum shown in table H are based on 1980 decennial census counts with some ad hoc adjustments. The distribution of the population across these strata is subject to seasonal, decennial, and long-term fluctuation. The oversampling rates are relative to the base sampling rate. They were determined following rules that have been demonstrated in the past to yield near-optimal allocation.

It was assumed that screening would be used in the other strata to create a uniform probability of selection for all households containing only eligible women who were not black. Thus, for example, two-thirds of the households with only white and Asian women in old construction stratum three would be dropped from the sample after screening. Although it would be possible to use screening to sample black and other women at different rates in all the strata, it was assumed that all households containing eligible women in strata with oversampling rates of one would be retained in the sample. Screening for black women in these strata is a very expensive proposition.

Having fixed the allocation across the strata, it is then necessary to fix the overall number of sample segments. The choice of 1,718 segments was made using the following considerations. About 1,550 area segments and 150 permit segments were selected in Cycle II (8). About 18 dormitory segments were selected in Cycle III. (There were no dormitory segments in Cycle II because unmarried nulliparous women were not eligible. There were no permit segments in Cycle III because it was so soon after a decennial census. There were only one-half as many area segments in Cycle III as in Cycle II because of greater attention to small domains for which intraclass correlation is not as troubling.) Table J shows some projections of

Table H. Percent distribution and oversampling rate of eligible women by stratum, according to race

Stratum	Percent			Percent other	Oversampling rate
	Total population	Black population	Other population		
Total	100.0	100.0	100.0	...	...
Old construction:					
Less than 10 percent black	66.0	14.7	74.0	97.3	1
10–30 percent black	5.0	11.0	4.5	78.1	1.4
30–60 percent black	4.0	16.0	2.4	52.0	3
60 percent or more black	7.0	44.0	0.7	8.7	4
New construction	15.7	12.0	16.1	88.9	1
College dormitories	2.3	2.3	2.3	86.7	1
Total population in millions	57.9	7.7	50.2	...	...

Table J. Second-stage contribution to design effect for Design A

Race	Effective number of segments	Number of interviewed women	Average segment size (λ_2)	Effect of variation in segment size (ξ_2)	Composite (ρ_2)	Contribution to design effect $(\lambda_2-1)\lambda_2\xi_2+\xi_2-1$
All races	1,714	11,763	6.9	1.28	0.03	0.23
Black	809	3,180	3.9	1.28	0.03	0.11
Other than black	1,329	8,582	6.5	1.28	0.03	0.21

how many of each of these types of segments would have at least one interviewed woman by race. The table also shows the contribution to the design effect that is implied by the resulting average segment sizes.

Effects of variation in weights

The design effect resulting from the oversampling of strata with high percentages of black persons was calculated by the standard formula:

$$b^2 = (\Sigma p_i k_i)\,(\Sigma p_i / k_i) - 1 \qquad (7)$$

where p_i is the proportion of domain population in a stratum and k_i is the oversampling rate for the stratum. This formula yielded values for b^2 of 0.43 for black households and 0.15 for total households. The value of b^2 for households is 0.0 because subsampling of households other than black discovered in screening would return the other-than-black household sample to a self-weighting sample.

Table K gives the population distributions across household sizes where size is defined as the number of age-eligible women living in the house. These distributions come from Cycle III estimates (derived from table 2–7 in the National Survey of Family Growth Cycle III, final report (9). The retention rate is the proportion of eligible women who would be designated for extended interviews. Given the NSFG rule of one respondent per household, the rates are obvious. The parameter a^2 reflects the increase in design effect due to subsampling within households as before.

The total design effects are obtained by summing the components derived in this section. Table L gathers all the numbers together in the same format as table F for ease of comparison. Note that the projected effective sample sizes do not quite line up between tables F and L. It is considerably more difficult to achieve the desired precision for black women in an unlinked design than for women other than black or total women. The compromise sample size for the unlinked design yields lower precision than the linked design for black women and higher precision for women other than black and total women. The only way to have matched precision by race would have been to use screening even in the stratum with the lowest concentration of black population.

Comparing these projected design effects to other NSFG design effects in table M, it is interesting to note that the unlinked design effect for all-race estimates is projected to be larger than that for black estimates, which is the reverse of what was observed with the actual Cycle IV design. This reversal is mainly the result of the fact that with the linked design, it was possible to oversample multi-eligible households and thus reduce the a^2 to near zero for persons other than black. The design effects projected in Waksberg and Northrup for an unlinked Cycle IV are not strictly comparable because they assumed different within-household sampling rules, total sample sizes, and analysis domains. Nonetheless, they are not too far off. (The smaller design effects projected for never-married women other than black are the result of the smaller cluster sizes that are naturally obtained for smaller domains.) Also interesting to compare, the Cycle III design effects are much larger, particularly for all races. The main reason for this is the extensive and inefficient oversampling of white teenagers that was conducted in Cycle III. Also a factor was the differential sampling of ever-married and never-married women.

Projected response rates

Having calculated projected design effects, it is possible to calculate the number of women that would have to be interviewed in an unlinked design to attain precision comparable to the linked Cycle IV design. The next step is to work backward to determine the numbers of housing units that would have to be selected from area-listing worksheets.

Table K. Percent distribution of population and retention rate by number of eligible women within household, according to race

Number of eligible women within household	Total	Black	Other	Retention rate
Total	100.0	100.0	100.0	...
1	66.8	56.9	70.5	1
2	23.5	28.8	21.5	0.5
3	8.0	11.0	6.9	0.33
4 or more	1.7	3.3	1.1	0.2
a^2	0.19	0.24	0.17	...

Table L. Components of design effects for Design A, by parameter and race

Parameter	*All races*	*Black*	*All others*
P	0.73	0.61	0.75
Number of NSR PSU's	60	30	60
λ_1	143.1	64.7	107.1
ρ_1	0.005	0.005	0.005
$P\lambda_1\rho_1$	0.52	0.20	0.40
Between first-stage stratum	0.02	0.02	0.02
Number of nonempty segments	1,714	809	1,329
λ_2	6.9	3.9	6.5
ρ_2	0.03	0.03	0.03
ξ_2	1.28	1.28	1.28
$\rho_2(\lambda_2-1)\xi_2+\xi_2-1$	0.23	0.11	0.21
a^2	0.19	0.24	0.17
b^2	0.15	0.43	0.00
a^2+b^2	0.34	0.67	0.17
c^2	0.12	0.00	0.09
Modeled design effect	2.23	2.00	1.89
Number of interviewed women	11,763	3,180	8,582
Modeled effective sample size	5,276	1,590	4,541

NOTE: NSR PSU is non-self-representing primary sampling unit.

Table M. Comparison of National Survey of Family Growth design effects, by race and various designs

				Unlinked			
	Linked Cycle IV			*Waksberg-Northrup projections*			
Race	*Original*[1]	*Revised*[2]	*Modeled*[3]	*Ever-married*	*Never-married*	*New projections*	*Actual Cycle III*
All races	1.56	1.49	1.68	...	...	2.23	3.00
Black	1.90	1.80	1.70	2.00	1.98	2.02	2.83
Other than black	...	...	1.40	1.79	1.43	1.89	...

[1] As reported in Judkins, Mosher, and Botman (5).

[2] As implied by table A of this report.

[3] As constructed in table F of this report.

Experience from Cycle II has shown that about 16 percent of listed units will turn out to be vacant, built since the last decennial census, or not currently intended, ready, and fit for human habitation (7). Of the housing units that are occupied, a certain percent will resist being interviewed (either passively through not answering the door or actively by refusing). In Cycle III, 5 percent resisted the screening interview that was used to determine race and sex of occupants, and 16 percent of the remainder resisted answering the detailed questions (8).

Having fixed design effects and response rates, the entries in table N illustrate the designated number of addresses that would have to be visited, the numbers of these that could be expected to be ineligible for screening, the number of occupied eligible households, the number of screener nonrespondents, the number of successfully screened households, the number of women to be designated from these households for the extended interview, the number of actually interviewed women, and the effective sample sizes.

Table N. Sample sizes for Design A, by race and type of household or respondent

Household or respondent	*All races*	*Black*	*All others*
Designated addresses	40,800	---	---
Less vacant or uninhabitable addresses	–6,500	---	---
Households	34,300	---	---
Less screener nonrespondents	–1,700	---	---
Screened households	32,600	---	---
Number of designated women	14,000	3,800	10,200
Less nonrespondents	–2,200	–600	–1,600
Number of interviewed women	11,800	3,200	8,600
Effective sample size[1]	5,300	1,600	4,500

[1] Taking design effects into account.

Chapter 5
Cost estimates and reconciliation with earlier projections

Table O summarizes the design specifications that were developed in the previous section. These features served as the basis for the cost estimates.

Table O. Summary design specifications for Cycle IV of the National Family Growth Survey

Specification	*Design A unlinked*	*Design B linked without delays*	*Design C Cycle IV linked*	*Design D linked to full NHIS*[1] without delays
Primary sampling units	80	156	156	198
Segments	1,700	3,100	3,100	5,000
Designated households	34,300	. . .	. . .	. . .
Designated women	14,000	10,000	10,300	9,500
Interviewed women	11,800	8,450	8,450	8,000

[1] National Health Interview Survey.

Table P shows cost estimates for four possible designs for the NSFG that would approximate the same precision. Because the purpose of this cost analysis is to show a comparison of costs for the different designs, it did not seem necessary to show indirect cost markups. Therefore, all cost estimates shown consist of direct costs only. All cost estimates assume that the main study data collection would be conducted in 1992.

Pretest costs have not been included for any of the designs because it is assumed that the cost of a pretest would be the same regardless of the design chosen. Costs for data handling for all designs were calculated assuming that the NCHS computer would be used.

For Designs A and C, the response rates are those actually experienced for Cycles III and IV, respectively. The response rates shown for Designs B and D were calculated by estimating the number of women who had moved in Cycle IV but would not have moved prior to the interview had it not been for the delay; it is also assumed that these women would have cooperated at the same rate as the rest of the sample. In Cycle IV, 35 percent of women had moved between the NHIS interview and the first attempt at the NSFG interview. According to information gathered about those who moved during Cycle IV, only 30 percent would have moved by the time of interview had it not been for the 8-month delay.

Waksberg and Northrup predicted a cost savings of 28–35 percent (relative to the Cycle III model) for a linked design with one-time interviewing of designated persons (with tracking), depending upon whether 200 or 100 NHIS PSU's were used. The new report indicates a cost savings of 22 percent (table P, Design C versus Design A).

The reasons for the difference are complex, involving changes in the objectives for Cycle IV, the procedure for oversampling black women in the 1985 redesign of the NHIS, cuts in the sample size for the 1985 and 1986 panels of the NHIS, the lag between NHIS and NSFG IV interviews, and improvements in the estimation of variance components. Unavoidably, variance on the variance and cost estimates also plays some role.

Table P. Direct costs for design options in 1992 dollars

Cost item	*Design A unlinked*	*Design B linked without delays*	*Design C Cycle IV linked*	*Design D linked to full NHIS*[1] *without delays*
Total	4,082,000	2,972,400	3,192,200	2,945,400
Professional labor	565,200	483,000	510,700	483,000
Clerical labor	470,800	350,600	373,300	315,400
Field labor	1,125,200	808,300	874,400	787,300
Travel	838,800	497,100	537,600	527,900
Other direct costs	1,082,000	833,400	896,200	831,800
Number of screeners	32,600	0	0	0
Number of interviews	11,800	8,450	8,450	8,000
Estimated response rate (percent)				
Screener	95	(2)	(2)	(2)
Interview	84	84	82	84

[1] National Health Interview Survey.

[2] Screener response rate comes from NHIS.

The impact of every factor was not quantified. Some were quantified and others were simply listed with an indication of the rough order of importance.

Major questionnaire revisions late in the survey planning process, resulting in postponement of the start of data collection, were the main cause of the downward revisions in the cost savings of a linked design. If a comparison is made between Designs B and A instead of C and A, there is a cost savings for the linked design of 27.2 percent instead of 21.8 percent. NHIS sample cuts also contributed. If the 1986 NHIS had been 100 percent and if the questionnaire had been approved in a timely fashion, the savings would have been an even better 27.8 percent (comparing Designs D and A).

Another major factor was a suboptimal procedure for oversampling black women in the redesigned NHIS (1985–94). At the time that Waksberg and Northrup prepared projections of design effects for the linked design, this factor was not anticipated. Waksberg and Northrup projected a design effect for black women of around 1.5 for a 2-year NHIS sample with 100 to 200 PSU's. The actual design effect for black women (table M) was 1.8. At least some of the difference is the result of the artificial limitations placed on the NHIS oversampling procedure (10), although improvements in variance estimation methodology in the current report or variance on both sets of variance estimates may also play a role.

Finally, when Waksberg and Northrup were writing their report, the then-current objectives for Cycle IV were much more stringent than final objectives. Specific reliability targets were set for ever-married and never-married women by race (black and other). The final objectives had specific reliability targets only by race, and these targets were considerably more relaxed. Having a single reliability target by race reduces the amount of screening necessary in an unlinked design and thereby reduces the savings of a linked design. Waksberg and Northrup called for a total sample size of roughly 10,500 interviewed women to achieve the then-current reliability targets. When the linked sample was finally selected, the goal was to obtain 8,500 interviews. Waksberg and Northrup thought that their results were fairly robust with respect to overall sample size, but in retrospect, it seems likely that the smaller sample size caused some inefficiencies of scale in the 156 NHIS PSU's that would not have affected an unlinked design as severely.

Working in the opposite direction, the relaxation of the goals allowed much more efficient sampling of women other than black. Originally, there had been a plan to oversample never-married women at twice the rate of ever-married women. When this plan was dropped, Waksberg had the idea to oversample households with multiple eligible women other than black (5). This allowed a considerable reduction in the variation in the probabilities of selection for women other than black, thereby reducing the value of a^2 considerably below that projected originally. (In reviewing this report, an error in Waksberg and Northrup was detected. In their table 13, the terms a^2 and b^2 are not defined but are of magnitudes that make it clear that the first is the amount to be added to the design effect due to differential sampling within households, and the second is the amount to be added due to differential sampling across block groups with different black population densities. Yet in chapter 5 of Waksberg and Northrup, a^2 and b^2 are discussed as if they had the reversed meanings. It is suggested that a^2 be substituted for b^2 and vice versa within chapter 5. The discussion will then be consistent with the numbers in their table 16 and with usage in this report.) The reduction in a^2 was possible only with the linked design because oversampling of multi-eligible households would have necessitated doubling the screening sample—a very expensive proposition in an unlinked design. The fact that this effect is important and in the wrong direction to explain the overprojection of savings with a linked design probably puts added emphasis on the projected effective sample size for black women in a linked design as the cause for the overestimation of the cost savings with a linked design. It remains to be emphasized, however, that the cost savings were still quite substantial.

References

1. Waksberg J, Northrup DR. Integration of sample design for the National Survey of Family Growth, Cycle IV, with the National Health Interview Survey. National Center for Health Statistics. Vital Health Stat 2(96). 1985.
2. Mathiowetz N, Northrup D, Sperry S, Waksberg J. Linking the National Survey of Family Growth with the National Health Interview Survey: Analysis of field trials. National Center for Health Statistics. Vital Health Stat 2(103). 1987.
3. Massey JT, Moore TF, Parsons VL, Tadros W. Design and estimation for the National Health Interview Survey, 1985–94. National Center for Health Statistics. Vital Health Stat 2(110). 1989.
4. Wolter KM. Introduction to variance estimation. New York: Springer-Verlag. 1985.
5. Judkins DR, Mosher WM, Botman S. National Survey of Family Growth, Sample Design, Estimation, and Inference. National Center for Health Statistics. Vital Health Stat 2(109). 1991.
6. Valliant R. Generalized variance functions in stratified two-stage sampling. J Am Stat Assoc 82:499–508. 1987.
7. Krug DN, Slobasky RF, Hendriks SK, Waksberg J. National Survey of Family Growth Cycle II, final report. Rockville, Maryland: Westat, Inc. 1977.
8. Hansen MH, Hurwitz WN, Madow WG. Sample survey methods and theory. New York: John Wiley & Sons, Inc. 1962.
9. Westat. National Survey of Family Growth Cycle III, final report. Rockville, Maryland: Westat, Inc. 1984.
10. Waksberg J. Oversampling of blacks in NHIS. Memorandum for Monroe Sirken, Andrew White, and Steve Cohen. Rockville, Maryland: Westat, Inc. 1985.

Appendixes

Contents

Appendix I
Notes on cost estimation

Professional labor is:

- Highest in Design A because there is additional field staff to be supervised during data collection, there is a listing to be organized and supervised, and there are more completed questionnaires to be processed.
- Next highest in Design C because the sample must be redrawn because of delay and additional tracking must be organized and supervised.

Clerical labor is:

- Ordered proportional to the number of questionnaires to be processed, except that
- Design C has higher clerical labor costs than Design B because of tracking.

Field labor is:

- Ordered proportional to the number of questionnaires to be completed, except that
- Design C has higher field labor costs than Design B because of tracking.

Travel costs are:

- Highest for Design A because there are more interviewers to travel to training and more field work that must be supported by out-of-town travel to build response rates.
- Next highest for Design C because there are more people who have moved to be tracked and interviewed at their new locations.
- Next highest for Design D because it is less clustered than Design B (i.e., has more PSU's as well as more segments).

Other direct costs are:

- Proportional to the amount of field labor because this category includes costs for local travel, postage, telephone, and supplies.

Appendix II
Improvements to measurement of components of variance

The measurement of the components of variance could be improved for future studies. The first area for improvement concerns measurement of within-segment variance when the number of women in the segment was odd. The second concerns the measurement of the effect of variation in segment size.

Segments with just one interviewed woman each should be dropped from the file when calculating within-segment variance. Of course, the balance of the sample would have to be reweighted to get the correct totals. The resulting within-segment variances would then have to be adjusted to compensate for the fact that the sample was larger than it appeared and that the weights were less variable. Also, special treatment should be given to the segments with three women each. For example, when half of a segment to be dropped consists of one woman, the remaining two should have their weights perturbed upward by 50 percent instead of 100 percent; similarly, when the half consists of two women, the remaining one should have her weight perturbed upward by 200 percent instead of 100 percent.

Measurement of the effect of variation in segment size on between-segment variance for substantive characteristics is more difficult. It would require the calculation of an additional set of replicate weights. This set would represent the between-segment variance if all the segments were equal in size (number of interviewed women). Variances could be calculated with this set of replicate weights and generalized. Label the new b parameter b_{WPE} for within-PSU equal-sized segments. Formula (5) would remain unchanged. However, formula (6) would be replaced by

$$\rho_2 = (b_{\mathrm{WPE}}/b_{\mathrm{WS}}-1)/\,(\lambda_2-1) \qquad (8)$$

and a new formula would be available to estimate ζ_2:

$$\xi_2 = b_{\mathrm{WP}}/b_{\mathrm{WPE}} \qquad (9)$$

Deriving the replicate weights for b_{WPE} would be fairly complicated. For simplicity of explanation, assume that two segments have been paired and that one contains two interviewed women and the other, three interviewed women. When the segment containing two women is dropped, the weights for the three women in the other segment would have their weights perturbed upward by 67 percent instead of 100 percent. When the segment containing three women is dropped, the weights for the two women in the other segment would have their weights perturbed upward by 150 percent instead of 100 percent.

*U.S. Government Printing Office: 1993 — 342-327/80001

Vital and Health Statistics series descriptions

SERIES 1. **Programs and Collection Procedures**—These reports describe the data collection programs of the National Center for Health Statistics. They include descriptions of the methods used to collect and process the data, definitions, and other material necessary for understanding the data.

SERIES 2. **Data Evaluation and Methods Research**—These reports are studies of new statistical methods and include analytical techniques, objective evaluations of reliability of collected data, and contributions to statistical theory. These studies also include experimental tests of new survey methods and comparisons of U.S. methodology with those of other countries.

SERIES 3. **Analytical and Epidemiological Studies**—These reports present analytical or interpretive studies based on vital and health statistics. These reports carry the analyses further than the expository types of reports in the other series.

SERIES 4. **Documents and Committee Reports**—These are 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 5. **International Vital and Health Statistics Reports**—These reports are analytical or descriptive reports that compare U.S. vital and health statistics with those of other countries or present other international data of relevance to the health statistics system of the United States.

SERIES 6. **Cognition and Survey Measurement**—These reports are from the National Laboratory for Collaborative Research in Cognition and Survey Measurement. They use methods of cognitive science to design, evaluate, and test survey instruments.

SERIES 10. **Data From the National Health Interview Survey**—These reports contain statistics on illness; unintentional injuries; disability; use of hospital, medical, and other health services; and a wide range of special current health topics covering many aspects of health behaviors, health status, and health care utilization. They are based on data collected in a continuing national household interview survey.

SERIES 11. **Data From the National Health Examination Survey, the National Health and Nutrition Examination Surveys, and the Hispanic Health and Nutrition Examination Survey**—Data from direct examination, testing, and measurement on representative samples of the civilian noninstitutionalized population provide the basis for (1) medically defined total prevalence of specific diseases or conditions in the United States and the distributions of the population with respect to physical, physiological, and psychological characteristics, and (2) analyses of trends and relationships among various measurements and between survey periods.

SERIES 12. **Data From the Institutionalized Population Surveys**—Discontinued in 1975. Reports from these surveys are included in Series 13.

SERIES 13. **Data From the National Health Care Survey**—These reports contain statistics on health resources and the public's use of health care resources including ambulatory, hospital, and long-term care services based on data collected directly from health care providers and provider records.

SERIES 14. **Data on Health Resources: Manpower and Facilities**—Discontinued in 1990. Reports on the numbers, geographic distribution, and characteristics of health resources are now included in Series 13.

SERIES 15. **Data From Special Surveys**—These reports contain statistics on health and health-related topics collected in special surveys that are not part of the continuing data systems of the National Center for Health Statistics.

SERIES 16. **Compilations of Advance Data From Vital and Health Statistics**—Advance Data Reports provide early release of information from the National Center for Health Statistics' health and demographic surveys. They are compiled in the order in which they are published. Some of these releases may be followed by detailed reports in Series 10–13.

SERIES 20. **Data on Mortality**—These reports contain statistics on mortality that are not included in regular, annual, or monthly reports. Special analyses by cause of death, age, other demographic variables, and geographic and trend analyses are included

SERIES 21. **Data on Natality, Marriage, and Divorce**—These reports contain statistics on natality, marriage, and divorce that are not included in regular, annual, or monthly reports. Special analyses by health and demographic variables and geographic and trend analyses are included.

SERIES 22. **Data From the National Mortality and Natality Surveys**—Discontinued in 1975. Reports from these sample surveys, based on vital records, are now published in Series 20 or 21.

SERIES 23. **Data From the National Survey of Family Growth**—These reports contain statistics on factors that affect birth rates, including contraception, infertility, cohabitation, marriage, divorce, and remarriage; adoption; use of medical care for family planning and infertility, and related maternal and infant health topics. These statistics are based on national surveys of childbearing age.

SERIES 24. **Compilations of Data on Natality, Mortality, Marriage, Divorce, and Induced Terminations of Pregnancy**—These include advance reports of births, deaths, marriages, and divorces based on final data from the National Vital Statistics System that were published as supplements to the *Monthly Vital Statistics Report* (MVSR). These reports provide highlights and summaries of detailed data subsequently published in *Vital Statistics of the United States*. Other supplements to the MVSR published here provide selected findings based on final data from the National Vital Statistics System and may be followed by detailed reports in Series 20 or 21.

For answers to questions about this report or for a list of reports published in these series, contact:

Scientific and Technical Information Branch
National Center for Health Statistics
Centers for Disease Control and Prevention
Public Health Service
6525 Belcrest Road, Room 1064
Hyattsville, MD 20782

(301) 436–8500

DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service
Centers for Disease Control and Prevention
National Center for Health Statistics
6525 Belcrest Road
Hyattsville, Maryland 20782

DHHS Publication No. (PHS) 93-1391, Series 2, No. 117