DATA EVALUATION AND METHODS RESEARCH

Series 2
Number 82

Small Area Estimation:

An Empirical Comparison of Conventional and Synthetic Estimators for States

This report describes the results of a large empirical investigation into the appropriateness of several competing methods for making estimates for States from the National Health Interview Survey. Because no method is superior, a strategy that should improve existing methodology is suggested.

DHEW Publication No. (PHS) 80-1356

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Public Health Service
Office of Health Research, Statistics, and Technology
National Center for Health Statistics
Hyattsville, Md. December 1979

Library of Congress Cataloging in Publication Data

United States. National Center for Health Statistics. Small area estimation.

(Vital and health statistics : Series 2, Data evaluation and methods research 2 ; no. 82) (DHEW publication ; no. (PHS) 80-1356)

Bibliography: p. 9.

1. Health surveys–Statistical methods. 2. Estimation theory. 3. Health surveys–United States–Statistical methods. I. Schaible, Wesley L. II. Title. III. Series: United States. National Center for Health Statistics. Vital and health statistics : Series 2, Data evaluation and methods research ; no. 82. IV. Series: United States. Dept. of Health, Education, and Welfare. DHEW publication ; no. (PHS) 80-1356.

RA409.U45 no. 82 312'.07'23s [614.4'2'0182] 79-22509

ISBN 0-8406-0176-X

NATIONAL CENTER FOR HEALTH STATISTICS

Vital and Health Statistics-Series 2-No. 82

DHEW Publication No. (PHS) 80-1356
Library of Congress Catalog Card Number 79-22509

PREFACE

This report presents the results of a large empirical investigation of several competing techniques for making small area estimates from the Health Interview Survey of the National Center for Health Statistics. The project was conducted by the staff of the Statistical Research Branch of the Office of Statistical Research, under the direction of Dr. Wesley L. Schaible, formerly the acting chief of the Statistical Research Branch.

In addition to internal reviews (by Dr. Paul E. Leaverton, Office of Statistical Research, and Mr. Dwight K. French, Office of Statistical Research), National Center for Health Statistics policy requires that methodological reports undergo an external review for technical merit and clarity. Dr. Steven Cohen of the National Center for Health Services Research performed this review and provided numerous constructive comments on an early draft of this report. Finally, the authors acknowledge the outstanding work of Mr. Barry W. Peyton of the Statistical Research Branch in preparing the computation for this project and Mr. Eugene Diggs for computer graphics.

CONTENTS

LIST OF TEXT FIGURES

LIST OF TEXT TABLES

SYMBOLS

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

SMALL AREA ESTIMATION: AN EMPIRICAL COMPARISON OF CONVENTIONAL AND SYNTHETIC ESTIMATORS FOR STATES

Wesley L. Schaible, Ph.D.,[a] Dwight B. Brock, Ph.D.,[b] Robert J. Casady, Ph.D.,[c] and George A. Schnack[b]

INTRODUCTION

Most large samples, for example, those of the Current Population Survey and Health Interview Survey, were originally designed to give national and regional estimates. However, estimates also are needed for States, Health Service Areas, counties, and other small areas. One way to meet this demand is to redesign these surveys, but this process can be both expensive and time consuming. If operational considerations make redesign difficult, then the immediate question becomes "How and under what conditions can reasonably accurate estimates for small areas be obtained from a large survey designed for national estimates?"

One approach, synthetic estimation, has received considerable attention. It was introduced in 1968 in a National Center for Health Statistics publication *Synthetic State Estimates of Disability*.[1] The authors stated that the sample size (and design) of the Health Interview Survey were inadequate to make direct State estimates by conventional procedures and suggested a synthetic approach that has since been extensively investigated. Levy[2] used mortality data to compute average relative errors of synthetic estimates for States. Gonzalez and Waksberg[3] calculated mean square errors averaged over all small areas and compared synthetic and direct estimates for selected standard metropolitan statistical areas. Gonzalez and Hoza[4] investigated synthetic-estimate errors by using unemployment data for counties from the CPS and the 1970 census. Namekata, Levy, and O'Rourke[5] investigated synthetic State estimates of work-loss disability in a similar manner. Schaible, Brock, and Schnack[6] compared the average squared errors of synthetic and direct estimates of unemployment rates for county groups in Texas. Levy and French[7] discussed properties of the nearly unbiased, synthetic, and regression-adjusted synthetic estimators and compared different synthetic estimators. Finally, Purcell and Kish[8] have given an excellent summary of the most recent literature on small area estimation.

If information from a national sample is used to make estimates for small areas and there are no sample units in the small area of interest, then, obviously, conventional estimation methods cannot be used and a synthetic approach is necessary. However, the sample size in a small area can be rather large; for example, the Health Interview Survey sample size in California is greater than 10,000 persons each year. Therefore, when estimating for areas with large sample sizes, should one ignore traditional procedures and use a synthetic approach? When the sample size of a small area approaches the size of the area's population, a conventional direct estimator becomes more desirable than a synthetic estimator. This statement is true whether

[a]U.S. Bureau of Labor Statistics.
[b]Office of Statistical Research, National Center for Health Statistics.
[c]Office of Mathematical Statistics, National Center for Health Statistics.

or not the sample was designed to produce estimates for small areas.

The purpose of this report is to compare a variety of direct and synthetic estimators for making State estimates and to investigate the relative performance of these estimators over different State sample sizes.

APPROACH

The appropriateness of an estimation procedure often depends on the sample design. Thus a brief description of the Health Interview Survey (HIS) design used from 1969 through 1971 is given below. A more complete description is given in a separate National Center for Health Statistics (NCHS) publication.[9] This HIS sample design is a multistage probability design which permits a continuous sampling of households from the civilian noninstitutionalized population of the United States. The first stage consists of a sample of 357 primary sampling units (PSU's) drawn from approximately 1,900 geographically defined PSU's that cover the 50 States and the District of Columbia. A PSU consists of a county, a small group of contiguous counties, or a standard metropolitan statistical area. Within PSU's smaller units called segments, each containing an expected six households, are selected. The usual HIS sample consists of approximately 8,000 segments, which yield a probability sample of about 134,000 persons in 42,000 households interviewed in a year. The number of sample persons residing in each State and the District of Columbia in 1970 and 1969-71 is given in table III.

The usual HIS estimation procedure is elaborate. Each responding sample person is assigned an estimation weight that is the product of the reciprocal of the probability of selection, two nonresponse factors, and two poststratification factors. A weighted estimate of a population aggregate is then made by weighting the observation of interest for each responding sample person by the corresponding estimation weight and then summing over all persons.

For the purposes of this study, a variety of direct and synthetic estimators are used with HIS sample data to estimate known population values for each State and the District of Columbia. This procedure allows calculation of actual errors to compare the estimators' performance. The known population values were obtained from data taken from the 5-percent sample questionnaire of the 1970 census. The actual data from the Public Use Sample Tapes contain a one-in-one-hundred sample of the total U.S. population. The variances of "estimates" from a sample this large are negligible, and the quantities computed from these tapes are treated as population values. Comparable variables were selected from the HIS, and estimates were calculated from the 1970 and 1969-71 data. The variables studied are (1) percent of the population less than 1 year old, (2) percent of the population married, (3) percent of the population separated, (4) percent of the population completing high school, and (5) percent of the population completing college. Hereafter, these variables will be referred to as "less than one," "married," "separated," "high school," and "college," respectively.

Sixteen estimates were calculated for each variable: four basic and four ratio adjusted, and each estimate was calculated with and without the HIS estimation weights. The four basic estimators are (1) the simple direct, (2) a poststratified, (3) a 16-cell synthetic, and (4) a 64-cell synthetic. The four basic estimators and an explanation of the ratio adjustments are given in the following section.

ESTIMATORS

Let $y_{ij\alpha}$ denote the observation of interest on the i^{th} unit in the j^{th} State in the α^{th} poststratification cell (or demographic class):

$$i = 1, 2, \ldots, n_{j\alpha}$$

$$j = 1, 2, \ldots, 51$$

$$\alpha = 1, 2, \ldots, k$$

where $n_{j\alpha}$ is the HIS sample size in the $j\alpha^{\text{th}}$ cell, and let $N_{j\alpha}$ denote the population size in the $j\alpha^{\text{th}}$ cell. NOTE: When $n_{j\alpha} = 0$, we define

$$\sum_{i=1}^{0} y_{ij\alpha} = 0$$

The usual dot summation convention will be used here; for example,

$$\sum_{\alpha=1}^{k} n_{j\alpha} = n_{j.}$$

Simple Direct Estimator

The simple direct estimator for the j^{th} State is

$$\bar{y}_{j.} = \sum_{\alpha=1}^{k} \sum_{i=1}^{n_{j\alpha}} y_{ij\alpha}/n_{j.}$$

Poststratified Estimator

The sample mean of the α^{th} poststratification cell in the j^{th} State is

$$\bar{y}_{j\alpha} = \sum_{i=1}^{n_{j\alpha}} y_{ij\alpha}/n_{j\alpha}$$

Thus the usual poststratified estimator for State j is

$$\bar{y}'_{j.} = \sum_{\alpha=1}^{k} N_{j\alpha}\bar{y}_{j\alpha}/N_{j.}$$

The variables used to define the 16 poststratification cells were age, sex, and color, as described for the synthetic estimator.

Synthetic Estimators

The sample mean of the α^{th} demographic class (poststratification cell) for the total U.S. is

$$\bar{y}_{.\alpha} = \sum_{j=1}^{51} \sum_{i=1}^{n_{j\alpha}} y_{ij\alpha}/n_{.\alpha}$$

The simple synthetic estimator for State j is

$$\bar{y}''_{j.} = \sum_{\alpha=1}^{k} N_{j\alpha}\bar{y}_{.\alpha}/N_{j.}$$

The variables that define the demographic classes are as follows:

1. Color: white, all other.
2. Sex: male, female.
3. Age group: under 17 years, 17-44 years, 45-64 years, 65 years and over
4. Family size: fewer than five members, five members or more.
5. Industry of head of family: Standard International Classifications: (1) forestry and fisheries, agriculture, construction, mining and manufacturing; (2) and all other industries.

The 64 classes produced by these variables were used in the 64-cell synthetic estimator. The 16 classes defined by the age, sex, and color groups were used for the 16-cell synthetic estimator.

Ratio-Adjusted Estimators

An additional four estimators were created by modifying each basic estimator by a regional ratio adjustment. The ratio adjustment for the simple direct estimator ($\bar{y}_{j.}$) illustrates the adjustment procedure.

Let

$$\bar{y}'_R = \sum_{j\epsilon R} N_{j.}\bar{y}_{j.} / \sum_{j\epsilon R} N_{j.}$$

and let $\bar{y}_R$ be the usual weighted HIS estimator, where R denotes the HIS region that includes the State of interest. The ratio-adjusted simple direct estimator for the j^{th} State is then:

$$\bar{y}'''_{j.} = \bar{y}_{j.} \frac{\bar{y}_R}{\bar{y}'_R}$$

The ratio adjustment for each of the remaining estimators is the same except in the denominator of the ratio where the simple direct estimators are replaced by the particular estimators being adjusted. This ratio adjustment forces the weighted sum of the State estimates, when each State estimate is weighted by the proportion of the regional population in that State, to be consistent with the usual HIS estimate for each region.

RESULTS

The average squared difference between each estimate and its corresponding State census value, that is, the average squared error, is shown for the 5 variables and 16 estimators in table A. The estimates used to compute these errors were based on the 1969-71 HIS sample data. The average squared error obtained when the HIS regional estimate is used for each State is also shown in this table. This average squared error of the regional estimator is an indicator of the variability of the State values within regions. This indicator is important because synthetic estimators tend to perform better when the estimated characteristic does not vary substantially among small areas. For example, the variable "married" shows little variability among small areas, and has a regional average squared error of 5.71 (see table A). On the other hand "high school," which varies considerably more from State to State, has a regional average squared error of 13.43. This average squared error for the regional estimator also can be used for comparison with the mean square errors of other estimators. A minimum expectation of a State estimator is to outperform this simplistic method of estimation.

Relatively large differences in average squared errors occur among the eight direct and eight synthetic estimators. This finding is partially explained by the size of the statistic. The synthetic estimators have smaller average squared errors than direct estimators when they are used to estimate "less than one" *because this variable differs little between the States within a region.* The average squared error of the regional estimate is 0.04, and in 1970 this percent ranged from 1.5 in Rhode Island to 2.7 in Alaska. For the variables "married" and "separated," the direct estimators have smaller average squared errors than do the synthetic. For the remaining two variables neither group of estimators has consistently smaller average squared errors.

Although major differences in average squared errors occur among the direct and syn-

Table A. Average squared errors of estimates for 50 States and the District of Columbia, by selected attribute variables and small area estimators: Health Interview Survey, 1969-71

Estimator	Variable				
	Percent less than 1 year old	Percent married	Percent separated	Percent completing high school	Percent completing college
	Average squared error				
Simple direct	0.16	1.81	0.05	12.36	1.81
Weighted	0.15	2.27	0.05	12.59	1.86
Ratio adjusted	0.17	1.94	0.05	12.72	1.84
Weighted, ratio adjusted	0.16	2.23	0.05	12.48	1.86
Poststratified	0.14	2.41	0.05	7.45	1.70
Weighted	0.14	2.34	0.05	6.83	1.72
Ratio adjusted	0.16	2.06	0.05	7.95	1.71
Weighted, ratio adjusted	0.15	2.26	0.05	7.18	1.74
Synthetic (16)	0.02	3.81	0.31	19.54	2.32
Weighted	0.01	3.81	0.31	19.51	2.33
Ratio adjusted	0.02	3.23	0.23	11.37	2.34
Weighted, ratio adjusted	0.02	3.21	0.23	11.27	2.32
Synthetic (64)	0.02	3.83	0.30	16.42	1.68
Weighted	0.02	3.81	0.30	16.42	1.69
Ratio adjusted	0.02	3.21	0.22	8.94	1.78
Weighted, ratio adjusted	0.02	3.19	0.22	8.89	1.76
Regional estimate	0.04	5.71	0.49	13.43	1.95

thetic estimators, some within-group differences also exist. The synthetic estimators that have the regional ratio adjustment generally produce smaller average squared errors than those that do not. The synthetic estimators that use the HIS estimation weights have average squared errors almost identical to those that do not. The use of the ratio adjustment with the direct estimators tends to increase the average squared error rather than decrease it. Furthermore, the use of estimation weights does not improve the performance of the direct estimators. In fact, when the 1970 data were used to produce estimates, the addition of estimation weights increased the average squared errors of the direct estimators (table I). In general, these results indicate that if direct estimators are used for producing State estimates, perhaps they should not be weighted or ratio adjusted and if synthetic estimators are used, they should be ratio adjusted. As a result of this evidence four estimators were chosen for further investigation: the simple direct and poststratified estimators, neither weighted nor ratio adjusted, and the 16- and 64-cell synthetic estimators, both weighted and ratio adjusted. The weighted synthetic estimators were chosen instead of the unweighted estimators because the weighted α-cell means used are generally more readily available than the unweighted means.

The plots of State estimates and State census values in figures 1-4 show the results when the two direct and two synthetic estimators were used to estimate the percent completing high school. Plots of results obtained by using these four estimators for each of the four remaining variables are presented in figures I-XVI. In each plot the census value is shown on the horizontal axis and the estimate on the vertical axis. Each State (and the District of Columbia) is represented by a point. The error in an estimate for a State is the vertical distance from the point for that State to the 45° line bisecting the plot.

All four estimators generally produce estimates that approximate the State census values. The 16- and 64-cell synthetic estimators (figures 3 and 4) both tend to overestimate State values that are low and underestimate those that are high. In the estimates produced by both synthetic estimators the largest error occurs in the

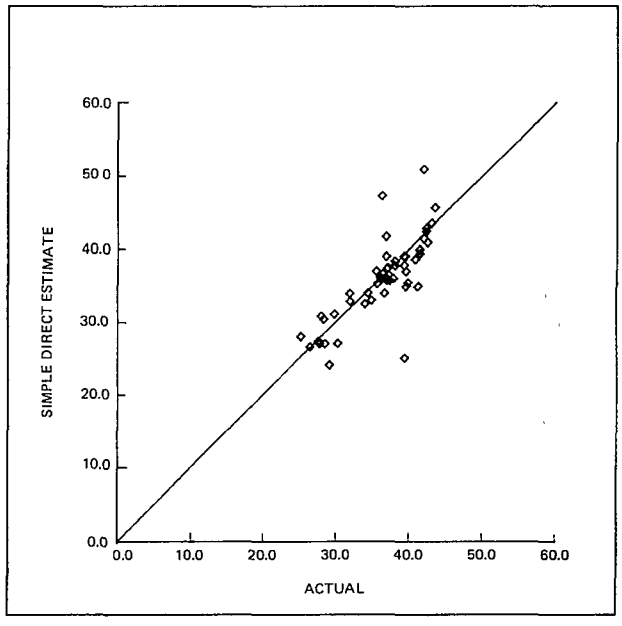

Figure 1. Percent of the population who have completed high school—simple direct estimates and actual values for 50 States and the District of Columbia: Health Interview Survey, 1969-71

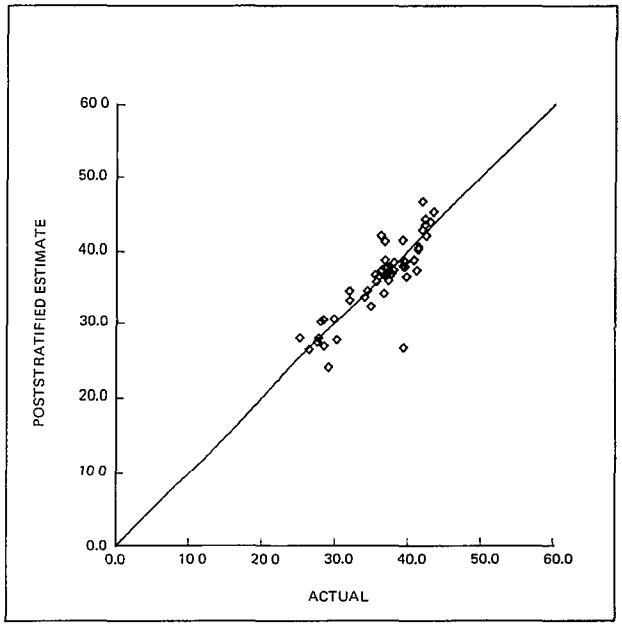

Figure 2. Percent of the population who have completed high school—poststratified estimates and actual values for 50 States and the District of Columbia: Health Interview Survey, 1969-71

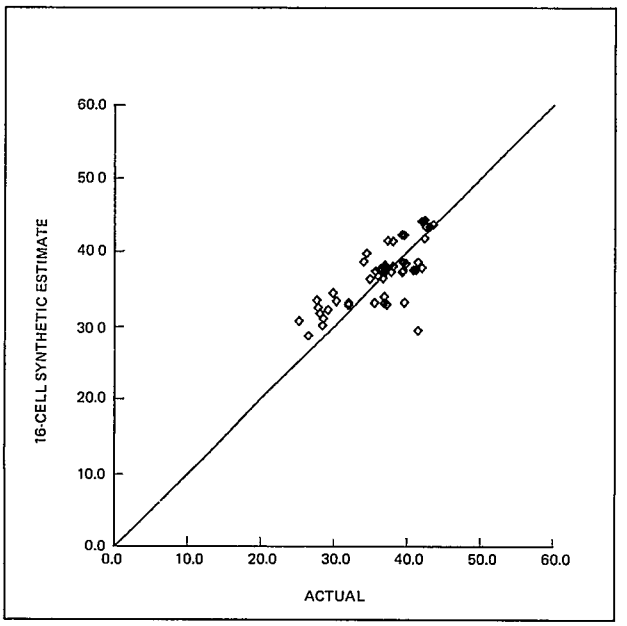

Figure 3. Percent of the population who have completed high school—16-cell synthetic estimates and actual values for 50 States and the District of Columbia: Health Interview Survey, 1969-71

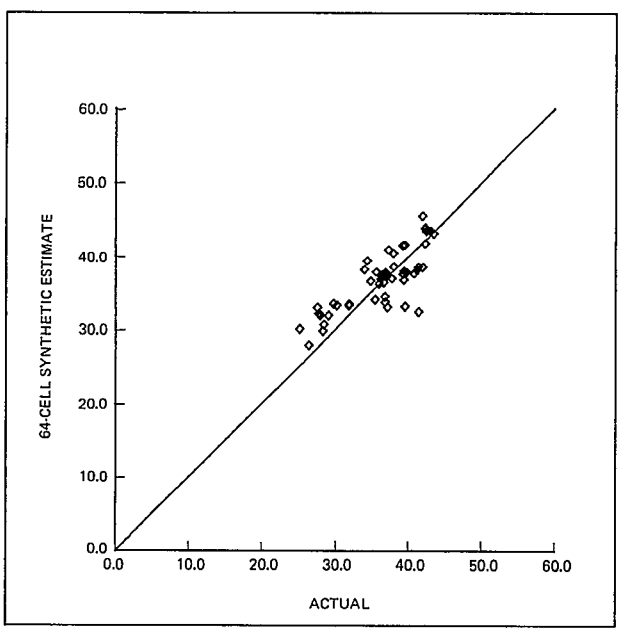

Figure 4. Percent of the population who have completed high school —64-cell synthetic estimates and actual values for 50 States and the District of Columbia: Health Interview Survey, 1969-71

estimate for the District of Columbia and the next largest error occurs in the estimate for Hawaii. These observations are generally true for the other four variables. In fact, the synthetic estimates for the District of Columbia and Hawaii are in error to the extent that, except for the variable "less than one," they dominate the average squared errors of the synthetic estimators presented in table A. If these two errors are removed from the average squared error calculations, the synthetic estimators generally have smaller average squared errors than the direct estimators. For example, excluding the District of Columbia and Hawaii reduces the average squared error of the 64-cell ratio adjusted synthetic estimator from 3.19 to 1.08 for "married," from 0.22 to 0.08 for "separated," from 8.89 to 6.72 for "high school," and from 1.76 to 1.15 for "college." If these results indicate what to expect when estimating other characteristics, then special care should be taken when interpreting synthetic estimates for these two locations.

In the simple direct estimates in figure 1, the three largest errors occur for States with the smallest sample sizes. One large error for the poststratified estimates in figure 2 occurs in Vermont, the State with the second smallest sample size. Figures 5 and 6 show the squared errors of the simple direct and poststratified estimators for each State by the combined 1969-71 HIS State sample size. Similar graphs for these two estimators and the remaining variables are shown in figures XVII-XXIV. As expected with any conventional estimator, the States with large sample sizes generally have smaller errors than the States with small sample sizes. This result is true for both estimators regardless of the variable considered.

The fitted curves were generated by the model, $y_i = A + B/n_i + e_i$, where y_i is the squared error of the estimate for State i, n_i is the sample size, e_i is a random error term, and A and B are unknown parameters. One imposed restriction was that A must equal zero when the sample size equals the average State population size. The parameters A and B were estimated from the data by minimizing the sum of the absolute values of the e_i's (table II).

Because the expected errors of the simple direct and poststratified estimators decrease as State sample size increases, it is interesting to ascertain at what sample size(s) these errors will become smaller than those expected from a synthetic estimator. The preceding model fits the squared errors of the direct estimates reasonably well; however, some difficulty arises in finding a model that adequately describes the squared errors of the State synthetic estimates as a function of State sample size. This difficulty is primarily due to the fact that, as noted by Schaible, Brock, and Schnack,[6] the squared error of the synthetic estimator is subject only to a small sampling variance inherent in the estimated large area mean but is usually dominated by a bias component which is independent of sample size.

Thus the squared error of the State synthetic estimates is described by a constant function, specifically the squared error averaged over States. However, in the variables investigated this model tended to overestimate the squared errors of those States with larger sample sizes. This overestimation was due to three interrelated factors. First, the State synthetic estimates have a smaller bias component when the State cell values approach the regional or national cell values. Second, the contribution of a State cell value to a regional or national value is in direct proportion to the size of the State's cell population, so that States with large populations tend to have actual cell values near the actual regional or national cell values. Third, the HIS sample is designed so that States with large populations have large sample sizes.

With the assumption that a more accurate function can be approximated by this simple one, the average squared errors (omitting the District of Columbia and Hawaii) of the 64-cell ratio adjusted synthetic estimator can be compared with the appropriate curves in figures 5 and 6 and figures XVII-XXIV. The sample sizes at which the two expected squared error functions intersect are approximately 8,000 for the variable "less than one," 300 for the variable "separated," and approximately 2,000 for the three remaining variables. Many State sample sizes are large compared with the values where

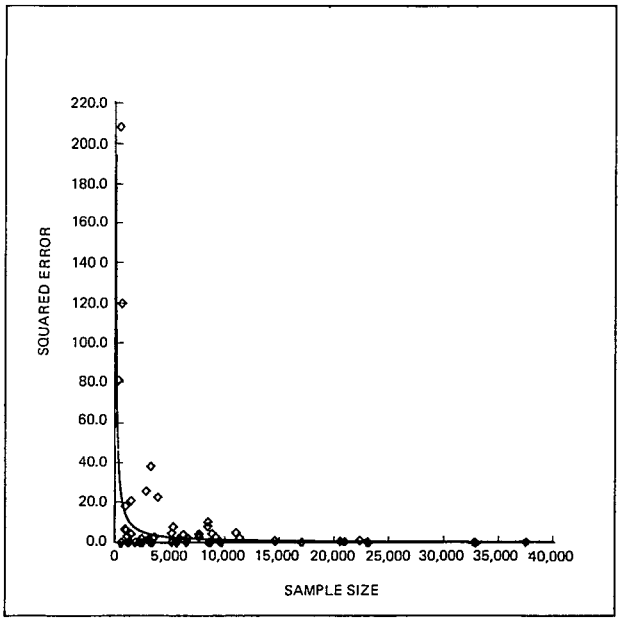

Figure 5. Squared errors of simple direct estimates of the percent of State population who completed high school, by sample size: United States, 1969-71

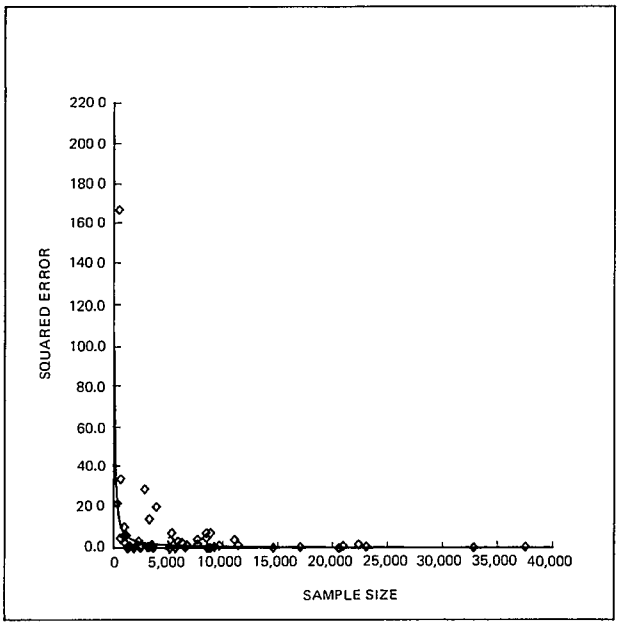

Figure 6. Squared errors of poststratified estimates of the percent of State population who completed high school, by sample size: United States, 1969-71

these functions intersect (table III). In fact, the average State sample size for the combined 1969-71 sample is 7,500. This observation suggests that in many States the direct estimators are expected to perform, as well as, if not slightly better than, the synthetic estimators.

Although the average squared error is used to evaluate synthetic estimators, this measure is not appropriate for all purposes. Generally, the average squared error is a measure of how well the estimate agrees with the parameter being estimated. However, in some cases, rather than estimating only one parameter, the difference between the parameters of two domains is estimated. In other cases, information is provided on the relative position of the parameters from several domains. When comparing estimators, the one with the smallest average squared error might be expected to perform best in estimating both relative positions and the level of a single parameter. This expectation is not necessarily justified. Table I shows that for the variable "college" the simple direct estimator has an average squared error of 3.50, which is much larger than the 1.72 of the 64-cell ratio adjusted synthetic estimator. However, the simple direct estimator for this variable is more highly correlated with the actual values than the synthetic estimator is (table B). In fact, the direct estimators correlate better with census values than the synthetic estimators do for all the variables except "less than one."

SUMMARY

Sixteen different estimators, eight direct and eight synthetic, were used with HIS data to estimate five different known census values for each of the 50 States and the District of Columbia. The effect of several different estimation techniques on the average squared errors of these estimators was noted. Estimators that used the HIS estimation weight designed for national estimates did not outperform unweighted estimators, in fact, in some instances the use of HIS estimation weights increased the average squared errors of direct estimators. A ratio adjustment to regional HIS estimates improved synthetic estimators but did not improve direct estimators. A 64-cell synthetic estimator outperformed a 16-cell synthetic estimator for two of the five variables considered. The average squared errors of these two estimators were essentially identical for the remaining three variables. A poststratified estimator generally produced smaller average squared errors than did the simple direct estimator. Moreover, as expected, the squared errors of the direct estimators decreased as the sample sizes in the States increased.

Although the preceding differences were noticeable, the major differences in the 16 estimators occurred between the direct and synthetic estimators. The direct estimators had smaller average squared errors with some vari-

Table B. Correlation coefficients between actual and estimated State values for 2 direct and 2 synthetic estimators, by selected attribute variables: Health Interview Survey, 1970 and 1969-71

Data years and estimator	Variable				
	Percent less than 1 year old	Percent married	Percent separated	Percent completing high school	Percent completing college
1970	Correlation coefficient				
Simple direct	0.36	0.77	0.93	0.65	0.55
Poststratified	0.36	0.72	0.94	0.78	0.60
Synthetic (16)—weighted, ratio adjusted	0.71	0.68	0.81	0.72	0.25
Synthetic (64)—weighted, ratio adjusted	0.69	0.68	0.82	0.78	0.42
1969-71					
Simple direct	0.44	0.88	0.96	0.79	0.69
Poststratified	0.40	0.86	0.96	0.86	0.71
Synthetic (16)—weighted, ratio adjusted	0.79	0.67	0.80	0.74	0.27
Synthetic (64)—weighted, ratio adjusted	0.76	0.67	0.81	0.79	0.45

ables, and the synthetic estimators had smaller average squared errors with others. However, the synthetic estimators produced estimates for the District of Columbia and Hawaii with such unusually large errors that, when these two places were omitted from the comparisons, the synthetic estimators generally had smaller average squared errors than the direct estimators.

When the correlation between an estimate and its corresponding census value was used as an evaluation criterion instead of the average squared error, the results were different because the direct estimators generally outperformed the synthetic estimators. These results suggest that the direct estimators might serve better when estimating differences in characteristics between States, or, similarly, when determining relative positions among States.

These considerations and previous studies[1,10] indicate that the selection of a single estimator to produce State estimates from HIS may not be the best choice. A superior estimator might be obtained by using a linear combination of estimators where the components are weighted according to their expected performance. Recently, these types of estimators, called composite estimators, were studied theoretically[11,12] and, on a limited basis, empirically.[13,14] A more extensive empirical investigation will be the subject of future research.

REFERENCES

[1] National Center for Health Statistics: *Synthetic State Estimates of Disability*. PHS Pub. No. 1759. Public Health Service. Washington. U.S. Government Printing Office, 1968.

[2] Levy, P. S.: The use of mortality data in evaluating synthetic estimates, in *Proceedings of the American Statistical Association 1971, Social Statistics Section.* Washington. American Statistical Association, 1974. pp. 328-331.

[3] Gonzalez, M. E., and Waksberg, J. E.: Estimation of the Error of Synthetic Estimates. Presented at the first meeting of the International Association of Survey Statisticians, Vienna, Austria, Aug. 1973.

[4] Gonzalez, M. E., and Hoza, C.: Small area estimation with application to unemployment and housing estimates. *J. Am. Stat. Assoc.* 73:7-15, 1978.

[5] Namekata, T., Levy, P. S., and O'Rourke, T. W.: Synthetic estimates of work loss disability for each State and the District of Columbia. *Pub. Health Rep.* 90:532-538, 1975.

[6] Schaible, W. L., Brock, D. B., and Schnack, G. A.: An Empirical Comparison of Two Estimators for Small Areas. Presented at the Second Annual Data Use Conference of the National Center for Health Statistics, Dallas, Tex., 1977.

[7] National Center for Health Statistics: Synthetic estimation of State health characteristics based on the Health Interview Survey, by P. S. Levy and D. K. French. *Vital and Health Statistics.* Series 2-No. 75. DHEW Pub. No. (HRA) 78-1349. Health Resources Administration. Washington. U.S. Government Printing Office, Oct. 1977.

[8] Purcell, N. J., and Kish, L.: Estimation for small domains. *Biometrics* 35:365-384, 1979.

[9] National Center for Health Statistics: Statistical design of the Health Household Interview Survey. *Health Statistics.* Series A-2. PHS Pub. No. 584-A2. Public Health Service. Washington. U.S. Government Printing Office, 1958.

[10] Royall, R. M.: Discussion of two papers on recent developments in estimation for local areas, in *Proceedings of the American Statistical Association 1973, Social Statistics Section.* Washington. American Statistical Association, 1974. pp. 43-44.

[11] Royall, R. M.: Prediction Models in Small Area Estimation. Presented at the NIDA-NCHS Workshop on Synthetic Estimates, Princeton, N.J., 1978.

[12] Schaible, W. L.: Choosing weights for composite estimators for small area statistics, to appear in *Proceedings of the American Statistical Association 1978, Survey Research Section.* Washington. American Statistical Association, 1979. pp. 741-746.

[13] Schaible, W. L., Brock, D. B., and Schnack, G. A.: An empirical comparison of the simple inflation, synthetic and composite estimators for small area statistics, in *Proceedings of the American Statistical Association 1977, Social Statistics Section.* Washington. American Statistical Association, 1978. pp. 1017-1021.

[14] Brock, D. B., and Peyton, B. W.: Small Area Estimation: An Application of Three Methods to the U.S. National Health Interview Survey. Presented at the 36th Annual Meeting of the United States–Mexico Border Health Association, Reynosa, Tamaulipas, Mexico, 1978.

APPENDIX

CONTENTS

LIST OF APPENDIX TABLES

LIST OF APPENDIX FIGURES

APPENDIX

SUPPLEMENTAL TABLES AND FIGURES

Table I. Average squared errors of estimates for 50 States and the District of Columbia, by selected attribute variables and small area estimators: Health Interview Survey, 1970

Estimator	Variable				
	Percent less than 1 year old	Percent married	Percent separated	Percent completing high school	Percent completing college
	Average squared error				
Simple direct	0.52	5.21	0.11	30.87	3.50
Weighted	0.41	6.73	0.16	33.68	3.50
Ratio adjusted	0.51	5.74	0.11	32.20	3.55
Weighted, ratio adjusted	0.45	6.56	0.16	32.64	3.49
Poststratified	0.46	5.63	0.09	13.84	3.00
Weighted	0.36	6.20	0.12	12.94	3.11
Ratio adjusted	0.48	4.80	0.09	14.90	3.05
Weighted, ratio adjusted	0.38	5.88	0.11	13.80	3.22
Synthetic (16)	0.02	3.83	0.31	20.22	2.39
Weighted	0.01	3.76	0.31	20.25	2.43
Ratio adjusted	0.03	3.14	0.23	13.52	2.47
Weighted, ratio adjusted	0.02	3.08	0.22	13.50	2.47
Synthetic (64)	0.02	3.84	0.31	17.19	1.72
Weighted	0.02	3.75	0.30	17.12	1.75
Ratio adjusted	0.02	3.14	0.22	10.97	1.89
Weighted, ratio adjusted	0.02	3.06	0.22	10.96	1.90
Regional estimate	0.04	5.85	0.47	16.69	2.02

Table II. Estimated parameters of curves[1] fitted to plots of squared errors of the simple direct and poststratified estimators for selected attribute variables: Health Interview Survey, 1970 and 1969-71

Data years, estimator, and parameter	Variable				
	Percent less than 1 year old	Percent married	Percent separated	Percent completing high school	Percent completing college
1970	Estimated parameters				
Simple direct					
A	-0.4200×10^{-4}	-7.5800×10^{-4}	-0.0684×10^{-4}	-24.2600×10^{-4}	-3.3700×10^{-4}
B	0.0168×10^{4}	0.3033×10^{4}	0.0027×10^{4}	0.9705×10^{4}	0.1349×10^{4}
Poststratified:					
A	-0.4199×10^{-4}	-9.5559×10^{-4}	-0.0416×10^{-4}	-6.4205×10^{-4}	-1.3076×10^{-4}
B	0.0168×10^{4}	0.3822×10^{4}	0.0017×10^{4}	0.2568×10^{4}	0.0523×10^{4}
1969-71					
Simple direct:					
A	-0.3446×10^{-4}	-5.8840×10^{-4}	-0.0486×10^{-4}	-33.9500×10^{-4}	-9.3670×10^{-4}
B	0.0138×10^{4}	0.2354×10^{4}	0.0019×10^{4}	1.3580×10^{4}	0.3747×10^{4}
Poststratified:					
A	-0.4360×10^{-4}	-6.0494×10^{-4}	-0.0573×10^{-4}	-17.0790×10^{-4}	-5.1814×10^{-4}
B	0.0174×10^{4}	0.2420×10^{4}	0.0023×10^{4}	0.6832×10^{4}	0.2073×10^{4}

[1] Model: $y_i = A + B/n_i + e_i$, see text for further explanation.

Table III. Health Interview Survey sample sizes by State for 1970 and 1969-71

State	1970 sample size	1969-71 sample size	State	1970 sample size	1969-71 sample size
Total	116,401	382,543	Mississippi	1,522	5,148
			Iowa	1,453	5,190
California	11,497	37,509	Oklahoma	1,260	3,903
New York	10,017	32,789	Colorado	1,106	3,617
Pennsylvania	6,967	23,003	Nebraska	1,045	3,275
Texas	6,653	22,328			
Ohio	6,433	20,941	Oregon	1,030	3,392
			Kansas	1,025	3,221
Illinois	6,274	20,551	Arizona	904	3,501
Michigan	5,261	17,023	West Virginia	901	3,086
New Jersey	4,581	14,576	Arkansas	886	2,846
Massachusetts	3,633	11,378			
Florida	3,156	11,035	Rhode Island	773	2,448
			Maine	761	2,507
Indiana	2,816	9,654	Utah	700	2,280
Virginia	2,794	8,846	New Mexico	581	1,844
Georgia	2,778	8,475	Hawaii	510	1,466
Missouri	2,610	8,569			
North Carolina	2,601	8,802	Delaware	447	1,319
			District of Columbia	429	1,463
Wisconsin	2,585	9,181	Idaho	376	1,227
Tennessee	2,515	8,470	North Dakota	332	1,131
Louisiana	2,403	7,662	New Hampshire	328	968
Maryland	2,226	7,682			
Minnesota	2,061	6,676	South Dakota	269	932
			Montana	267	1,015
Connecticut	1,956	6,522	Alaska	224	635
Kentucky	1,901	5,649	Wyoming	177	602
Alabama	1,844	5,885	Vermont	136	463
Washington	1,705	6,229	Nevada	49	321
South Carolina	1,643	5,308			

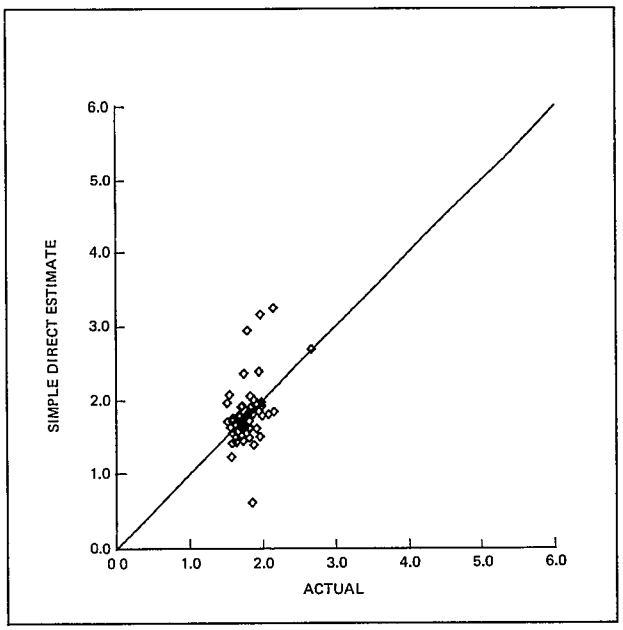

Figure I. Percent of the population under 1 year of age—simple direct estimates and actual values for 50 States and the District of Columbia: Health Interview Survey, 1969-71

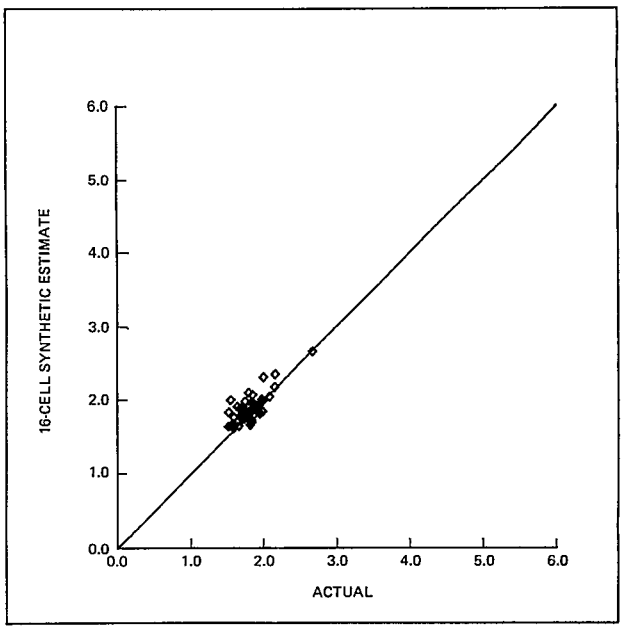

Figure III. Percent of the population under 1 year of age—16-cell synthetic estimates and actual values for 50 States and the District of Columbia: Health Interview Survey, 1969-71

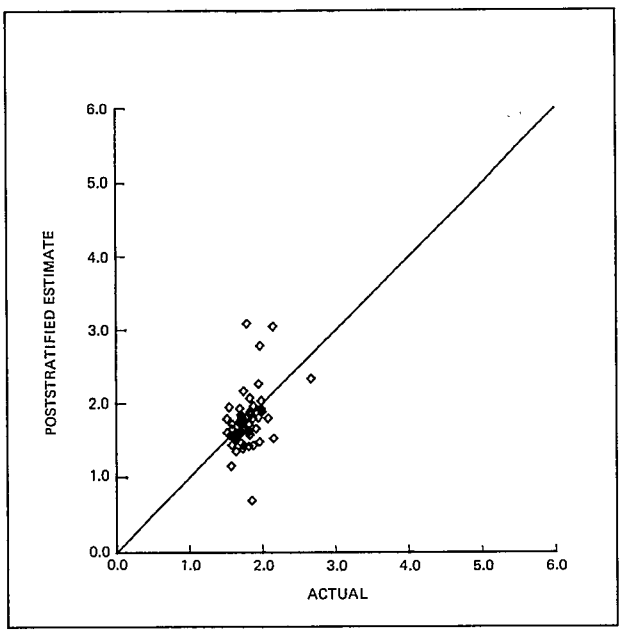

Figure II. Percent of the population under 1 year of age—post-stratified estimates and actual values for 50 States and the District of Columbia: Health Interview Survey, 1969-71

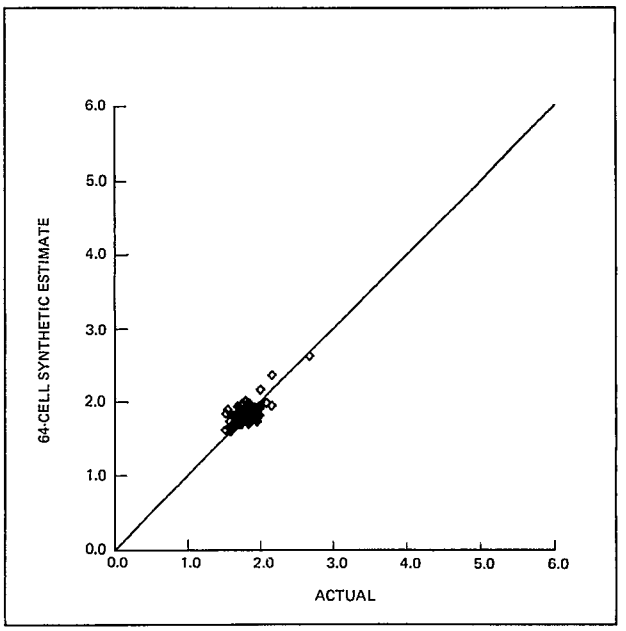

Figure IV. Percent of the population under 1 year of age—64-cell synthetic estimates and actual values for 50 States and the District of Columbia: Health Interview Survey, 1969-71

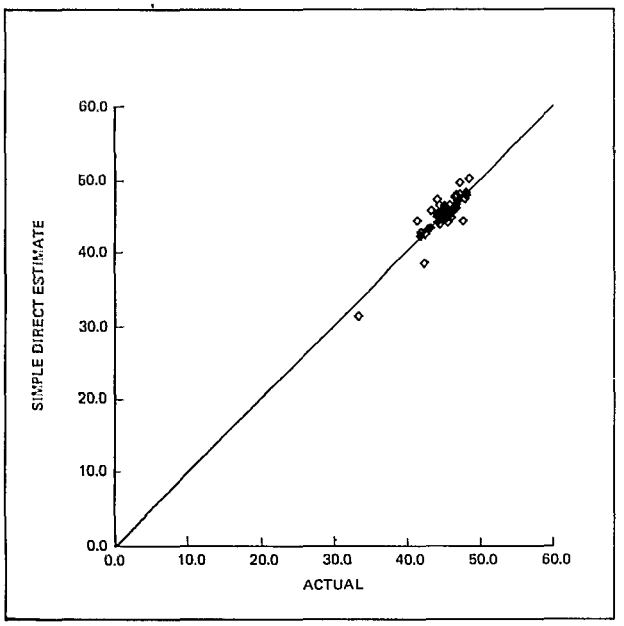

Figure V. Percent of the population who are married—simple direct estimates and actual values for 50 States and the District of Columbia: Health Interview Survey, 1969-71

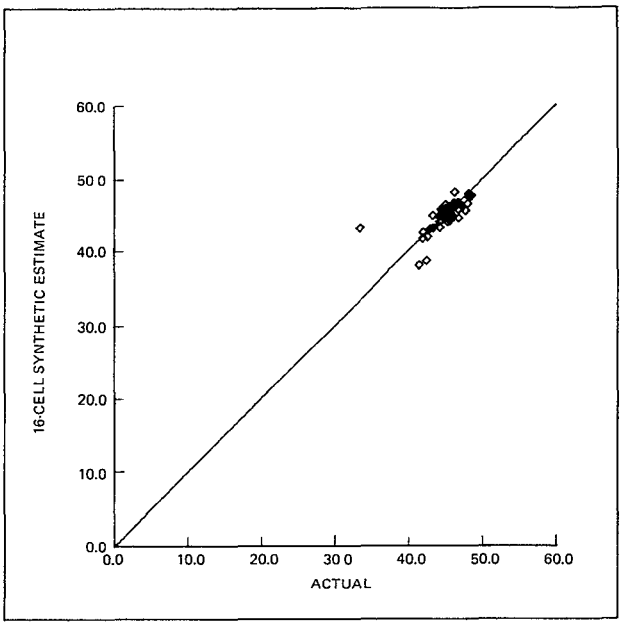

Figure VII. Percent of the population who are married—16-cell synthetic estimates and actual values for 50 States and the District of Columbia: Health Interview Survey, 1969-71

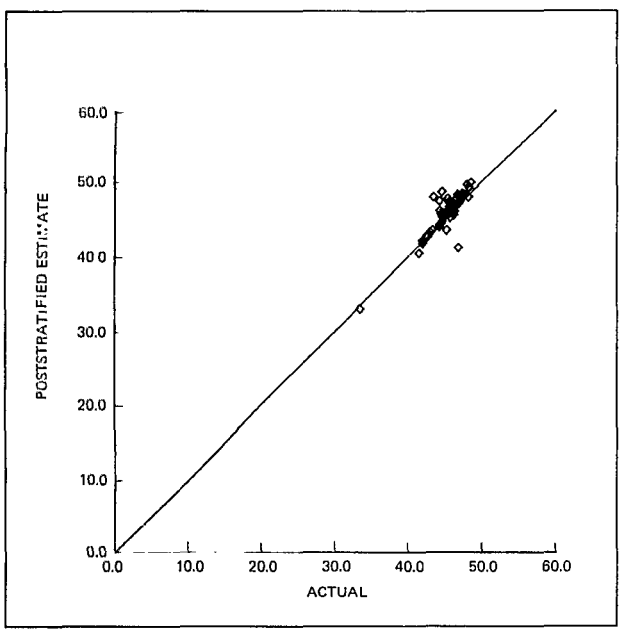

Figure VI. Percent of the population who are married—post-stratified estimates and actual values for 50 States and the District of Columbia: Health Interview Survey, 1969-71

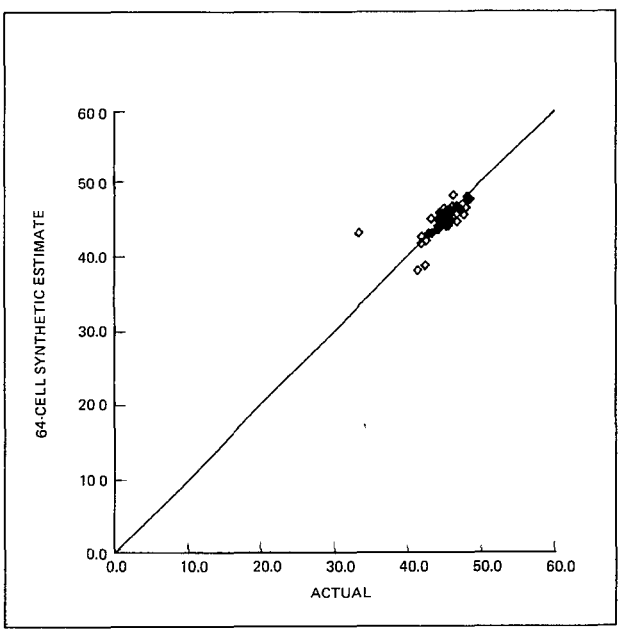

Figure VIII. Percent of the population who are married—64-cell synthetic estimates and actual values for 50 States and the District of Columbia: Health Interview Survey, 1969-71

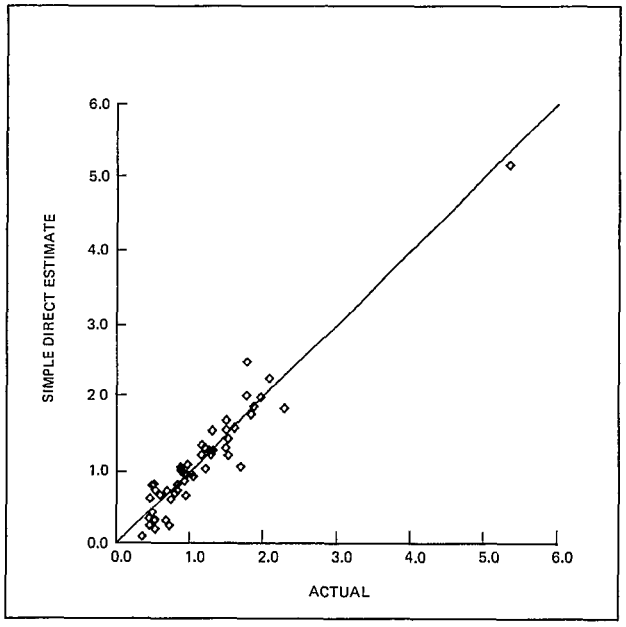

Figure IX. Percent of the population who are separated—simple direct estimates and actual values for 50 States and the District of Columbia: Health Interview Survey, 1969-71

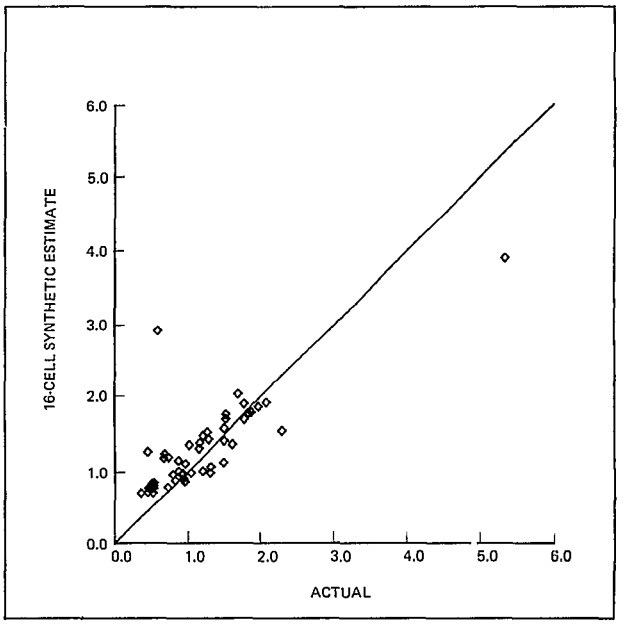

Figure XI. Percent of the population who are separated—16-cell synthetic estimates and actual values for 50 States and the District of Columbia: Health Interview Survey, 1969-71

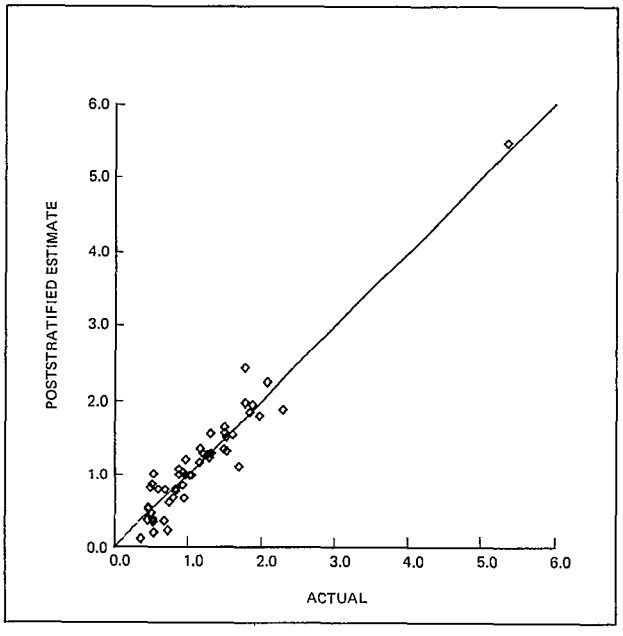

Figure X. Percent of the population who are separated—post-stratified estimates and actual values for 50 States and the District of Columbia: Health Interview Survey, 1969-71

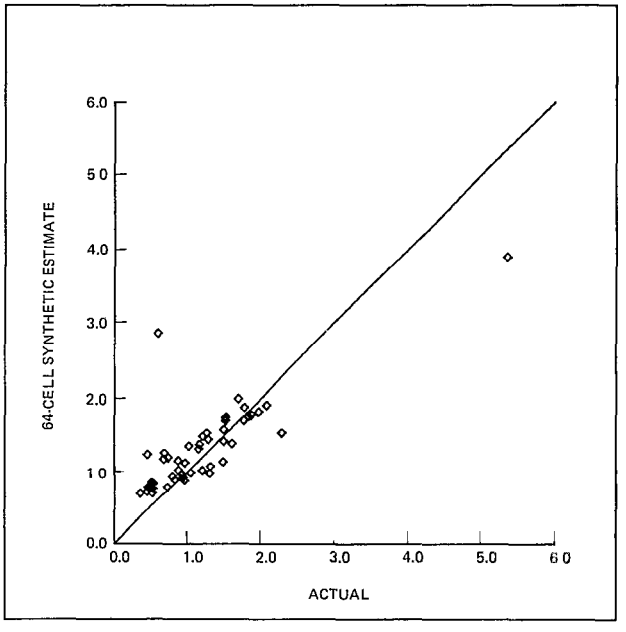

Figure XII. Percent of the population who are separated—64-cell synthetic estimates and actual values for 50 States and the District of Columbia: Health Interview Survey, 1969-71

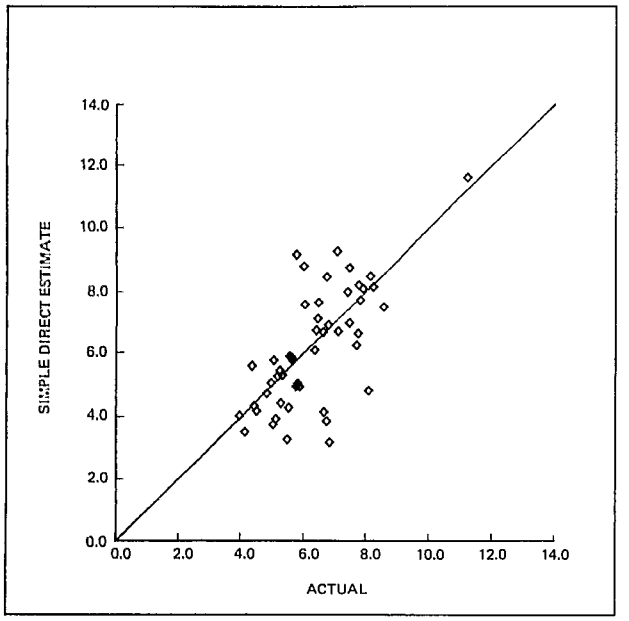

Figure XIII. Percent of the population who have completed college—simple direct estimates and actual values for 50 States and the District of Columbia: Health Interview Survey, 1969-71

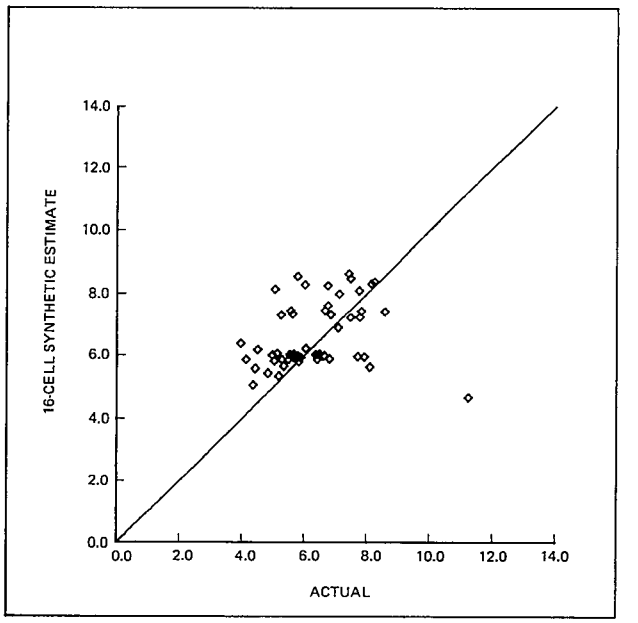

Figure XV. Percent of the population who have completed college—16-cell synthetic estimates and actual values for 50 States and the District of Columbia: Health Interview Survey, 1969-71

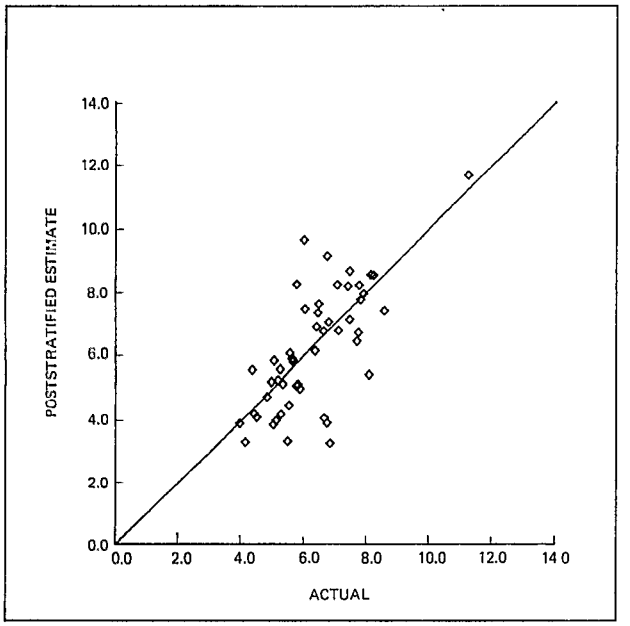

Figure XIV. Percent of the population who have completed college—poststratified estimates and actual values for 50 States and the District of Columbia: Health Interview Survey, 1969-71

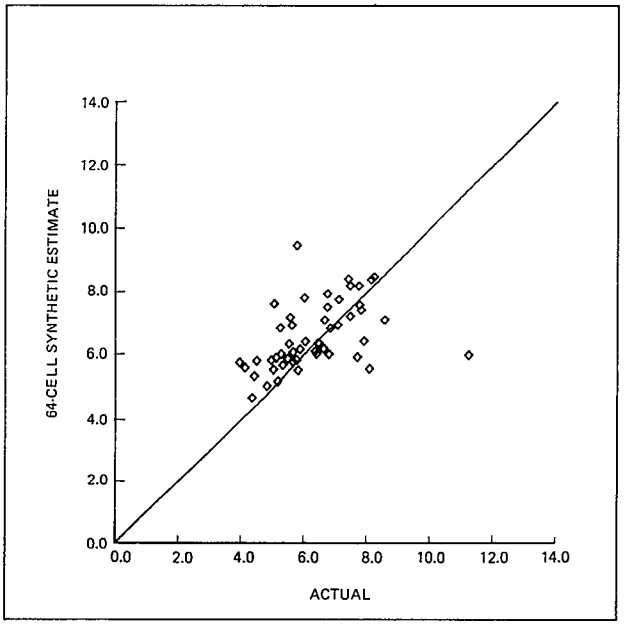

Figure XVI. Percent of the population who have completed college—64-cell synthetic estimates and actual values for 50 States and the District of Columbia: Health Interview Survey, 1969-71

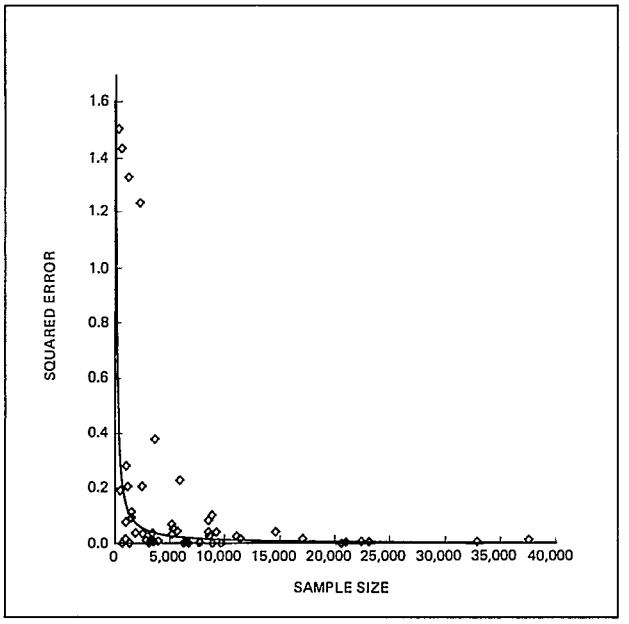

Figure XVII. Squared errors of simple direct estimates of the percent of State population less than 1 year of age, by sample size: United States, 1969-71

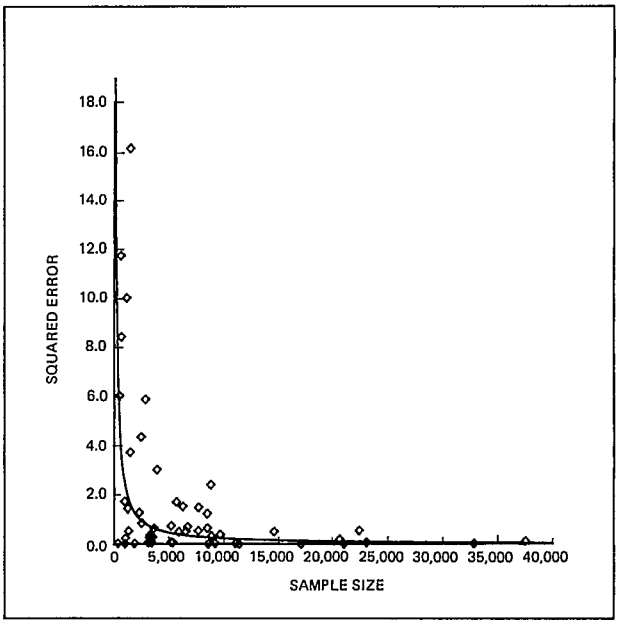

Figure XIX. Squared errors of simple direct estimates of the percent of State population who are married, by sample size: United States, 1969-71

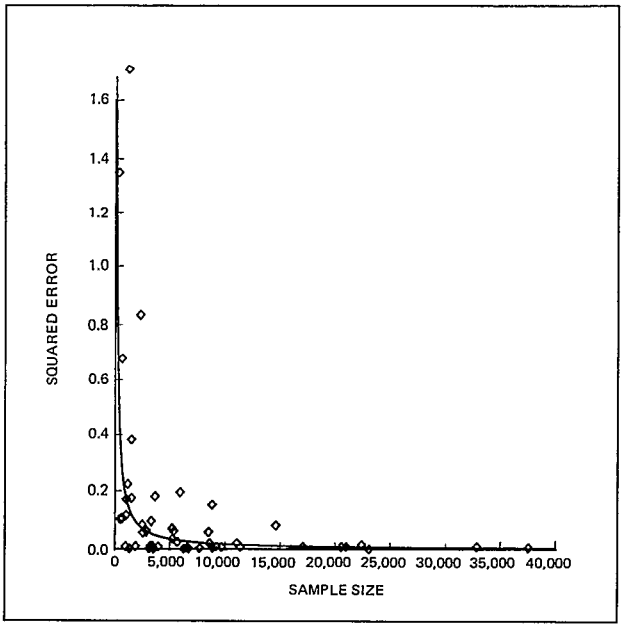

Figure XVIII. Squared errors of poststratified estimates of the percent of State population less than 1 year of age, by sample size: United States, 1969-71

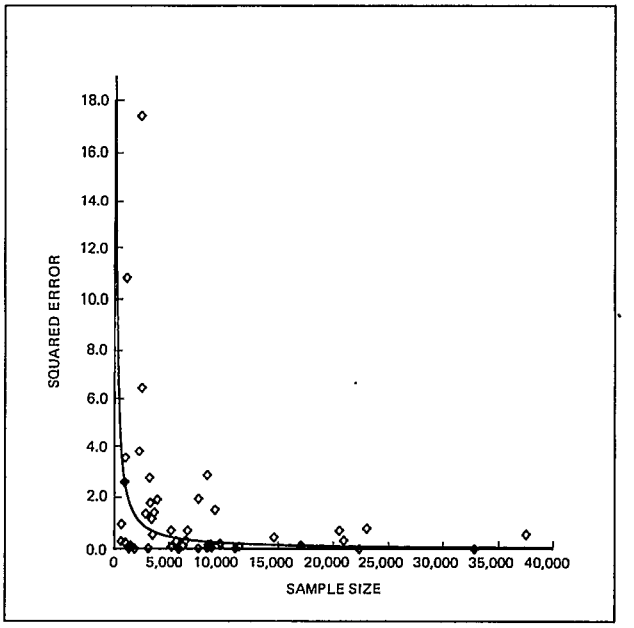

Figure XX. Squared errors of poststratified estimates of the percent of State population who are married, by sample size: United States, 1969-71

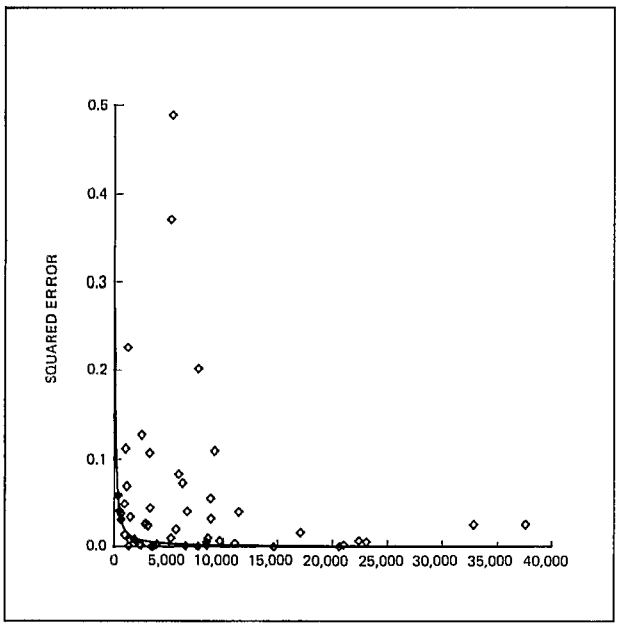

Figure XXI. Squared errors of simple direct estimates of the percent of State population who are separated, by sample size: United States, 1969-71

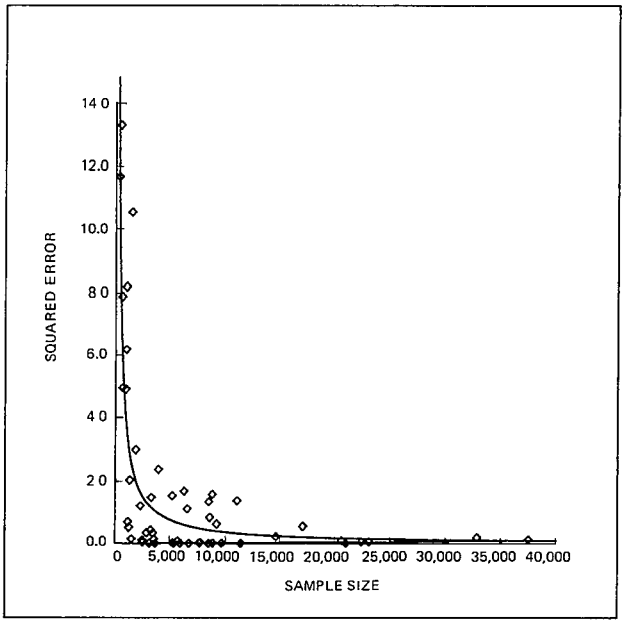

Figure XXIII. Squared errors of simple direct estimates of the percent of State population who completed college, by sample size: United States, 1969-71

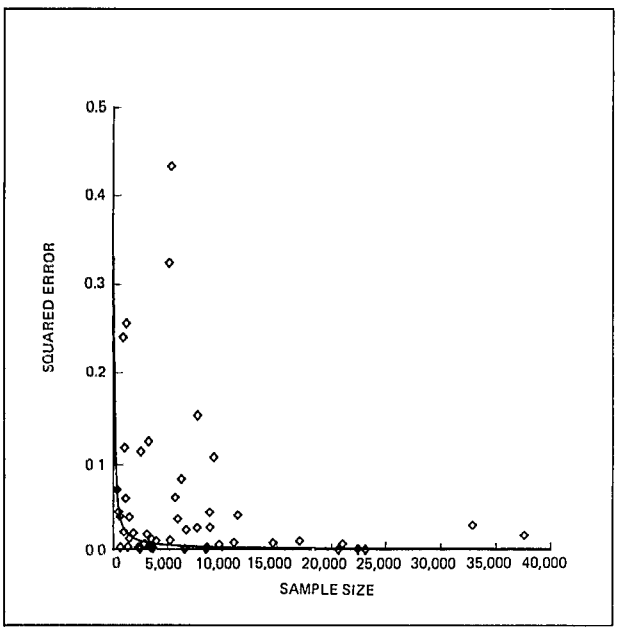

Figure XXII. Squared errors of poststratified estimates of the percent of State population who are separated, by sample size: United States, 1969-71

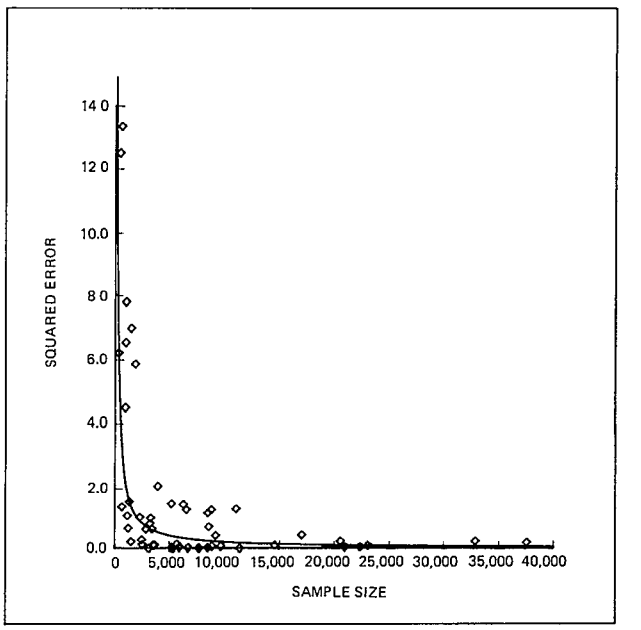

Figure XXIV. Squared errors of poststratified estimates of the percent of State population who completed college, by sample size: United States, 1969-71

*U.S. GOVERNMENT PRINTING OFFICE : 1979 0-311-240/38

VITAL AND HEALTH STATISTICS Series

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

DHEW Publication No. (PHS) 80-1356
Series 2-No. 82

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