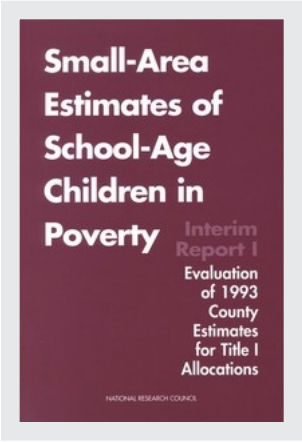

Small-Area Estimates of School-Age Children in Poverty: Interim Report 1, Evaluation of 1993 County Estimates for Title I Allocations

DETAILS

96 pages | 6 x 9 | PAPERBACK
ISBN 978-0-309-05832-2 | DOI 10.17226/5885

AUTHORS

Panel on Estimates of Poverty for Small Geographic Areas, National Research Council

Small-Area Estimates of School-Age Children in Poverty

Interim Report I: Evaluation of 1993 County Estimates for Title I Allocations

Constance F. Citro, Michael L. Cohen, Graham Kalton, and
Kirsten K. West, *Editors*

Panel on Estimates of Poverty for Small Geographic Areas

Committee on National Statistics

Commission on Behavioral and Social Sciences and Education

National Research Council

NATIONAL ACADEMY PRESS
Washington, D.C. 1997

NATIONAL ACADEMY PRESS • 2101 Constitution Ave., N.W. • Washington, D.C. 20418

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This report has been reviewed by a group other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

The National Academy of Sciences is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce Alberts is president of the National Academy of Sciences.

The National Academy of Engineering was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. William A. Wulf is president of the National Academy of Engineering.

The Institute of Medicine was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Kenneth I. Shine is president of the Institute of Medicine.

The National Research Council was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce Alberts and Dr. William A. Wulf are chairman and vice chairman, respectively, of the National Research Council.

The project that is the subject of this report is supported by Contract RN96131001 between the National Academy of Sciences and the U.S. Department of Education. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the organizations or agencies that provided support for this project.

Copies of this report are available for sale from the National Academy Press, 2101 Constitution Avenue, N.W., Box 285, Washington, D.C. 20055. Call 800-624-6242 or 202-334-3313 (in the Washington Metropolitan Area). This report is also available on-line at **http://www.nap.edu.**

Library of Congress Catalog Card Number 97-68115

International Standard Book Number 0-309-05832-5

Printed in the United States of America.

PANEL ON ESTIMATES OF POVERTY FOR SMALL GEOGRAPHIC AREAS

Contents

	PREFACE	vii
	EXECUTIVE SUMMARY	1
1	INTRODUCTION	5
2	POVERTY ESTIMATES BASED ON CENSUS AND CPS DATA	8
3	MODEL-BASED ESTIMATES OF POOR SCHOOL-AGE CHILDREN	17
4	PANEL ASSESSMENT OF THE METHODOLOGY	25
5	RECOMMENDATION FOR TITLE I ALLOCATIONS	36
6	NEXT STEPS	41
	APPENDICES	
A	THE TITLE I ALLOCATION PROCESS	49
B	COMPARISON OF CENSUS AND CPS ESTIMATES OF POVERTY	51
C	CENSUS BUREAU'S METHODOLOGY FOR MODEL-BASED ESTIMATES	63
D	POPULATION ESTIMATES	69
E	FUTURE RESEARCH	72
F	SPECIAL CASE: ESTIMATES FOR PUERTO RICO	80
	REFERENCES	82
	BIOGRAPHICAL SKETCHES, PANEL MEMBERS AND STAFF	84

Preface

The Committee on National Statistics convened the Panel on Estimates of Poverty for Small Geographic Areas to conduct a study mandated by the reauthorization of the Elementary and Secondary Education Act (The Improving America's School Act of 1994). One purpose of the act is to distribute Title I funds for education programs for disadvantaged children on the basis of estimates of school-age children in poverty (ages 5-17) that are more up to date than estimates from the 1990 census. This interim report is the first from the panel, focusing on 1993 estimates for counties. The panel's future work will include reviewing further updates for county estimates and estimates for school districts.

It was originally intended that the panel would start work soon after the passage of the Act and that, after 18 months, the panel would produce an interim report on the Census Bureau's methodology for county estimates. It was envisioned that, following this report, the Census Bureau would release the county estimates, and the panel would issue a brief report to the Secretaries of Commerce and Education on the appropriateness and reliability of these estimates. The Secretaries are required to base their decision on whether to use those estimates for allocating Title I funds for the 1997-1998 and 1998-1999 school years on the panel's assessment.

The contract for the study was delayed, however, so the panel could not begin its work until June 1996 and had considerably less time than originally planned. The panel was faced with producing a report on both its evaluation of the methodology and its assessment of the estimates for their intended purpose in less than a year, if the report was to be of assistance to the Secretaries of Commerce and Education in reaching their decision. Since school districts make deci-

sions on hiring and retaining teachers in the spring, and they need to know what funds are available, the date for the decision could not be changed. The panel would have liked more time for obtaining more information and preparing this report, but the panel wants to respond to its mandate to help the Department of Education meet the pressing needs of school districts.

The panel reviewed the work of Census Bureau staff at three meetings—in June, October, and December 1996. During this time, the Census Bureau was still conducting its work. It provided provisional estimates to the Department of Education and our panel on January 7, 1997. Later that month, the Department of Commerce informed the panel that those preliminary estimates were to serve the purposes of the legislation.

We thank the staff of the Department of Education and the Bureau of the Census for their untiring assistance to us, especially under the pressures of time. Many people helped us with evaluations we requested and with careful reviews of technical descriptions in our report. We also thank TerriAnn Lowenthal, of the Rothleder-Lowenthal Group, for providing information on the legislative history of small-area poverty estimates and our study.

I thank my colleagues on the panel for their valuable contributions to our deliberations, investigations, and report and for doing so under a very demanding situation. We were fortunate to have been assisted by a very able staff. Above all, we are appreciative of research associate Kirsten K. West, who prepared many drafts of the panel's report. Michael L. Cohen served as interim study director and contributed to the report in many ways. Constance Citro began working with us during the intensive process of report revisions, and we are delighted that she will be the study director for the next phase of the project. Meyer Zitter, who served as a consultant, investigated a number of technical issues for the panel. Eugenia Grohman, associate director for reports of the Commission on Behavioral and Social Sciences and Education, edited the report and also assisted in our deliberations. We also benefited from the helpful advice of Miron L. Straf of the Committee on National Statistics. Margaret Gill and Kathleen Saslaw provided administrative support for our study, and Candice Evans assisted in the production of the report. To all we are grateful.

Graham Kalton, *Chair*
Panel on Estimates of Poverty for
Small Geographic Areas

Executive Summary

Estimates of numbers of school-age children in poverty are used by the U.S. Department of Education under Title I of the Elementary and Secondary Education Act to allocate federal funds to school districts for programs to aid disadvantaged children. Until now that allocation has been based on the numbers and percentages of school-age children in poverty by county from the most recent decennial census. In 1994 Congress authorized the Bureau of the Census to provide updated estimates of the numbers of school-age children in poverty, first for counties and subsequently for school districts. The use of these estimates for the Title I allocations is required unless the estimates are determined to be inappropriate or unreliable. Congress also authorized a study of the Census Bureau's program for producing these small-area poverty estimates. That study is being carried out by the Panel on Estimates of Poverty for Small Geographic Areas of the Committee on National Statistics. This is the panel's first report.

For the fiscal 1997 and 1998 allocations, the Census Bureau has developed county-level estimates of the numbers of children aged 5-17 in 1994 who were living in and related to a family in poverty in 1993. Its estimation procedure uses a statistical model that combines data from several sources, including the March Current Population Survey (CPS), food stamp program records, income tax return records, and county population estimates. The panel has assessed the Census Bureau's work in order to fulfill its mandate to advise the Secretaries of Commerce and Education on the advisability of using those estimates for the Title I allocations to counties for the 1997-1998 and 1998-1999 school years. The panel concludes that the Census Bureau's model-based estimates represent a significant step toward the provision of more up-to-date estimates of poverty for small geo-

graphic areas but that these estimates have not yet been sufficiently evaluated to serve as the sole basis for allocating funds under Title I.

The panel strongly endorses a model-based approach for county-level estimates of school-age children in poverty and commends the Census Bureau for working to develop a specific model for this purpose. In comparison with the continued use of 1990 census estimates, the use of the Census Bureau's model-based estimates has a clear advantage of employing more up-to-date information. In selecting a model, however, it is important to question the assumptions it uses to see that they are reasonable, to examine predictions to see that they contain no identifiable systematic errors, and to compare the selected model to alternative models. The Census Bureau's model has not yet been sufficiently evaluated in these respects. Many aspects of the model's performance need to be more extensively tested before the panel can recommend basing the Title I allocations solely on estimates from it.

Yet it is not desirable to continue to base Title I allocations solely on estimates of poverty in 1989 from the 1990 census data. Although those estimates have the advantage of being based on a much larger sample than the CPS data used in the model, they are missing the major changes in the distribution of poverty that occurred between 1989 and 1993. For the immediate purpose of Title I allocations, the panel has had to balance its concerns about using the Census Bureau's model-based estimates against its concerns about using estimates based on the 1990 census data. The panel concludes that a solution that takes advantage of the Census Bureau's work on model-based estimates but reduces the impact of possible limitations in those estimates is the most appropriate approach at this time. Therefore, the panel's recommendation uses estimates of the number of children aged 5-17 in families in poverty in each county from both the Census Bureau model and the 1990 census data.

The panel recommends to the Secretaries of Commerce and Education that funds under Title I for fiscal 1997 be allocated on the basis of estimates that are obtained by averaging two poverty rates and then applying the average rate to the 1994 population estimate. First, calculate an average poverty rate for a county as a simple average of (1) the rate based on the number of related school-age children in the 1990 census who were in poverty in 1989 and (2) the rate based on the Census Bureau's model-based estimates of the number of related school-age children in 1994 who were in poverty in 1993. Then, obtain the number of related school-age children in poverty by multiplying this average rate by the Census Bureau's estimate of the number of related school-age children (ages 5-17) in the county in 1994.

The panel's recommendation takes some account of the changing number and geographic distribution of children in poverty by using the model-based estimates, but it also uses the decennial census estimates to moderate the results from a model that has not yet been fully evaluated. This solution will smooth the transition to model-based estimates for subsequent allocations, after further re-

search has been conducted. This recommendation is made in response to the need of the Secretary of Education to make an immediate decision regarding allocation of Title I funds for fiscal 1997. The panel stresses that the recommendation pertains only to Title I allocations to be made now. It is not a recommendation for any other purpose.

For the future, the panel encourages the Census Bureau to continue the research it has begun. The need to produce estimates in time for the fiscal 1997 allocations limited the Census Bureau's ability to thoroughly evaluate its model-based estimates, compare its model with alternative ones, and further investigate other approaches to developing updated estimates of poverty for small areas. Future work should take advantage of new research, some of it already under way at the Census Bureau, on statistical methods for making estimates for small geographic areas. The Census Bureau also needs to determine whether it is possible to develop acceptably reliable model-based estimates at the school district level. Finally, it is critical that any official estimates of the numbers of children in poverty for counties or school districts released by the Census Bureau be accompanied by formal documentation of methods, estimates of reliability, and other detailed evaluations.

1

Introduction

Estimates of income and poverty for states and smaller geographic areas are used by many federal and state agencies to allocate funds for a variety of programs, including the Community Development Block Grant Program (U.S. Department of Housing and Urban Development); the Job Training Partnership Act (U.S. Department of Labor); the Head Start Program (U.S. Department of Health and Human Services); the Community Investment Program (Federal Housing Finance Board); and the Rural Housing Program (Farmers Home Administration). In fiscal year 1994, $30 billion was allocated for all such programs by the federal government. Some allocations to states use estimates from the Annual Demographic Supplement to the Current Population Survey (CPS)—commonly known as the March Income Supplement to the CPS—or estimates from a combination of data from the decennial census, the March CPS, and administrative records. Other allocations, particularly for such small areas as counties and school districts, are based on estimates from the census.

The largest program that uses small-area estimates derives from Title I of the Elementary and Secondary Education Act; it accounts for about 20 percent of the total, more than $6 billion in fiscal 1993. Under the program, the U.S. Department of Education allocates funds for compensatory education programs to meet the needs of educationally disadvantaged children. The funds are allocated at the county level, and the states then distribute these funds among school districts within each county (see Moskowitz et al., 1993). The county allocations are based on the numbers of eligible children, who are predominantly children aged 5-17 in families with incomes below the poverty level,[1] but also include children

[1]The poverty status of individuals is determined by comparing the before-tax money income of their family to the appropriate poverty threshold. The poverty thresholds vary by family size and are updated by the change in the Consumer Price Index each year. See Citro and Michael (1995) for an

in foster homes, children in families above the poverty level that receive Aid to Families with Dependent Children (AFDC), and children in local institutions for neglected and delinquent children. The allocations depend primarily on the numbers of eligible children, but they also depend on the proportions of school-age children who are eligible. The allocations also take into consideration the state's average per-pupil expenditure, and the formula includes a hold-harmless provision to cushion the impact of decreases in allocations. Details of the allocation process are provided in Appendix A.

The poverty estimates for the Title I program used by the Department of Education are provided by the Bureau of the Census. The practice until recently had been to use poverty estimates based on decennial census data. Since the rates of poverty, as well as the actual number of children in poverty, change over time, however, Congress recently authorized the Bureau of the Census to provide updated estimates of poverty for the county allocations and, subsequently, for school districts (formally known as local educational agencies, LEAs). Having the most up-to-date estimates is important so that resources can be directed toward areas that are most in need.

Congress also authorized a study—through the Department of Education—by the National Research Council's Committee on National Statistics to review the Census Bureau's program for small-area poverty estimates. The statute requires that the Department of Education use the updated estimates unless the Secretaries of Commerce and Education determine that they are "inappropriate or unreliable" on the basis of the committee's study ("Improving America's Schools Act of 1994," P.L. 103-382, and 1996 continuing resolution).

The Panel on Estimates of Poverty for Small Geographic Areas was set up under the committee to carry out the authorized study. The panel is charged with a broad review of the Census Bureau's postcensal poverty estimates for small geographic areas and their utility for Title I allocations, including: the methodology for producing and publishing those data; alternative methods of producing such data; the availability of alternative indicators of poverty that could be used for comparison; the reliability of the data, including comparisons with similar data; and the usefulness of the estimates for federal programs that allocate funds to state and substate areas.

It was the intent of Congress that the panel study be conducted as the Census Bureau began its small-area estimates program, so that the program could benefit from the study. The legislation specifying the use of updated Census Bureau poverty estimates was signed into law on October 20, 1994. However, funding for the panel study was not provided to the Department of Education until the fiscal 1996 continuing funding resolution, and the Department signed a contract for the study on March 15, 1996. The panel to conduct the study began its work

---

evaluation of the current official poverty measure and a proposed alternative measure; the issue of how poverty should be defined is not considered in this report.

in June 1996 and is scheduled to work through 1998, producing a final report at that time and such interim reports as are needed.

The first task for the Census Bureau was to produce estimates for counties of the numbers of children aged 5-17 from families with incomes below the poverty level in 1993, to be used by the Department of Education for the 1997-1998 and 1998-1999 Title I allocations.[2] Within 90 days of the release of those estimates by the Census Bureau, the panel is required to provide a report to Congress with an assessment of the reliability and utility of the estimates. This interim report of the panel responds to the legislative mandate for an assessment of updated poverty estimates.

In order to ensure that the Department of Education would have appropriate information in time for the allocations to be made in the spring of 1997, it was originally assumed that the Census Bureau would release its estimates in the fall of 1996 and that the panel would provide its assessment shortly thereafter. However, because the Census Bureau had not formally released its estimates by the end of 1996 and had not set a firm date for doing so, this report assesses the preliminary estimates that were provided to the panel and the Department of Education on January 7, 1997, and some draft descriptions of the methodology for them that were provided earlier.[3]

The panel's work has been limited by the information made available to it and the time available for its analysis. We would have preferred to delay this report until the Bureau of the Census provided more evaluation materials about its estimation program and we had time to study those materials. However, the next Title I allocations must be made in the spring of 1997, and the panel believes that its initial assessment will be useful to the Department of Education in deciding what data to use for those allocations.

Our report contains five other sections and six appendices. Section 2 assesses the advantages and disadvantages of using decennial census data and CPS data for small-area poverty estimates and looks at the differences between the census and the CPS as data sources for income and poverty. Section 3 describes the Census Bureau's methodology for producing updated small-area poverty estimates for 1993. Section 4 provides the panel's assessment of that methodology and the resulting estimates. Section 5 presents the panel's recommendation regarding the use of the estimates for the Title I fund allocations for 1997-1998. Section 6 summarizes the panel's suggestions of directions for future research. The appendices provide descriptive and technical information.

---

[2]More precisely, the Census Bureau's estimates pertain to poor related children aged 5-17 (whom we refer to as "poor school-age children"). Related children include the following family members in a household: own children under 18 years of age (a never-married son or daughter by birth, a stepchild, or an adopted child of the householder) and all other persons under 18 years of age, regardless of marital status, who are related to the householder, except the spouse of the householder. Foster children are not included since they are not related to the householder (Bureau of the Census, 1993a).

[3]The Bureau of the Census released the updated estimates on March 26, 1997.

2

Poverty Estimates Based on Census and CPS Data

CENSUS DATA

Traditionally, the decennial census has been the source of small-area income and poverty estimates, with each census being used until data from the next census are available. The 1980 census data, covering income and poverty for 1979, were used for Title I allocations through the 1993-1994 school year; the 1990 census data with 1989 income and poverty data were not available for use until the 1994-1995 school year. As a result of changes over time, the data become increasingly outdated over the course of a decade or so and do not reflect current socioeconomic conditions and demographic distributions in the population.

For small geographic areas, the changes in a decade, or less, can be substantial. Over the course of a few years a county can experience rapid population growth from new suburban development and expansion, or rapid population loss from outmigration in response to a decrease in employment opportunities. Likewise, poverty rates can increase or decrease substantially because of a rise or decline of an industry, migration, or changes in other economic and social conditions. Census data that may have accurately represented the population at the time the census was taken will not reflect subsequent socioeconomic and demographic changes. As a consequence, areas that experience either large demographic or economic shifts or both over the decade may be disproportionately overfunded or underfunded under Title I allocations that are based on census estimates.

Concerns about using decennial census income data that become outdated were reinforced by changes observed between the 1980 and 1990 censuses. Nationally, the number of poor children aged 5-17 rose by 5 percent over the 10-

year period, from 7.7 million to 8.1 million. At the state level, there was considerable variability: 24 states and the District of Columbia experienced declines in the number of poor children aged 5-17 of up to 34 percent; 15 states saw increases of up to 25 percent; 8 states had increases ranging between 25 and 50 percent; and 3 states had increases between 50 and 67 percent (Moskowitz et al., 1993:71). When considering the use of 1990 census data for current allocations, there is similar concern. Income data collected in the 1990 census are referenced to 1989; they do not capture the recession that began in 1990 or the changes in the proportion and geographic distribution of people below the poverty level that resulted from the subsequent rise in unemployment.

In addition to not being current, decennial census data on income are themselves estimates, and as such they are subject to sampling error because the data are collected from only a sample of households. In the 1990 census, income data were collected on the "long form" that was mailed to about 1 out of every 6 households—or about 15 million households in the United States. Sampling rates varied from 1 in 2 for very small counties and places (with an estimated 1988 population of less than 2,500) to 1 in 8 for very populous census tracts (or equivalent areas). Although sampling errors are relatively small for large geographic areas, such as states, the sampling errors for smaller geographic areas can be large relative to the estimate.

Table 2-1 provides information on the amount of error due to sampling variability in the estimated numbers of poor school-age children (related children aged 5-17) by county from the 1990 census. For example, for 63 counties, the margin of error due to sampling variability is less than 5 percent of the estimated number of poor school-age children.[1] The estimates for these counties are thus fairly precise. Moreover, these counties, although a small percentage (2%) of all 3,138 counties in 1990, are large ones: they contained 37 percent of the nation's poor school-age children estimated by the 1990 census. However, for 1,405 counties, the margin of error due to sampling variability is 25 percent or more of the estimated number of poor school-age children. Although these counties contained only 6.4 percent of the poor school-age children in the nation estimated by the 1990 census, the imprecision in their estimates is of concern for the Title I allocation.

[1]The margin of error is expressed in Table 2-1 as the relative width of the 90-percent confidence interval; that is, the width of the interval as a percentage of the estimated number. Confidence intervals for a sample estimate are ranges that include the average result of all possible samples with a known probability; they are constructed from the estimate and its standard error (the measure of the magnitude of sampling variability of the estimate). The 90-percent confidence interval for an estimate is 1.645 standard errors below the estimate to 1.645 standard errors above the estimate: there is a 90-percent chance that the 90-percent confidence interval includes the average estimate of all possible samples.

TABLE 2-1 Distribution of Counties by Relative Widths of the 90-Percent Confidence Interval for the Estimated Number of Poor Related Children Aged 5-17 in 1989: 1990 Census

Relative Width of Confidence Interval[a]	Counties		Poor Children	
	Number	Percent	Number	Percent
All Counties	3,138	100.0	7,544,737	100.0
Less than 5%	63	2.0	2,818,997	37.4
5 to 10%	236	7.5	1,846,546	24.5
10 to 15%	466	14.9	1,258,897	16.7
15 to 20%	538	17.1	761,149	10.1
20 to 25%	430	13.7	372,733	4.9
25 to 50%	1,061	33.8	449,464	6.0
50 to 75%	238	7.6	31,585	0.4
More than 75%	106	3.4	5,366	(Z)[b]

NOTE: Three counties with no poor related children aged 5-17 in the sampled households are excluded from the table.

[a]The relative width of the confidence interval is the percentage that the width of the 90 percent confidence interval represents of the estimated number of poor related children aged 5-17 in a county. The 90 percent confidence interval is 3.29 times the standard error of the estimate.

[b]Less than .05 percent

SOURCE: Data from Bureau of the Census.

CPS DATA

The CPS is designed primarily to provide monthly estimates of labor force participation, employment, and unemployment. Every March, the CPS collects additional data on income for the prior calendar year from which poverty rates can be determined. The CPS is therefore a more timely data source on poverty than the census. Indeed, the annual March Income Supplement to the CPS provides the official national measure of poverty.[2] The March Income Supplement also serves as a basis for some federal fund allocations (Office of Management and Budget, 1993).

The CPS sample size is not large enough to produce detailed information on the changes that occur over time in the geographical distribution of the popula-

[2]The Survey of Income and Program Participation (SIPP) is another source of up-to-date income and poverty data. Two Committee on National Statistics panels have recommended that SIPP become the official source of annual national poverty estimates in place of the March CPS (see Citro and Kalton, 1993; Citro and Michael, 1995).

TABLE 2-2 Change in the Total Estimated Number of Poor Persons between 1989 and 1993 for Selected States: March CPS Data

State	Change in Poverty (in percent)	Standard Error of Estimate (in percentage points)[a]
United States	24.5	1.8
Florida	52.3	12.8
California	44.2	9.4
New Jersey	35.8	15.3
Ohio	27.2	12.3
Pennsylvania	26.8	12.1
New York	26.1	8.7
North Carolina	23.0	11.4
Massachusetts	19.6	13.9
Michigan	19.0	10.6
Illinois	7.3	10.1
Texas	4.1	8.0

[a]3.29 times the standard error gives the 90-percent confidence interval.

SOURCE: Data from Bureau of the Census.

tion in poverty, but the survey can provide some useful indicators. They can illustrate how large changes can occur over short periods of time and how different areas can experience substantially different rates of change. Consider, first, the changes in the distribution in the number of poor people of all ages between 1990 and 1994 (income in 1989 and 1993). The CPS sample is sufficiently large to estimate such changes for 11 states, although the estimates are subject to large sampling errors; see Table 2-2.[3] Overall, the estimated total number of poor people in the country increased by 24.5 percent, but with a wide range across states: 52 percent for Florida and 44 percent for California, but only 7 percent for Illinois and only 4 percent for Texas. Statistical sampling error affects the precision of these estimates, but it is still clear that there were changes over the period and that they differed among states.

The CPS data, when grouped by selected categories of counties and averaged over 3 years to improve precision, show similar changes in the estimated number of school-age children in poverty, which increased for the nation as a whole by 19.6 percent between 1989 and 1993; see Table 2-3. The increase is evident for counties in all regions of the country, in metropolitan and nonmetropolitan areas,

[3]For these 11 states, the sample was designed to meet reliability requirements for consecutive monthly changes in the unemployment rate.

TABLE 2-3 Estimated Number of Related Children Aged 5-17 in Poverty by Selected Categories of Counties: 1989 and 1993, March CPS Data

County Category	Children in Poverty, Income Year 1989[a]	Children in Poverty, Income Year 1993[b]	Change in Poverty between 1989 and 1993 (in percent)
U.S. Total	8,036,000	9,613,000	19.6*
Metropolitan			
Central	5,608,000	6,853,000	22.2*
Other	362,000	471,000	30.1*
Nonmetropolitan	2,066,000	2,289,000	10.8*
Region[c]			
Northeast	1,312,000	1,636,000	24.7*
Northcentral	1,754,000	1,986,000	13.3*
South	3,396,000	3,813,000	15.7*
West	1,674,000	2,178,000	30.1*
Population Size			
Under 9,999	202,000	243,000	20.3
10,000-49,999	1,489,000	1,538,000	3.3
50,000-99,999	759,000	927,000	22.2*
100,000-499,999	2,143,000	2,448,000	14.2*
500,000-999,999	1,229,000	1,510,000	22.9*
1 million and over	2,214,000	2,947,000	33.1*

*Statistically significant difference from 0 using a 10 percent significance level.

[a]The estimates are 3-year centered averages. For 1989 estimates, averages of March 1989, 1990, and 1991 CPS data were used (reported income in 1988, 1989, and 1990, with population controls derived from the 1980 census).

[b]The estimates are 3-year centered averages. For 1993 estimates, averages of March 1993, 1994, and 1995 CPS data were used (reported income in 1992, 1993, and 1994, with population controls derived from the 1990 census, including an adjustment for the estimated undercount beginning with the March 1994 CPS).

[c]The Census Bureau regions are as follows: Northeast: Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, and Pennsylvania; Northcentral: Ohio, Indiana, Illinois, Michigan, Wisconsin, Missouri, Minnesota, Iowa, North Dakota, South Dakota, Nebraska, and Kansas; South: Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida, Kentucky, Tennessee, Alabama, Mississippi, Arkansas, Louisiana, Oklahoma, and Texas; West: Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah, Nevada, Washington, Oregon, California, Alaska, and Hawaii.

SOURCE: Data from Bureau of the Census.

and in all population size categories, but there is substantial variation in the size of the increase. The largest increases are for counties with a population size in the category of 1 million or more (33.1%), "other" (noncentral) counties in metropolitan areas (30.1%), and counties in the West region (30.1%). The smallest increases are for counties with a population size in the category of 10,000-49,999 (3.3%, not statistically significantly different from zero), counties in nonmetropolitan areas (10.8%), and counties in the Northcentral region (13.3%).[4]

Although the CPS provides more current data than the decennial census, its much smaller sample size limits its ability to produce estimates for smaller areas. The March CPS collects data from only about 60,000 households (50,000 beginning in 1996), containing about 28,000 children aged 5-17, compared with about 15 million households for the 1990 census.[5]

For all but a few very large counties, the CPS sample size is too small to produce reliable estimates. In fact, there is no CPS sample in over one-half of U.S. counties; only about 1,300 counties of 3,143 counties (in 1994) are represented in the sample. And for those counties for which CPS sample data are available, the estimates of poverty and of the population aged 5-17 are, as a rule, extremely imprecise because of small sample sizes. However, as discussed in Section 3, a model-based approach that combines CPS estimates with administrative data in a statistical model can be used to yield estimates for counties that are more up to date than census estimates and have acceptable prediction errors. The Census Bureau's county-level model increases the CPS sample size for counties by combining 3 years of data.[6]

DIFFERENCES BETWEEN CENSUS AND CPS DATA

The census and the CPS differ in other ways besides sample size. Even for a census year, the decennial census and the CPS do not produce identical results

[4]The increases in the number of poor school-age children between 1989 and 1993 are the result of increases in the number of school-age children, as well as of increases in the poverty rate for this group. Consequently, for the United States as a whole, the poverty rate for school-age children increased by less than the increase in the number of poor school-age children (11.1% versus 19.6%). The increase in the poverty rate for school-age children, like the increase in their number, varied across regions of the country and types of counties.

[5]In turn, the SIPP sample size, currently 37,000 households, is smaller than that of the March CPS.

[6]By combining 3 years of data from the March 1993, 1994, and 1995 CPS to produce estimates for 1993, the number of counties represented in the sample increases from about 1,300 to about 1,500. A new 1990 census-based sample design was introduced beginning in the April 1994 CPS; some counties are included in both the new design and the old (1980 census-based) design, but other counties are included in only one design. The average number of sample households for counties represented in one or more of the 3 years is 113; for counties with populations under 10,000, the average number of sample households is 28, and for counties with 500,000 or more people, the average number of sample households is 701. However, several hundred (mostly small) counties with CPS sample households lack any sample households with poor school-age children (see Coder et al., 1996:Tables 1, 3).

TABLE 2-4 Census and March CPS Estimates of Related Children Aged 5-17 in Poverty in 1989, by Selected Categories of Counties

County Category	Children in Poverty, 1990 Census	Children in Poverty, March CPS[a]	Percentage Difference: CPS – Census as Percent of Census
U.S. Total	7,545,000	8,036,000	6.5*
Metropolitan			
Central	5,021,000	5,608,000	11.7*
Other	347,000	362,000	4.4*
Nonmetropolitan	2,177,000	2,066,000	–5.1*
Region[b]			
Northeast	1,180,000	1,312,000	11.2*
Northcentral	1,641,000	1,754,000	6.8*
South	3,174,000	3,396,000	3.9*
West	1,550,000	1,674,000	8.0*
Population Size			
Under 9,999	197,000	202,000	2.5
10,000-49,999	1,489,000	1,489,000	0
50,000-99,999	843,000	759,000	–9.9*
100,000-499,999	1,990,000	2,143,000	7.7*
500,000-999,999	1,124,000	1,229,000	9.3*
1 million and over	1,901,000	2,214,000	16.5*

*Statistically significant difference from 0 using a 10 percent significance level.

[a]The CPS estimates are 3-year centered averages of the March 1989, 1990, and 1991 CPS data (reported income in 1988, 1989, and 1990, with population controls derived from the 1980 census).

[b]See Table 2-3 for the states in each region.

SOURCE: Data from Bureau of the Census.

with regard to children in poverty. Table 2-4 shows the differences between the 1990 census (1989 income) estimates of the number of poor school-age children and the 1989 CPS estimates for the nation as a whole and for various subcategories of counties. (The CPS estimates are averages of income data for 1988, 1989, and 1990; averaging is used to improve precision given the small CPS sample size in smaller areas; see Appendix Table B-5 for a similar comparison of poverty rates.)

Overall, for the U.S. population, the CPS provides an estimate of the number of poor school-age children that is 6.5 percent higher than the decennial census.[7]

[7]Some portion of the differences shown for the United States and various kinds of subnational areas may be due to the use of 3-year centered averages for the CPS-based estimates, which included

For most groups of counties, the CPS estimate is also higher than the census estimate, and there is a suggestion of a pattern in which the ratio of the CPS estimate to the census estimate of poor school-age children in 1989 may increase as a function of county size. The panel conducted an analysis to determine whether there were statistically significant differences among the CPS-census ratios for counties grouped by population size and other characteristics, but did not find such differences. However, this work was very preliminary and needs to be extended.[8]

Though not yet fully researched and understood, differences between census and CPS estimates of poverty may result from the different ways the income data are obtained. The census and CPS use the same official poverty thresholds[9] to determine poverty status, income is counted in both as annual money income received in the previous calendar year, and both are intended to measure the same kinds of income. However, the CPS questionnaire asks respondents to provide income amounts for many more detailed categories than does the census questionnaire. For example, the decennial census asks respondents to provide a combined income amount for Supplemental Security Income (SSI), Aid to Families with Dependent Children (AFDC), and other public assistance or public welfare payments; the CPS asks separately for SSI, AFDC, and other public assistance or public welfare (including the source). Methodological research suggests that more detailed questions elicit more complete income reports (see Citro and Michael, 1995:402-405); however, the extent to which questionnaire differences affect the responses in the CPS and the census is not known.[10]

The CPS and the census also use somewhat different rules for defining the universe to which poverty applies (see Appendix B). For example, the CPS includes students living in college dormitories as family members in their parental households; the census considers the dormitory the place of residence and excludes residents of college dormitories from the poverty universe. The result is that somewhat more families with college students may be estimated as living in poverty in the CPS than in the census because a college student in a family in-

---

a year (1990 from the March 1991 CPS) in which the poverty rate for school-age children was higher than in either 1989 or 1988. The difference between the 1990 census and the single-year March 1990 CPS in the number of poor school-age children for the United States in 1989 is 4.9 percent, compared with 6.5 percent for the 3-year average figure.

[8]Table 2-4 indicates that the differences between the CPS and census estimates of poor school-age children in 1989 are statistically significant (i.e., significantly different from 0) for all county groups except those with small sample sizes. This finding is not surprising given the large national difference in the two estimates; however, it does not support a conclusion that differences between the ratios of CPS estimates to census estimates are statistically significant across county groups. A different comparison would be needed to establish such differences.

[9]For example, for a family of four the 1993 (weighted average) poverty threshold level was $14,763.

[10]Another difference is that the 1990 census questionnaire, but not the March CPS questionnaire, included a "total income" question. The intent of this question was to permit respondents to enter a single amount if they could not provide amounts by source.

creases its size and therefore its poverty threshold but likely does not add appreciably to its income.

The way the data are collected may also result in differences. In the CPS, data are collected through personal contacts (mostly by telephone) made by trained field representatives. In contrast, the census primarily relies on respondents to complete and return a questionnaire by mail. Appendix B discusses in more detail these and other differences between the two data sources, such as coverage errors and the treatment of missing data. Such differences are important to consider in evaluating the appropriateness of moving from use of the decennial census to the CPS as the basis for developing estimates of poor school-age children for the purpose of the Title I allocations.

TIMELINESS OF ESTIMATES

The CPS provides more timely data than the decennial census; however, estimates of poor school-age children for counties that are derived from the CPS (or from another survey, such as SIPP) will still lag the allocation year by 1 or more years, depending on how quickly the data from the various sources used in the estimation model become available. The estimates that the Census Bureau produced with CPS data for the 1997-1998 Title I allocations are for the numbers of school-age children in counties in 1994 who were poor in 1993. In addition to data availability, the time lag is also caused by the decision to use 3 years of CPS data in the Census Bureau's model to improve the precision of the estimates (see Section 3). Consequently, the estimates will not capture any changes in the extent and distribution of poverty among school-age children that may have occurred since the year to which they apply.

Published CPS data indicate that poverty among school-age children for the nation as a whole increased from 17.4 percent in 1989 to 20.1 percent in 1993 and then declined to 18.3 percent in 1995 (Bureau of the Census, 1990:Table 18; 1995a:Table 8; 1996:Table 2).[11] Data for 1996 are not yet available, and no data are readily available with which to estimate the changes in the distribution of poverty among school-age children across states and counties.

We were asked to evaluate the accuracy of the updated county-level estimates that the Census Bureau was able to produce with available data. We addressed the question of the accuracy of the estimates for the estimation year (1993), not the question of how well the estimates for 1993 predict poverty among school-age children in 1997. There are conceptual and operational issues involved in considering how one could make estimates of poor school-age children yet more timely. We briefly outline some possibly useful directions for research on timeliness in Section 6.

[11]These estimates are for related children aged 6-17; estimates are not published for related children aged 5-17.

3

Model-Based Estimates of Poor School-Age Children

Past reliance on the most recent decennial census to allocate federal funds to counties and other small areas primarily reflects the absence of alternative data sources with comparable or superior reliability. As discussed in Section 2, the CPS can provide reasonably reliable annual estimates at the national level of such population characteristics as the number and percent of poor children, but it cannot produce estimates for counties. Nonetheless, the CPS data may serve as the basis for creating usable estimates through the application of statistical estimation techniques to develop "model-based" or "indirect" estimates. Indirect estimators use data from other areas, time periods, or data sources to "borrow strength" and improve precision. In contrast, direct estimators use only the data from one source for the area and time period in question. A model-based approach is useful when there is no single source of information that can provide direct estimates, but relationships among several variables across various data sources can be used to provide estimates with acceptable precision.

The Census Bureau has been mindful of the need for updated small-area estimates. Even as Congress charged the Census Bureau to develop postcensal estimates of poor school-age children for counties and school districts, the Census Bureau was organizing a program to study methods for producing postcensal income and poverty estimates for states and counties during the 1990s. The Census Bureau launched this program in late 1992 with financial support from a consortium of five federal agencies. The program faces a challenging task. In particular, there is no single administrative or survey data source that provides all of the information required to develop reliable estimates of the number of school-age children in poverty by county, including income information that is detailed and precise at the county level.

The Census Bureau's research suggests that, for state and county estimates of poor school-age children, the best procedure to follow is a model-based approach. Previously, the Census Bureau used this strategy to develop estimates of median family income for states (Fay et al., 1993) and, in part, to develop population estimates for states and counties (see Spencer and Lee, 1980).[1]

This section briefly describes the model-based approach as applied by the Census Bureau to estimate the number of school-age children in poverty by county. It also describes the Census Bureau's model for states, which estimates the state-level proportions of school-age children in poverty. (Appendix C provides a more detailed, technical explanation of both models; see also Coder et al., 1996; Fay, 1996.) Finally, the section describes how the estimated numbers of poor school-age children for the counties in each state are adjusted to agree with the corresponding estimates from the state-level model. This procedure requires that the estimated proportions from the state-level model are converted to estimated numbers of poor school-age children in each state.

Both the county-level and state-level models were estimated for school-age children in 1994 who were poor in 1993 for purposes of the Title I allocations and for school-age children in 1990 who were poor in 1989 for purposes of model validation by comparison with estimates from the 1990 census. This validation exercise is discussed in Section 4, which provides the panel's assessment of the Census Bureau's methods.

COUNTY-LEVEL MODEL

Development of the Census Bureau's county-level model for estimates of the number of poor school-age children involved several steps: determining what administrative and other data sources are available for all counties that can be used in a prediction equation; specifying and estimating an equation that relates the predictor variables to a dependent variable from 3 years of the March CPS; using the estimates from the equation, together with direct estimates for counties for which they are available, to develop estimates for all counties; and, finally, adjusting the county estimates for consistency with estimates from a separate state-level model. The state-level model and the final adjustment of the county estimates are discussed following the description of the county-level model.

[1]The model-based population estimates (total noninstitutionalized population for counties and noninstitutionalized population under age 65 for states) were produced as components of one of the methods that the Census Bureau used to estimate state and county population totals for use in allocating funds under the general revenue sharing program. The final estimates were developed by averaging the results of three separate methods.

Administrative Data Used

The first step in developing model-based estimates of school-age children in poverty by county is to bring together administrative data that are related to poverty and that are available for all counties on a consistent basis (i.e., that are obtained using the same definitions and procedures). The Census Bureau examined a variety of administrative records and selected two sources as most nearly meeting these criteria: counts of the number of people receiving food stamps in each county from the U.S. Department of Agriculture (USDA) and state food stamp agencies;[2] and county estimates from Internal Revenue Service (IRS) data of the number of child exemptions (assumed to be under age 21) in families who reported income below the poverty threshold on their federal income tax return. Neither of these two data sources gives the number of school-age children in poverty as measured by the March CPS or by the census, but this is not a problem for model-based estimation: it is necessary only that the variables chosen to be used in the model can provide good predictions of that number.

For the Food Stamp Program, the total number of recipients is available annually for states and counties, and eligibility requirements are generally uniform across all states (with some exceptions for Alaska and Hawaii). Two key eligibility requirements are that households must have gross income before deductions that is below 130 percent of the applicable poverty guideline and net countable income that is below 100 percent of the applicable guideline.[3] (The gross and net income limits for eligibility and the ceilings on allowable deductions are higher in Alaska and Hawaii than in the other states due to their higher cost of living.) Although the program is generally administered uniformly across all states, participation rates—the proportion of eligible households that apply for and receive benefits—are not the same. Also, the information obtained for each county is not always the same: in most counties, the counts of food stamp recipients pertain to July; for some counties, they are an average of the monthly counts for the year.

Information from federal income tax returns can be used to construct family units and to compare the income of such units with the applicable poverty threshold. Individual tax returns are assigned to counties on the basis of their address information. There are three major advantages of data from tax returns: (1) coverage of a very large proportion of the population, (2) coverage of a very large proportion of the income received by families, and (3) data availability on an annual basis.

The number of child exemptions reported on tax returns for families with incomes below the poverty threshold, like the number of food stamp recipients, is

[2]USDA counts of food stamp recipients were not complete for all counties; the Census Bureau contacted individual state agencies to obtain missing information.

[3]The poverty guidelines used for determining program eligibility are derived by smoothing the official poverty thresholds for families of different sizes (see Fisher, 1992).

an imperfect measure of poverty for school-age children. Not all people file tax returns, especially those with very low incomes or income mostly from nontaxable sources. In addition, "income" as defined on tax returns does not include all the sources of income that are used in the official measure of poverty, and tax filing units are not totally consistent with the Census Bureau's definition of families. Moreover, the address on a tax return does not always correspond to a filer's residential address. Nonetheless, tax information, like counts of food stamp recipients, is a useful variable to develop predictions of poverty for school-age children.

Model Specification

The second step in developing a model-based estimate of the number of school-age children in poverty by county is to specify and estimate a formula, or prediction equation, that relates the administrative data and other "predictor" variables to the dependent or "outcome" variable, which is an estimate of the number of school-age children in poverty from the March CPS. The CPS estimate is chosen as the object of prediction because the CPS provides the largest and most up-to-date data set that is available with which to estimate poverty among school-age children.

A key decision in the specification of the county-level model was to use the CPS estimate of the *number* of school-age children in poverty as the dependent variable.[4] Another choice would have been to model the *proportion* of school-age children who were poor and then convert the estimated proportions to estimated numbers of poor school-age children. (This approach was in fact adopted for the state-level model—see below.) The Census Bureau decided against the approach of estimating proportions and converting them to numbers because of a concern that the county population estimates of school-age children that would form the basis for converting the estimated proportions to numbers were of uncertain quality. Hence, it would be difficult to construct estimates of the precision of the estimated numbers of poor school-age children, which play the most important role in the Title I allocation formula.

Another key decision was to estimate the number of school-age children in each county who were poor at a particular time (e.g., the number in 1994 who were poor in 1993) and not to estimate the change in the number since the 1990 census. This decision reflects the Census Bureau's conclusion that the available administrative data were likely measured more consistently across areas at a given time than they would be over time, given changes in tax and transfer program rules. Both of these decisions are discussed further in Section 4, which presents the panel's assessment of the Census Bureau's methods.

[4]As noted in Section 1, for this application school-age children are defined as related children aged 5-17.

The county-level model includes five predictor variables. In addition to the two variables described above, the number of food stamp recipients and the estimated number of child exemptions reported by families in poverty on tax returns, the variables are: the total number of child exemptions on tax returns, the total population under age 21 from the Census Bureau's postcensal population estimates program,[5] and the number of school-age children in poverty from the most recent census.

In the county-level model, the CPS estimate is a 3-year centered average, and all the variables are measured on a logarithmic scale. A reason to use logarithms was the wide variation in the CPS estimate and the values of the predictor variables among counties: transforming the variables to logarithms made their distributions more symmetric and the relationships between some of them and the dependent variable more linear. A reason for the decision to combine 3 years of CPS data for county estimation and thereby improve precision was the small CPS sample sizes in individual counties. (For the 1993 county-level model, the CPS estimate is an average of data from the March 1993, 1994, and 1995 CPS, representing measured poverty in 1992, 1993 and 1994.) Given that only a subset of counties is represented in the March CPS sample, the relationships between the predictor variables and the dependent variable in the model are estimated solely on this subset of counties. This subset includes proportionately more large counties and proportionately fewer small counties than the distribution of all counties.[6]

By calculating the relationships among the predictor variables and the CPS estimates of school-age children in poverty for the subset of counties that have households in the March CPS sample with poor school-age children, it is possible to obtain a good estimate of an equation for predicting the number of poor school-age children in a county, even though the CPS estimate for any specific county has a measurable level of uncertainty that is large for many small counties.[7] The prediction equation can then be used to predict the number of school-age children

[5]The population estimates for people under age 21 are the estimated resident population under age 21 derived from demographic analysis minus the estimated population in institutions and military barracks for that age group; see Appendix D. The estimates pertain to July 1 following the income year (e.g., July 1994 for the 1993 model). Including the estimated population under age 21 and the estimated number of total child exemptions on income tax returns as variables in the model is intended to provide a measure of the number of people not covered on tax returns, most of whom are at the low end of the income distribution.

[6]Because values of 0 cannot be transformed into logarithms, a number of counties whose sampled households contain no poor school-age children are excluded from the estimation; see Section 4 for discussion.

[7]The regression coefficients on the predictor variables that express the relationships with the dependent variable in the county-level model are estimated using weighted least squares. The weights used are the reciprocal of the sum of the estimated sampling variance of the logarithm of the number of poor school-age children in a given county plus the estimated variance of model error, assumed to be constant across counties; see Appendix C.

in poverty from the food stamp, IRS, population estimates, and census predictor variables for each county, whether or not the county is in the March CPS sample.

For counties that have households with poor school-age children in the March CPS sample, a weighted average of the model prediction and the estimate based on data from the sample households (the direct estimate) is used to produce an estimate for that county. The weights that are given to the model prediction and the direct estimate depend on their relative precision. (See Appendix C for how these weights are derived.) For a county with very few sample households in the CPS and hence a high level of sampling variability in the direct estimate, most of the weight will be given to the model prediction and little to the direct estimate. For a county with a large number of sample households in the CPS, more weight will be given to the direct estimate and less to the model prediction.[8] In either case, assuming that the weights have been well estimated, the combined estimate of the number of school-age children in poverty will be at least as accurate as the better of the separate predictions (from the model or the CPS). For counties that lack households with poor school-age children in the CPS sample, the prediction from the model is the estimate. In both cases, after first transforming the logarithmic values back to numbers, the county estimates are adjusted for consistency with the estimates from the state-level model, as described below.

STATE-LEVEL MODEL

The Census Bureau's state-level model for estimates of poverty among school-age children is similar in general approach to the county-level model. However, it differs in a number of respects:

- The state-level model uses as the dependent variable the proportion of school-age children in poverty: that is, the dependent variable is a poverty ratio rather than the number of poor school-age children, as in the county-level model.[9] The numerator for the ratio is the CPS estimate of poor school-age children in a state (i.e., the estimate of the number of poor related children aged 5-17); the denominator is the CPS estimate of the total number of noninstitutionalized children aged 5-17 in the state.[10]

[8]The variation in the difference between the model prediction and the actual number of school-age children in poverty is assumed to be the same, on a proportional basis, for all counties with households in the March CPS sample. This difference is termed model error: as used in statistics, "error" is the inevitable discrepancy between the truth and an estimate due to variability in measurements and the fact that modeled relationships are not precise.

[9]The predicted variable is termed a ratio because the denominator is not exactly the same as that for the official published poverty rates.

[10]A different denominator—noninstitutionalized school-age children rather than the slightly smaller universe of related school-age children—is used for consistency with the denominator that is used to convert the estimated poverty ratios to estimated numbers of poor school-age children.

- The state-level model uses four predictor variables for each state: the estimated percentage of child exemptions in families who reported incomes below the poverty threshold on their federal income tax return; the estimated percentage of people under age 65 who did not file an income tax return;[11] the percentage of the population that received food stamps; and the residuals from a regression of poverty rates for school-age children from the prior decennial census on the other three independent variables (see Appendix C).
- The dependent variable in the state-level model is derived from 1 year of CPS data (the March 1994 CPS for the 1993 model), rather than a 3-year centered average as in the county-level model. This decision assumes that the sample size for states permits estimating the model with reasonable accuracy and, implicitly, that it is preferable when possible to have estimates that pertain directly to the income year.
- Since all the variables in the state-level model are proportions rather than numbers, they need not be transformed to a logarithmic scale as is done with the numbers in the county-level model. Such a transformation is not needed because the distributions of the estimated proportions for the predictor variables are more symmetric and have a more linear relationship with the dependent variable than is the case for the distributions of the estimated numbers.

All states have sample households in the CPS; however, the variability associated with estimates from the CPS is large for some states. As is done in the county-level model, the state-level model weights the direct estimate for a state and the estimate from the model according to their relative precision to produce estimates of the proportion of poor school-age children in each state. To produce estimates of the number of poor school-age children in each state, the estimates of the proportion poor from the model are multiplied by estimates of the total number of noninstitutionalized school-age children. For the 1993 model, these estimates are derived from the Census Bureau's program of demographic population estimates (see Section 4 and Appendix D). Finally, the state estimates of the number of poor school-age children are adjusted to total the CPS national estimate of school-age children in poverty. The national estimate pertains to related children aged 5-17 so that, at this final stage, the state estimates are consistent with the county estimates in that both sets represent estimates of the numbers of related children aged 5-17 in poverty.

ADJUSTMENT OF COUNTY ESTIMATES TO STATE CONTROLS

The county-level model described above produces an initial set of estimates of the number of poor school-age children in each county in the United States.

[11]This percentage is obtained by subtracting the estimated number of exemptions for people under age 65 on income tax returns from the estimated total population under age 65 derived from demographic analysis; see Appendix D.

The final estimates for counties are produced by "benchmarking" the initial county estimates to the final adjusted state estimates: for each state, the estimate for every county in that state is multiplied by a constant factor to make the sum of the resulting county estimates equal the state estimate. For example, if the estimated state total is 5 percent higher than the sum of the county estimates for that state, the estimate for each county in that state is multiplied by 1.05. If the estimated state total is 5 percent lower than the sum of the county estimates for that state, the estimate for each county in that state is multiplied by 0.95. The rationale for this last step is that the state estimates are more reliable because they are based on more data (larger samples) than are available for most counties. For example, if the county-level model tends to underpredict for counties in a particular state, the state as a whole is not affected by that error because its total is determined by the state-level model.

The county-level model predicts the *number* of school-age children in poverty. Estimates of county poverty *rates* for school-age children, which play an important but secondary role in the Title I allocation formula, are obtained by dividing the estimated number of school-age children in poverty from the county-level model by an updated estimate of the county noninstitutionalized population aged 5-17, adjusted to represent related school-age children. These estimates are produced from the Census Bureau's population estimates program (see Appendix D).

The county and state estimation procedures described in this section are based on the CPS. Therefore, the county estimates represent estimates of poverty for school-age children as measured by the March CPS, not as measured by the decennial census. The issues raised by this shift in the underlying source of data for the estimates are considered in Section 4.

4

Panel Assessment of the Methodology

Any set of model-based estimates requires thorough evaluation of the assumptions underlying the model, the quality of the input data, the variability of the resulting estimates, and other features of the estimation procedure. For the purposes of Title I allocations, the primary concern is with the quality of the estimates of poverty among school-age children for counties. Thus, the discussion in this section focuses largely on the county-level model, but it also considers the state-level model and the Census Bureau's population estimates, both of which enter into the final county estimates.

The ideal evaluation of the Census Bureau's (or any) methodology for estimating the number of poor school-age children for counties would start by comparing 1993 estimates from the model-based procedure to the true numbers of poor school-age children for some or all counties in 1993 (or, at least, to measurements known to be highly accurate). One could then determine how close the estimates are to the "true" values. Unfortunately, the truth is not known, and no measurements known to be highly accurate are available. Because the ideal evaluation is impossible, the Census Bureau and the panel have approached the problem of evaluation from a number of different directions. Although no single evaluation is conclusive, the various evaluations have enabled us to form preliminary conclusions, which serve as the basis of our recommendation, about the degree of confidence we can have in various parts of the estimation procedure and the final product. The development of model-based estimates for counties is a major research effort for which extensive evaluation is required. Our conclusions are preliminary because the Census Bureau has not yet had time to conduct all of the assessments that the panel believes are necessary to fully evaluate the quality of the estimates and the suitability of the selected model.

The assessments that should be done include additional evaluations of the current procedures, the development of other competing models, and comparisons of the performance of the Census Bureau's models (both county- and state-level models) with other, similar models. Such models include a county-level model that predicts rates instead of numbers of school-age children in poverty; models that predict change in poverty over time or that use change-related predictor variables (e.g., changes in the number or proportion of child exemptions reported by families in poverty on tax returns); models that include additional predictor variables constructed from the available data; and models that allow a more flexible approach by using such statistical estimation procedures as generalized linear modeling. The Census Bureau has begun work to estimate and evaluate some alternative models, but the work is incomplete (see further discussion in Section 6).

In the rest of this section, we present a nontechnical summary of the evaluations of the county-level and state-level models that have been conducted to date (see also Appendix E; Coder et al., 1996). We first consider several approaches to statistical evaluation of the county- and state-level models, beginning with consideration of the reasonableness of the form (specification) of the models used—both the variables included in the models and the mathematical formulas that are used to express the relationships among the variables. We next consider standard statistical tests to show the significance of the relationships between the predictor variables used in the model and the dependent variable. We also consider the relationship between the state- and county-level models. We describe the evidence on the performance of the Census Bureau's estimation procedure when used to estimate school-age children in poverty in 1989 instead of 1993. We call that the "1989 model" to distinguish it from the "1993 model" that estimates school-age children in poverty in 1993. (The estimation procedure is applied to 1989 because the latest available decennial census poverty estimates pertain to 1989 and can be used as a standard of comparison.) We also examine the presence of systematic under- or overestimation for particular groups of counties. We then consider the implications for the estimates of the fact that the March CPS and the decennial census represent somewhat different approaches to measuring poverty, using the same definition but not the same data collection or estimation procedures. Lastly, we consider the reliability of the postcensal estimates of population that are used in the estimation process.

MODEL SPECIFICATION

The use of a model-based approach to estimation is one the panel fully supports. Because of the nature of the available data, there is no alternative at this time to the use of models to develop poverty estimates for intercensal years. The decision to use a weighted combination of model and direct estimates for each area is widely accepted as the appropriate practice for small-area estimation prob-

lems of this type (see, e.g., Platek et al., 1987). However, several of the more detailed decisions about the Census Bureau's modeling strategy are not necessarily widely accepted or the best possible choices.

We find the specification decisions made in the state-level model relatively straightforward: to use certain predictor variables and not others, to derive the dependent variable from 1 year (not 3 years) of March CPS data, to use a linear model, to express the model in terms of rates (or ratios) rather than absolute numbers, and to use a particular model to smooth the variance estimates for state data. However, it would nonetheless be useful to develop some alternative models and compare them with the selected model. We were interested in whether a model in which all variables were entered as changes from decennial census year to model year would be more accurate. We believe, however, that the effects that would be captured by this approach are at least in part represented in the Census Bureau's model through inclusion of the residual from a 1989 model as a predictor in the 1993 model.

We have more concerns about the specification of the county-level model. An early decision was made to specify this model in terms of the number of school-age children in poverty, rather than in terms of a poverty rate or ratio. The expressed rationale for this decision is that, although there are postcensal estimates of the school-age population by county with which to connect the estimated rates to estimated numbers of poor school-age children, there are no variance estimates for these population estimates (Coder et al., 1996). The consequence of this decision is that changes in the number of children in poverty due to changing poverty rates and due to changing overall population growth (or decline) are all captured in the same regression model; postcensal population estimates affect the estimates only as a predictor variable in the model, just like the number of food stamp recipients and other variables. We are not confident that this approach gives the most precise estimates for counties.

The decision to specify the model in terms of numbers also necessitated the use of a loglinear model, that is, one in which all variables are transformed to a logarithmic scale and the relationship between these transformed variables is assumed to be linear. The properties of this model are more complicated than those of the linear model used at the state level. In particular, as noted above, a number of counties represented in the CPS had no children in poverty in the sample households; these counties were simply deleted from the regression computation, as their data could not be transformed to the logarithmic scale.[1] Although some calculations suggest that the magnitude of the effect of these deletions on the estimates may not be very large, we would have more confidence in a modeling strategy that did not require this exclusion of part of the data. Statistical methods that avoid this problem are available and preferred.

[1]Of the 1,529 counties in the March 1993, 1994, and 1995 CPS data used for the 1993 model, 345 were deleted for this reason; most of these counties had very small samples.

There are two other major concerns about the specification of the county-level model (analyzed in more detail in Appendix E).[2] One is that the model is almost entirely cross-sectional, making no use of variables that measure changes in poverty or program participation since the decennial census base year. The rationale for this specification is that the administrative data used in the model are more consistently measured across different areas than across time. It is certainly true that changes in tax and transfer program rules will affect the comparability of administrative data over time, but differences in program participation rates and administration may affect comparability across areas.[3] Although the quantitative evaluation summaries we have seen of variations of the Census Bureau's basic model did not demonstrate superior accuracy for models that used more change variables, the fact that a type of change variable was important in the state-level model makes us believe that it would have been helpful in some form in the county-level model as well. In addition, a 1989 model was deemed essential for evaluation purposes, but it was not possible to fit a change model for that year because of the lack of historical data on some key variables.

Another concern is that the county-level model does not scale up uniformly with county size: for example, if one doubles the size of the county while keeping its composition the same, so that the population estimates, the number of food stamp recipients, the number of child exemptions reported by families in poverty on tax returns, and the other count variables are all doubled, the predicted number of children in poverty more than doubles. This effect is quite substantial over the wide range of sizes of county populations. This feature of the model may reflect some real trends with respect to county size (see Appendix E). Thus, large counties may tend to have some characteristics, not included in the model, that are associated with higher poverty rates than are predicted from food stamp, income tax return, and prior census poverty estimates alone. What these variables might be is not known, and it may be difficult to include additional variables in the model, given the lack of suitable data from other sources. However, some work could be done to analyze the characteristics of counties for which a model that scales up uniformly with county size (achieved by constraining the coefficients of the predictor variables to sum to 1) produces better or worse estimates than the Census Bureau's unconstrained model.[4] It is possible that such work would

[2]A more minor concern is that the variables in the county-level model are highly correlated, which can be a problem given that the county-level model does not use a representative set of counties from the CPS (see Appendix E).

[3]Such differences may become more prevalent in the future with increasing devolution of program responsibility to the states.

[4]The Census Bureau's unconstrained model and a constrained model for 1989 were estimated and the resulting estimates compared with 1990 census estimates of the number of school-age children in poverty as part of a broader evaluation for 1989 (see below; see also Appendix E). However, there has been no analysis as yet of types of counties for which the constrained model performed better or worse.

identify additional variables that it would be useful to include, such as interactions of variables already in the model or characteristics that could be obtained from census data—such as whether the county experienced high population growth or includes a central city.

In summary, the panel is relatively satisfied with the specification of the state-level model, although it should be further evaluated. We are less satisfied with and would have liked to pursue more alternatives for the county-level model. In fairness to the county-level effort, we recognize that the availability and accuracy of data at the county level are limited in comparison with what are available for states, while the degree of variation to be explained for counties is greater. It is therefore not surprising that the specification of the model at this level is more challenging and controversial.

EVALUATIONS OF MODEL SPECIFICATION

Formal statistical testing to assess the significance of the predictor variables in the models has been complicated by the complexity of the components of the county-level model. A formal hypothesis test performed for the state-level model (Fay, 1996) supported the conclusion that the state-level regression model using administrative records data improved on the estimates of poverty rates for 1993 that could be obtained by using only 1989 poverty rates from the decennial census.[5] Thus, this test provided evidence to support basing estimates on a statistical model rather than on decennial census data. The Census Bureau also conducted statistical hypothesis tests to show the statistical significance of the predictor variables in the county-level model, but the same type of hypothesis test that was used for the state-level model to demonstrate the superiority of the model-based estimates was not performed.

CONSISTENCY OF STATE- AND COUNTY-LEVEL MODELS

The panel compared estimates of the number of school-age children in poverty by state that were obtained directly from the state-level model to those obtained by adding within each state the estimates from the county-level model before they were calibrated to match the state estimates. If the ratio of state estimates from the state-level model to state estimates aggregated from the county-level model were the same in every state, this result would indicate that the county-level model captures all of the effects captured by the state-level model, making the latter model superfluous. We did not find this result. Rather,

[5]The test assumes that the objective is to predict poverty rates that reflect the CPS measurement of poverty and not the decennial census measurement. As discussed in Section 2, the CPS and census do not provide the same measures of poverty because of differences in data collection methods and other features (see also Appendix B).

we found substantial variability in this ratio, from 0.822 to 1.332, with one-half the states lying between 1.017 and 1.128 (see also Table E-1).[6] This finding suggests that there may be substantial state effects that are not captured by the county-level model.

We would like to determine whether a county-level model that also incorporates state effects will give substantially different estimates from those obtained from the present two-stage approach, but the data and analysis are not now available for such a determination. More generally, we would like to investigate the possible benefits (including evaluation opportunities) of approaches that provide for greater consistency in the specification of the state- and county-level models (e.g., in the predictor variables that are included and in the specification of the dependent variable).

ACCURACY OF 1989 PREDICTIONS

Although one cannot compare 1993 model-based county estimates to any measure of "truth" for 1993, one can make a comparison for 1989 by estimating the model for 1989 and comparing the results with the 1990 census. One can also compare the performance of the 1989 model estimates in predicting the 1990 census with the performance of simpler models that rely entirely, or much more heavily, on data from the 1980 census. This comparison is relevant because, in the absence of model-based estimates, the estimates from the last census have been used for Title I allocations. Such a comparison can also help clarify the advantages and disadvantages of the Census Bureau's model.

We compared county estimates from the 1990 census of the number of school-age children in 1990 who were poor in 1989 with county estimates for 1989 from four models, described below. Because Title I funds are distributed from a fixed budget and therefore are largely driven by the share of the national total of poor school-age children in each county (rather than the absolute number of poor school-age children in each county), we adjusted the counts from each model proportionally to match the 1990 census estimated national total number of school-age children in poverty in 1989.

Model 1 is the Census Bureau's county-level model estimated for 1989. In this model, the dependent variable is the 3-year centered average of poor school-age children from the March 1989, 1990, and 1991 CPS (for income years 1988, 1989, and 1990), and the predictor variables are the 1989 counts of food stamp

[6]Our comparison of the state estimates from the state- and county-level models for 1993 was performed with data that were made available to the panel in January 1997. Subsequently, the Census Bureau discovered errors in the input data for a few counties that somewhat changed the 1993 estimates from the county-level model. (The state estimates were unchanged.) However, the general findings hold true that the use of control totals from the state-level model results in large adjustments to the estimates for counties in some states and that the adjustments vary widely across states.

recipients, the 1989 estimates of child exemptions reported by families in poverty on tax returns and total child exemptions reported on tax returns, the 1990 census estimates of the population under age 21, and the 1980 census estimates of the number of poor school-age children. The county estimates are controlled to state estimates from the Census Bureau's state-level model for 1989.

Model 2 is the 1989 county-level model (Model 1) that is not controlled to state totals.

Model 3 is a model in which the 1980 census estimates of the number of poor school-age children are updated to reflect the change in the total number of school-age children in each county from 1980 to 1990.

Model 4 is a model in which the 1980 census estimates (Model 3) are used without any updating (except, as in the other methods, to adjust all of the estimates proportionally so that they equal the 1990 census national total estimate of poor school-age children). In other words, Model 4 assumes that the distribution of poor school-age children by county did not change between 1980 and 1990.

As a summary measure of accuracy, we used the average of the absolute differences between estimates from each model and the 1990 census estimates of school-age children in poverty by county for 1989. Our comparison of Model 4 (the 1980 census estimates) with the 1990 census gave an average absolute error of 543 school-age children in poverty per county, which is almost one-fourth of the 1990 census average estimate of about 2,400 school-age children in poverty per county in 1989. When we used Model 3 (updating for the change in county population), the average absolute error was 415. Using Model 2 (the 1989 county-level model without controls to state totals), the average error was 292. The prediction from Model 1 (full Census Bureau 1989 county-level model) had an average error of 270. These results suggest that updating the previous census for population shifts improves the estimates (comparing Model 3 with Model 4), but that the greatest benefit comes from the use of the county-level model (Model 1), which captures both population shifts and changes in the incidence of poverty.[7]

These results provide evidence in favor of the model-based approach and against using the estimates from the previous census. However, we note that such a comparison might turn out differently if one had a measure of truth and could perform the comparison for 1993 instead of 1989[8] because the shorter interval (4

[7]Our analysis of estimates of the number of school-age children who were in poverty in 1989 (described here and the next part of Section 4 and in Appendix E) was performed with data that were made available to the panel in fall 1996. Subsequently, the Census Bureau discovered errors in the input data for a few counties that somewhat changed the estimates from its 1989 county-level model. However, the general findings continue to hold.

[8]In this instance, Models 1 and 2 would represent the Census Bureau's 1993 county-level model adjusted and not adjusted, respectively, to the estimates from the 1993 state-level model; Models 3 and 4 would be based on the 1990 census, updated and not updated, respectively, for changes from 1990 to 1994 in the total school-age population of each county.

years versus 10) and the particular patterns of change during the two time periods affect the comparison unpredictably.

The 1989 comparison represents only one replication. A similar comparison in which the 1980 census would provide the measure of "truth" was not possible because the administrative data required to estimate a county-level model of the number of school-age children in 1980 who were poor in 1979 were not all available. However, a comparison could be developed in which the 1990 census provides the measure of "truth" and the model estimates that are compared with it include the results of the 1989 county-level model and the results of estimating the county-level model for other years in the 1980s (say, 1985 or 1986). Such comparisons could provide insight into the effects of differing time lags on the quality of the model predictions. The Census Bureau has begun work to compare the 1990 census estimates with the results from the Census Bureau's 1989 county-level model and the results from alternative models for 1989 (e.g., a county-level model that estimates rates). Such work, which needs to be completed, can help determine if there is a specification for the county-level model that performs appreciably better than the selected specification.

SYSTEMATIC UNDER- OR OVERESTIMATION FOR GROUPS OF COUNTIES

Another important assessment is whether the model-based estimates tend to systematically over- or underestimate the number of school-age children in poverty for groups of counties with particular characteristics, a distinct question from the differences between model and truth for individual counties that are considered in the preceding section. Systematic error (bias) is important for two reasons. First, it harms or helps certain types of counties and, therefore, certain groups of people on the basis of their characteristics; in contrast, predictions that are reasonably free of systematic error avoid harming or helping identifiable groups. There is still random error, but its effect on any particular area is unpredictable and, to some extent, may average out over time. Second, identifying systematic error indicates how to improve a model (e.g., by including the specific characteristics in the model as predictors) and, thus, how to obtain more accurate estimates.

For groups of counties, we compared the model estimates to census estimates and CPS direct estimates for 1989, as well as the model estimates to CPS direct estimates for 1993, averaging each type of estimate across member counties in each group. (Because we averaged over large enough groups of counties, the CPS sampling error is not excessively large.) Statistical hypothesis tests suggest that there is a tendency for the model to underpredict for counties with smaller populations, although the pattern is not completely consistent for the years (1989 and 1993) we examined. Tests involving a few other groupings of counties (e.g., by metropolitan/nonmetropolitan status) showed no consistent trends.

The panel is concerned about the effect of a possible bias by county size because the Census Bureau's estimates of school-age children in poverty in 1993 show a shift toward larger counties, compared with the distribution in 1989 (as measured by the 1990 census).[9] This shift reflects some combination of actual changes between 1989 and 1993, differences between the CPS and decennial census measurements of poverty, and effects due to the properties of the model. The panel has so far only conducted a preliminary investigation to separate these three components, which has not enabled us to draw any firm conclusions about their relative importance; this issue merits further investigation.

EFFECTS OF DIFFERENCES BETWEEN CENSUS AND CPS POVERTY MEASURES

We are concerned that there may be subtle but important differences between census and CPS poverty measures, and, hence, we are concerned about switching from one measure to the other for Title I allocations before these differences have been more thoroughly studied. As noted above, the county-level model estimates poverty as measured by the March CPS, which represents a somewhat different measurement from that of the decennial census (see Appendix B). Qualitatively, it is plausible to expect that there would be differences between the CPS and census measures of poverty due to the many differences in data collection; quantitatively, however, it is not easy to show what systematic differences exist between the two measures.

There are statistically significant differences between the March CPS (3-year average) and 1990 census estimates of the number of school-age children in poverty in 1989 for the nation as a whole and for groups of counties characterized by metropolitan status, region, and population size. For most categories, the CPS estimates of the number of poor school-age children are larger than the census estimates (see Table 2-4). Also, the pattern of differences by size of county suggests that the ratios of the CPS estimates to the census estimates may exhibit systematic differences; specifically, that these ratios may increase as a function of county size. Yet preliminary tests by the panel have not been able to establish statistically significant differences between the CPS-census ratios among groups of counties defined by population size and other characteristics. (As an example, there does not appear to be a statistically significant difference between the CPS-census ratio for counties with population size 100,000-499,999 population and the ratios for counties with smaller or larger population sizes.) However, we regard these analyses as very preliminary.

It is important to note that the evaluations described above of the 1989 estimates of school-age children in poverty use the 1990 census as the "gold stan-

[9]These comparisons were performed with the county estimates that were provided to the panel on January 7, 1997.

dard." Therefore, they provide valid evidence in favor of the CPS-based model even if one takes the census as the standard for measurement of poverty. Presumably, if the CPS (which is currently the basis for official poverty estimates) were treated as the standard, the CPS-based model would be evaluated still more favorably.

USE OF POSTCENSAL POPULATION ESTIMATES

The process for estimating school-age children in poverty at the county level and the Title I allocation formulas for using those estimates require population totals by age in noncensus years for two purposes: as a variable in the county-level model regression equation (population under age 21) and as the denominator (population aged 5-17) of the poverty rate in the allocation formula.[10] Population totals by age are also required for the state-level model. The population totals by age must be estimated and, as estimates, are subject to errors. (See Appendix D for a description of the Census Bureau's population estimates program, which uses demographic analyses to update the previous census.) The variability of the estimated poverty rates of school-age children that is attributable to error in the denominator is not modeled explicitly by the Census Bureau. Analyzing this variability is difficult and has not yet been done.

The Census Bureau does have an active program to develop and review the performance of its population estimates, including evaluating the estimates at 10-year intervals by comparing them with decennial census figures. These comparisons provide an indication of the errors, but they are not complete measures of accuracy and precision because the standard (i.e., the decennial census) itself is flawed, notably from net population undercount, which varies by age group across time and place (see Robinson et al., 1993).

The Census Bureau's methods for producing postcensal population estimates have generally improved over time, but three features continue to apply to the county and state estimates. First, the relative errors are larger on average for small areas than for large ones. (This relation is practically inevitable.) Second, the relative errors tend to be larger for areas in which the population is changing rapidly than for areas that are more stable. Third, the relative errors for age groups tend to be higher than those for the total population.

For county estimates of the total population, the average absolute percentage error improved between 1980 and 1990: in 1980, it was 4.1 percent unweighted and 3.1 percent weighted (by size of county); in 1990, it was 3.6 percent un-

[10]More precisely, the denominator of the poverty rate in the allocation formula is the estimated population of related children aged 5-17 in each county. These estimates are developed by adjusting the estimates from the Census Bureau's population estimates program for the noninstitutionalized population aged 5-17 (see Appendix D).

weighted and 2.3 percent weighted. Population size markedly affects the accuracy of the estimates. For counties with populations of more than 100,000 in 1980, the average absolute error in the 1990 estimate was 2.0 percent, but it was 4.6 percent for counties with populations of 2,500-5,000 and 7.7 percent for the smallest counties, those with less than 2,500 population.

Growth rate is another important factor in the errors. The average absolute percentage error for the fastest growing counties (25% growth between 1980 and 1990) was 4.9 percent in 1990. For counties that grew very little (0-5%), the error was only 3.0 percent. The pattern is not monotonic: counties that lost population had a somewhat larger average absolute error, 3.5 percent (Davis, 1994).

The Census Bureau has not yet completed its evaluation of county-level estimates for the population group aged 5-17 and the one under age 21. It is therefore difficult to assess the effect of the errors in these estimates on the updated estimates of school-age children in poverty or on the corresponding poverty rates. However, preliminary evaluations (Bureau of the Census, private communication) indicate that the average absolute error for the group aged 5-17 was 6.4 percent across all counties when estimates for 1990 that were derived from updating the 1980 census figures are compared with the 1990 census figures.

5

Recommendation for Title I Allocations

The Panel on Estimates of Poverty for Small Geographic Areas has examined the Census Bureau's assumptions, methodology, and evaluation tests in developing the 1993 county estimates that were produced for the Department of Education to consider in allocating Title I funds for the 1997-1998 and 1998-1999 school years. The panel focused its attention on the advisability of using those estimates for determining fiscal 1997 allocations to counties. It is the panel's assessment that considerable progress has been made in developing model-based estimates of school-age children in poverty for states and counties. The panel recognizes the efforts to produce these results and commends the Census Bureau for its work in developing these estimates.

The alternative to the use of model-based estimates for Title I allocations is the use of 1990 census estimates of school-age children who were poor in 1989. As we discuss in Section 2, the census estimates are seriously out of date. The major changes in the distribution of children in poverty that have occurred since the 1990 census make the use of those estimates for current allocations highly problematic. There was a substantial increase in the number of children in poverty between 1989 and 1993, and there is clear evidence that the geographic distribution of such children has changed markedly. Thus, the census estimates are subject to sizable error if used for current allocations. In addition, the census estimates are based on sample data and are therefore subject to sampling error. The magnitude of the sampling error is fairly large for many small counties (see Table 2-1).

The Census Bureau's model-based estimates have a clear advantage of being more up to date. However, they are also subject to error. Such error is inevitable and is not a reason for rejecting model-based estimates. Provided that the errors

are not too large and, after thorough investigation, are not found to exhibit any specific patterns, the panel would strongly endorse the use of those model-based estimates.

The development of the state-level model is an extension of work conducted and tested by the Census Bureau over many years, and the task of producing reliable estimates for states is less challenging than it is for counties. The panel considers the model-based estimates for states to be clearly preferable to the census estimates, although further work to evaluate the state-level model is desirable. The county estimates are constrained to sum to the state totals, and, as discussed in the previous section, some of the evaluations that have been conducted of the county estimates provide evidence that their use for Title I allocations may be preferable to continued use of the 1990 census estimates. However, other evaluations of the county estimates, as well as the form of the model, raise concern.

In general, the panel believes that the county-level model and resulting estimates have not yet been sufficiently evaluated, especially in comparison with alternative models, for use without modification for such an important purpose as allocating funds under Title I. The necessary evaluations need to be completed before the panel would be confident in basing current Title I allocations solely on the 1993 model-based county estimates. Further research may also lead to improvements in the county-level model (see Section 6).

The panel is also concerned about the effect on the allocations caused by the difference between poverty estimates obtained from the census and from the CPS. In the aggregate, the CPS produces estimates of children in poverty that are about 7 percent larger than estimates from the census, but little is known about how the differences between the estimates from the two sources may affect the relative size of the estimates for different types of counties.

In time, many questions pertaining to model-based estimates can be answered. And we can expect a better understanding of the reliability of these estimates, of the effects of the differences between the census and the CPS, and of the advantages of some models over others. And alternative models, which appear to be equally viable, can be more fully explored. But the panel is faced with making a recommendation now to the Secretaries of Commerce and Education on whether to use the Census Bureau's model-based county estimates for allocating Title I funds for the 1997-1998 school year. For the current allocations, there is no time to complete the research needed to address the questions that the panel believes need to be answered.

For the purpose of the Title I allocations, the problem is one of weighing the panel's concerns about the more up-to-date model-based estimates for 1993 against its concerns about using out-of-date decennial census estimates. To the panel, the Census Bureau's model-based county estimates appear, in many regards, preferable to the out-of-date decennial census estimates. The panel concludes, however, that a solution that takes advantage of the Census Bureau's work on model-based estimates but reduces the impact of possible limitations in

that work is the most appropriate approach for the immediate purpose of Title I allocations.

The panel's solution to this problem is to recommend that the Department of Education use an average of the two estimates for each county, as described below. This procedure is recommended only for the current Title I allocations and not for any other purpose. The rationale for our recommendation is that the averaging will take some account of the changing number and geographic distribution of children in poverty provided by the model-based estimates but will dampen any distortions that might occur with these estimates. (As an example, our recommendation will likely reduce the effects of any possible bias in the model-based estimates for different sized counties.) It will also dampen the effect of the differences between the census and the CPS poverty estimates. The procedure can be viewed as smoothing the transition to the model-based estimates for subsequent allocations, after further research has been conducted.

RECOMMENDATION AND AVERAGING PROCEDURE

The panel recommends to the Secretaries of Commerce and Education that funds under Title I for fiscal 1997 be allocated on the basis of estimates that are obtained by averaging two poverty rates and then applying the average rate to the 1994 population estimate.

The panel recommends the following precise form of averaging: take the average of the poverty rates for school-age children (i.e., related children aged 5-17) that are obtained from the census and from the model-based estimates for a given county and multiply this average rate by the Census Bureau's 1994 population estimate of the number of children aged 5-17 in that county as adjusted to represent related children aged 5-17 (see Appendix D). The number resulting from this calculation is the estimated number of related children aged 5-17 in poverty in the county. Expressed algebraically, for each county, use:

$$r = \left[\frac{a}{b} + \frac{c}{d}\right] / 2$$

as the rate, r, of school-age children in poverty, where a is the 1990 census estimate of the number of related children aged 5-17 who were poor in 1989, b is the number of related children aged 5-17 in the 1990 census, c is the model-based estimate of the number of related children aged 5-17 in 1994 who were in poverty in 1993, and d is the 1994 population estimate of the number of related children aged 5-17. For the estimated number, n, of school-age children in poverty, use $n = dr$.

The panel stresses that the above averaging procedure is designed to meet the need for an immediate decision on fund allocations. The solution we describe

should be supplanted by a set of fully model-based estimates as improvements in the model and evaluations take place over time.

At this time, however, we believe that averaging the model-based and 1990 census estimates of school-age children in poverty is appropriate for the purpose of fiscal 1997 fund allocations. We note that it is common practice in various areas of applied statistical analysis to average two (or more) estimates with different properties to produce improved estimates. It is, for example, standard practice in small-area estimation to average a direct survey estimate and a model-based indirect estimate to produce a composite estimate for an area (see Ghosh and Rao, 1994). This form of averaging is used by the Census Bureau in both the state- and county-level models. Other examples of its use in federal statistical programs are given in Office of Management and Budget (1993).

In most applications of composite estimation, a weighted average of the estimates is used, with the weights chosen to produce the overall estimate of highest quality. For example, the estimates may be weighted inversely proportional to their mean square errors in order to produce an overall estimate with minimum mean square error.[1] In order to produce such weighted estimates, it is necessary to know the relative quality of the individual estimates. When there is considerable uncertainty about the relative quality, as in the present case, then it is appropriate to use an equal (i.e., 50-50) weighting. As an example, the Census Bureau for many years developed population estimates for states and counties by averaging on an equal basis the results of three independent estimating procedures. The resulting estimates were used in making allocations under the State and Local Fiscal Assistance Act of 1972 (general revenue sharing), the Comprehensive Employment and Training Act of 1972, and the Housing and Urban Development Act of 1972 (Bureau of the Census, 1980; Zitter and Shryock, 1964).

In this situation, we know that both the 1993 model-based and 1990 census estimates are subject to errors from sampling variability and other sources. The 1990 census estimates have relatively low sampling variability (for all but the smallest counties) but are in error to an unknown but presumably large amount in that they are out of date. The 1993 model-based estimates, while more up to date than the 1990 census estimates, are subject to sampling variability and other potential sources of error, which have not yet been fully analyzed.

Given that there is uncertainty about the magnitude of the errors in the two estimates and that there was insufficient time and information with which to analyze the errors, we recommend a simple averaging of the two estimates. This approach is a conservative way of dealing with uncertainty. It takes advantage of the Census Bureau's work to develop more up-to-date estimates of poor school-age children but guards against possible limitations in that work.

---

[1]Mean square error combines the errors due to sampling variability and the errors due to bias in a single statistic.

SPECIAL CASE: PUERTO RICO

Puerto Rico is also included in the Title I allocations. While the commonwealth's 1990 decennial census provides the requisite estimates for 1989, no estimates of Puerto Rican children in poverty can be made for 1993 with the Census Bureau's model because the appropriate food stamp and IRS data are not available for Puerto Rico. The Census Bureau has computed 1993 estimates for Puerto Rico on the basis of an experimental Special Family Income Survey that was conducted in the commonwealth in February and March 1995, but several adjustments had to be made to produce the estimates of school-age children in poverty in 1993.

The approach adopted by the Census Bureau seems a reasonable one given the data available (see Appendix F for details). However, insufficient information was available to the panel about the quality of the underlying survey data. If, after further analysis, it is determined that the quality of the experimental income survey in Puerto Rico is adequate, then the panel would recommend that the Census Bureau's 1993 estimates of poor school-age children in Puerto Rico derived from the experimental survey be treated as equivalent to the 1993 U.S. model-based county estimates in determining the fiscal 1997 Title I allocations. In other words, the 1993 Puerto Rico estimate of the poverty rate for related children aged 5-17 from the experimental survey would be averaged with the 1989 estimate of the poverty rate for related children aged 5-17 from the 1990 census of Puerto Rico. The resulting average rate would be applied to the 1994 population estimate of the number of related children aged 5-17 to obtain the estimated number of poor school-age children in Puerto Rico for use for the fiscal 1997 Title I allocations.

6

Next Steps

The panel commends the Census Bureau for establishing a research program to develop methods to provide county estimates of school-age children in poverty that are more timely than those from the decennial census so that Title I funds can be more appropriately allocated, as Congress intended. The work that has been completed by the Census Bureau makes a strong case that model-based county estimates can be produced that are preferable to the estimates derived from the decennial census.

Though we unequivocally support a model-based approach, more analytical work is needed before we can endorse a specific model for allocating Title I funds. For this important purpose, there should be evidence that the model has been evaluated as fully as possible and that there are no systematic biases that might favor one population group over another. Before users can be confident in the model-based small-area estimates and before the panel can unequivocally recommend them as the sole basis for fund allocation, the behavior of the selected model and the resulting estimates should be well understood, and alternative models should have been thoroughly evaluated.

Evaluation of estimates can typically be separated into two kinds of activities. First, in an "external" evaluation, estimates can be compared with "comparison" values or values that serve as substitutes for true values. The Census Bureau designed its model and expended substantial resources for the purpose of enabling a fair comparison with 1990 decennial census estimates of school-age children in poverty in 1989. Unfortunately, there are no additional comparison values for other time periods that can be used for this purpose, thereby limiting this type of evaluation to 1989.

Second, an "internal" assessment would include, to the extent possible: (1) a comparison of the estimated variances of the updated estimates with variances and estimated biases of the decennial census estimates; (2) an evaluation of the model's assumptions, including the assumptions underlying the above variance estimates; and (3) a determination to the extent possible of the systematic errors (biases) in the model's estimates and how these biases could be reduced. The Census Bureau has performed a number of these internal assessments, including examination of estimated variances for the estimates (developed with variance estimation methods that reflect recent contributions to the literature in this area); residual plots and other techniques to investigate homogeneity assumptions; and regional indicator variables to investigate the possibility of regional biases. A substantial start has been made.

As part of both external and internal evaluations, it is useful to compare the performance of the selected model with alternative models, not only to determine whether the selected model is preferred to the alternatives, but also to gain more understanding of the performance of the selected model. This kind of analysis using alternative models is particularly needed in this case. Some of this activity has been carried out by the Census Bureau for some models that predict change in poverty over time, for models that predict poverty rates, for models that constrain the coefficients of the predictor variables (on the logarithmic scale) to sum to 1, and for simple improvements to the decennial census poverty estimates (e.g., controlling them to the Census Bureau's state estimates of the number of school-age children in poverty); however, more needs to be done to evaluate these and other models in comparison with the selected model.

Finally, more evaluation using comparative analyses with other sources of information is needed. The Census Bureau's estimates indicate substantial changes in the number of school-age children in poverty by counties since 1989. There has been little analysis to determine whether these changes correspond to what is known about those counties, both locally and regionally.

FURTHER EVALUATION OF THE COUNTY-LEVEL MODEL AND ESTIMATES

We outline here some steps that could be taken in the near future to achieve a greater understanding of the properties of the new county estimates and the model that generated them (see Appendix E for more details).

Alternative Modeling Approaches Alternative approaches to modeling the number of children living in poverty could be more fully examined. As noted in Section 3, the Census Bureau adopted a cross-sectional approach rather than a change model in order to avoid difficulties arising from changes in tax and transfer programs. For the time period in question, no substantial programmatic changes occurred. Therefore, it is reasonable to examine more fully a change

model for production of 1993 estimates. The Census Bureau also decided to model poverty levels rather than poverty rates to facilitate estimation of variances of model predictions. However, the panel believes that adequate variance estimates can be produced for a model of poverty rates and that the primary job of producing estimates of the highest quality dominates the need to produce precise estimates of variances. Therefore, we believe that models of rates also need to be examined more fully.

Other kinds of models that could be investigated include models that borrow information from similar counties or close-by years and models that use information on state poverty levels as part of an integrated state-county modeling approach, rather than controlling the county estimates from the model to agree with the state estimates from a different model. Interaction terms for the independent variables and other indicator variables—especially to account for geographic heterogeneity, growth in poverty, or other potential sources of bias—could be fully examined. In addition, bias resulting from the log transformation could be examined through the use of generalized linear models that link estimates of poverty numbers or rates to the predictor variables.

Other Evaluation Activities Five other kinds of further analyses would be useful. First, CPS-census differences could be more fully examined to understand the effect of a change from one measurement system to the other. In addition to studying the quantitative effect of the differences, the qualitative differences between the two measurement systems could also be evaluated. This examination would help to explain the degree to which differences between 1989 model-based estimates and the 1990 decennial census estimates for 1989 were due to CPS-census differences and would help support measurement error models (discussed below).

Second, the weights used in the weighted regressions for the county estimates are based on assumptions concerning model error and sampling variability that could be more fully explored. Although moderate failure of these assumptions may not be a concern, severe failure could result in lowering the quality of the estimates. At least two exercises could be carried out: (1) in validating the county-level estimation of the CPS sampling variances, the directly estimated sampling variances for large counties could be compared to those determined by subtracting out the model error from the county-level model; (2) on the basis of directly estimated variances, models could be developed to estimate sampling variances for all counties, and these estimates could be used to estimate the model error for 1989 and 1993. These estimates could then be compared to the model error from the 1990 census regression (see Appendix E). This method, and the method used by the Census Bureau, should produce similar weights. If they do not, the reason for the difference needs to be understood.

Third, some anomalies could be further examined: (1) The sums of the initial county estimates, before controlling to the state estimates, often differ from

the state estimates to a substantial extent, and these differences should be investigated; (2) When the county-level model is reestimated as a model of poverty rates, the residual is correlated with population size. This feature has a beneficial impact on the model for reasons that are not understood.

Fourth, counties that experienced large changes in the estimated number of children in poverty from 1989 to 1993 could be examined to see whether there are other indications that such changes actually occurred. In addition, for the counties included in the 1989 and 1993 CPS regressions, the counties with the largest residuals (especially from a robust fit) could be subject to further review to see whether there are characteristics that these counties share. This technique is often helpful for identifying additional useful predictor variables in regression models. It is important, however, to note that with any model, there will be some units of observation—in this case, counties—for which the predictions are relatively poor. This fact is simply a property of the methodology and is not a basis for rejecting the results. The finding of several dozen counties with poor predicted values is to be expected and is acceptable, as long as these counties have nothing in common. Random error tends to balance out over time. In contrast, systematic error persists: the primary goal is to root out systematic error, which is indicated by a pattern of similar counties with poor predicted values that err in the same direction.

Fifth, the county- and state-level models make use of predictor variables from administrative records, but the quality of these variables is unclear and so is of concern for this approach. The Census Bureau has performed careful analysis of the definitional, coverage, and nonresponse issues raised through the use of the administrative records (see Coder et al., 1996). More work could be done, however, to further understand the uniformity of the relationships to poverty across counties and the likelihood of misresponse. In this regard, we note that in 10 percent of counties, the number of reported child exemptions on income tax returns exceeded that of the estimated total population under age 21. Although this anomaly is not entirely unexpected because of differences between reporting on income tax returns and reporting in the census, which is the basis of population estimates, it needs further study.[1]

The panel expects that after all the above analyses have been carried out, it will not only be clear that model-based county estimates are preferable to the decennial census estimates for current allocations, but it will also be relatively clear which particular model—the Census Bureau's current model or one of the alternatives suggested in this report—is preferable for providing updated estimates of school-age children in poverty. We say relatively clear because there is

[1]For example, addresses on tax returns are not always the county of residence as defined for the census (e.g., the address may be that of the tax preparer); tax filers may report exemptions for children who do not reside with them; and some child exemptions are for children aged 21 or older.

at present only a single opportunity for the use of comparison values. Therefore, there will always be some degree of uncertainty as to which particular set of model-based estimates is preferable.

OTHER RESEARCH

The panel encourages the Census Bureau to proceed along lines of innovative research that it has already begun. The rush to produce estimates for use in fiscal 1997 and 1998 allocation forced the Census Bureau to put off some interesting work related to a time-series approach that uses correlations for 5 years of CPS data to produce state estimates. It is possible that this approach could be used to produce county estimates as well. These kinds of time-series models have only recently been proposed in the literature and have not been sufficiently tested. The Census Bureau is commended for its initial research into this area and is encouraged to continue, although such work will likely represent a long-term effort and not a contribution to the immediate problem of evaluating and improving the county estimates in the short term.

The Census Bureau has also begun work on a measurement error model to provide more understanding of CPS-census differences. This is an important effort. In addition to assisting in the understanding of the comparison of the 1989 model estimates with the 1990 decennial census estimates, such work could also be useful when developing estimates for 1999 by providing a way to produce poverty estimates from the decennial census that are consistent with the way poverty is measured by the CPS.

Another area for investigation concerns timeliness of the estimates. Although the Census Bureau's model-based estimates of poor school-age children are more up to date than the 1990 census estimates, the estimates for 1993 lag the allocation year (1997) by several years. It would be useful to consider possible methods to make model-based estimates yet more timely. Such methods could range from basing the county estimates on 1 year of March CPS data, instead of 3 years (which could reduce the lag by a year at the cost of increased variability in the estimates), to research into the dynamics of poverty across time and geographic areas that would make it possible to project the model-based estimates several years forward. Such research is likely to be difficult to carry out, but it and other methods to improve timeliness should be identified along with their possible advantages and disadvantages as a first step in developing a research plan in this area.

Still another area for investigation includes evaluation and refinement of the methods used to develop postcensal population estimates for children aged 5-17, which for 1994 represented relatively crude adjustments of estimates from the 1990 census. Such work should include research on methods to estimate the errors in the population estimates. There should also be related work to evaluate and improve the estimates of total and poor school-age children in Puerto Rico

with regard to their quality and comparability with the estimates for U.S. counties.

A major thrust of the next phase of the Census Bureau's work will be to determine whether updated estimates of poor school-age children at the level of school districts can be produced that are of adequate quality and preferable to decennial census estimates. Certainly, given the small CPS sample sizes for most school districts, the paucity of good administrative record data, the frequent revisions of school-district boundaries, the dynamics of school district-level poverty, and other factors, this work will be more difficult than county-level estimation, and evaluation will be critically important. The panel looks forward to working with the Census Bureau on this task as part of its continuing study.

CONCLUSION

The Census Bureau has made an excellent start on a difficult problem. The updated estimates it has produced of the number of school-age children in poverty for counties represent substantial progress in meeting an important need. It is the panel's strong belief that with the evaluations and other analyses we have outlined, a model-based approach will produce updated poverty estimates that are superior to those from the decennial census. What the panel has outlined represents a considerable amount of work, but such work is necessary to properly support the use of model-based estimates for the important purpose of allocating funds under Title I.

We stress that it is critical that users be provided with information that will enable them to fully understand the properties of model-based estimates for use for fund allocations and other purposes. Therefore, when official estimates of the number of school-age children in poverty are released by the Census Bureau, they should be accompanied by formal documentation that details the methods used, the analyses conducted, the estimates of reliability, and the detailed results of evaluations, in order to inform users of the basis of the estimates and to give the research community opportunities for further analyses.

Appendices

APPENDIX
A

The Title I Allocation Process

Title I allocations are based primarily on Census Bureau estimates of the number of children aged 5-17 in families with income below the poverty level. The formula also counts three other categories of children aged 5-17 (about 5 percent of all formula-eligible children): children in foster homes, in families above the poverty level that receive Aid to Families with Dependent Children (AFDC), and in local institutions for neglected and delinquent children. The Census Bureau estimates are combined with factors defined in the Title I statute to determine the allocations to each state and school district. Title I funds are provided to school districts through a two-stage process: the Department of Education allocates funds to the county level, and the states then suballocate the funds to school districts within each county.

The statute contains four formulas for allocating Title I funds—basic grants, concentration grants, targeted grants, and the Education Finance Incentive Program—but Congress has to date appropriated funds only for the basic and concentration formulas. Basic grants have existed since the program's inception in 1965; concentration grants were added in 1978 to provide additional funds to school districts with high concentrations of school-age children in poverty (Moskowitz et al., 1993). The fiscal 1997 appropriation was $6.195 billion for basic grants (86 percent of the total) and $999 million for concentration grants.

The basic grant formula allocates funds to all counties based on each county's number of formula-eligible children and the state's average per-pupil expenditures, a factor intended to compensate for state differences in the cost of education. County allocations are calculated in an iterative process. The number of formula children in each county is multiplied by 40 percent of the state's per-pupil expenditure (which is limited to between 80 and 120 percent of the national

average); the resulting allocations are then proportionally reduced so that the total matches the total appropriation for basic grants. Allocations are then adjusted to meet hold-harmless and state minimum grant provisions (see below). Although there are no minimum eligibility criteria for counties, school districts must have at least 10 formula-eligible children, *and* the number of eligible children must exceed 2 percent of the district's population aged 5-17, in order to receive a basic grant.

The concentration grant formula allocates funds only to counties and school districts with high numbers *or* percentages of poor school-age children. To be eligible to receive a concentration grant, a county or district must have at least 6,500 formula-eligible children or more than 15 percent of the children in the county or district must be formula eligible. For concentration grants, allocations for eligible counties are calculated based on numbers of formula-eligible children, the state's per-pupil expenditure, and a state minimum provision.

The Title I formula includes a hold-harmless provision to cushion the impact of decreases in allocations. Historically, there has been no hold-harmless provision for concentration grants, but in fiscal 1996, a hold-harmless provision applied to both formulas at a rate of 100 percent: that is, a county could not receive less funds than it had received in fiscal 1995. Beginning in fiscal 1997, the hold-harmless provision applies to basic grants at variable rates, with a higher rate for higher poverty counties and school districts: counties and districts with 30 percent or more poor school-age children are guaranteed at least 95 percent of the prior year's grant; the guarantee is 90 percent for counties with 15-30 percent poor school-age children and 85 percent for counties with fewer than 15 percent poor school-age children. For fiscal 1997 and subsequent years, there is no hold-harmless provision for concentration grants.

A state minimum grant provision applies to each of the two formulas as well. For basic grants, the state minimum grant is equal to the lesser of (1) 0.25 percent of total funds available for Title I basic grants and (2) the average of 0.25 percent of total funds and 150 percent of the national average per-pupil grant payment multiplied by the number of poor children in the state. There is some added complexity for the state minimum for concentration grants (Moskowitz et al., 1993).

APPENDIX B
Comparison of Census and CPS Estimates of Poverty

There are many similarities and some differences between the CPS and the census systems for measuring poverty. They both base poverty estimates on the official definition of poverty for the nation, which compares the income of a family (or an unrelated individual) to a given poverty threshold, but they differ in the amount of data they collect, how they collect and process the data, and the frequency of data collection.

1990 CENSUS

A census of the U.S. population is conducted once every 10 years, and 1990 is the most recent one. In the 1990 census, income data—the basis for measuring poverty—were collected from a sample of 15 million households: a sample of about 1 in 6 households spread systematically across the country, except that very small counties and places (with estimated 1988 populations under 2,500) were sampled at a 1-in-2 rate, and very populous census tracts (or equivalent areas) were sampled at a 1-in-8 rate.

Data are collected in the census mainly by self-enumeration, whereby respondents fill out questionnaires received in the mail. Enumerators follow up those households that fail to return a questionnaire and collect the information through direct interviews. In 1990 approximately 74 percent of U.S. households returned their questionnaires with some or all of the requested information (Edmonston and Schultze, 1995:189). Data from the balance of the population were obtained by personal interviews. The follow-up enumerators are usually inexperienced temporary workers who are given very limited training.

The income data in the 1990 census are based on eight questions on various

components of income. (The census form also included a total income question, which was intended to permit respondents to enter a single amount if they could not provide amounts by source.) Nonresponse rates are higher for income than for most other items in the census. When household income information is missing, the Census Bureau uses statistical techniques to impute it on the basis of nearby households with similar characteristics. For the 1990 census, on average, 19 percent of aggregate household income was imputed (Edmonston and Schultze, 1995:387).

All censuses are subject to undercount—that is, failure to count everyone. There are no direct estimates of the undercount for poor children. For 1990, the net undercount was estimated at 1.8 percent for the total population, but there were substantial differences among population groups. For example, the net undercount was estimated at 5.7 percent for blacks and 1.3 percent for nonblacks. The net undercount also varied significantly by age: black girls and boys aged 5-9 were missed at a rate of 7.5 percent and 7.7 percent, respectively. Almost two-thirds of the estimated omitted population consisted of two age groups: children under age 10 and men aged 25-39 (Robinson et al., 1993:13). The undercount was also higher in large cities than in other areas, and it was disproportionately concentrated in the inner areas of those cities. These are also the areas where poverty is high. Thus, it seems likely that the undercount for poor children aged 5-17 is larger than the undercount of all children aged 5-17.

CURRENT POPULATION SURVEY

The CPS is a monthly labor force participation survey. For the period from 1990 to 1994, about 60,000 housing units were eligible for interview every month, and about 57,400 of them were found to be occupied by households eligible for interview.[1] Of these 57,400 households, an interview was not obtained for various reasons for about 2,600 households—a noninterview rate of 4.5 percent.

Part of the CPS sample is changed each month: in the rotation plan, three-fourths of the sample is common from one month to the next, and one-half is common for the same month a year earlier. Each March, supplementary questions are asked about money income received the previous year. To obtain more reliable income data for the Hispanic-origin population, all November CPS households with one or more Hispanic persons are reinterviewed in March if they still include a Hispanic person. This procedure adds about 2,500 Hispanic households to the sample in March.

The CPS sample design, which is a multistage probability sample design, is revised about once every 10 years on the basis of the results of the latest census.

[1]Starting in 1996, about 50,000 households nationwide (a sample of about 1 in 2,000 households) were eligible for interview every month—a reduction of about 17 percent from the early 1990s.

From 1986 to 1994, the CPS sample design included 729 sample areas consisting of about 1,300 counties. These areas were chosen on the basis of 1980 census data to represent all 3,141 counties (in 1990) and independent cities in the 50 states and the District of Columbia. A design based on the 1990 census was phased in between April 1994 and July 1995: it included 792 sample areas consisting of about 1,300 counties, chosen to represent all 3,143 counties (in 1994) and independent cities in the 50 states and the District of Columbia.

In general, larger states have larger CPS sample sizes. The largest states, however, have CPS sample sizes that are smaller than their proportionate share of the U.S. population, and the smallest states have proportionately larger sample sizes. For example, California, with 12.2 percent of the U.S. population, has 9.9 percent of the CPS sample; Wyoming, with 0.18 percent of the U.S. population, has 1.3 percent of the CPS sample. This sample design means that estimates of poverty rates in large states are generally more precise than those in smaller states. The largest states, however, have larger relative errors due to sampling variability than would be expected if the CPS sample were allocated to the states in proportion to their population; the reverse holds true for smaller states.

The sample is designed to meet specific reliability criteria for the nation, each of the 50 states and the District of Columbia, and the substate areas of New York City and the Los Angeles–Long Beach metropolitan area. *The CPS is not designed for direct county estimates*. More than one-half of U.S. counties do not have sample households in the survey.

The CPS is carried out by permanent, experienced, and well-trained interviewers, initially by personal direct interviews, with subsequent interviews by telephone. For the March Income Supplement, the CPS asks household respondents about their money income received during the previous year, using a detailed set of questions for identifying about 28 different sources. About 20 percent of aggregate household income is imputed (about the same percentage as in the census)—that is, the data are missing and therefore constructed from information from similar households (Citro and Kalton, 1993:Table 3-6).

Like other household surveys, the CPS exhibits population undercoverage at higher rates than the census itself. The coverage ratios for the CPS show the magnitude of the population undercoverage relative to the census. Coverage ratios are defined as the estimated survey population before ratio adjustment to census-based population controls divided by the census-based population controls. (Beginning with the March 1994 CPS, the population controls reflect an adjustment for the undercount in the census itself.) For March 1994, the ratio of the CPS estimated population to the population control total (all ages) was 92 percent; for the age group 0-14 and the age group 15-19 years, the ratios were 94 percent and 88 percent, respectively (Bureau of the Census, 1996:Table D-2).

CPS undercoverage is corrected by ratio adjustments to the survey weights that bring the CPS estimates of population in line with updated national population controls by age, race, sex, and Hispanic origin. However, the ratio adjust-

ments do not correct for other characteristics on which the undercovered population might be expected to differ from the covered population. For example, the ratio adjustments reweight equally the sample households within an age-race-sex-Hispanic origin category, when research suggests that it is likely that lower income households within a category are more poorly covered than higher income households.

DIFFERENCES BETWEEN CENSUS AND CPS DATA

In comparing the census and the CPS as data sources for estimates of income and poverty, one difference is the definition of the universe for which the numbers in poverty are estimated; see Table B-1. Residence rules differ for the two data sources: in the census, students attending colleges away from their parental homes are counted at their college location; in the CPS, they are usually counted at their parental home. Also, the census excludes all unrelated individuals under age 15 in households from the poverty universe; the CPS excludes only those unrelated individuals under age 15 in households who are not part of an unrelated subfamily. (Unrelated subfamilies are made up of people who are related to each other but not to the householder, such as the family of a resident employee.) For

TABLE B-1 Poverty Universes for the 1990 Census and the March 1990 CPS

Component	1990 Census	March 1990 CPS
Total resident U.S. population	248,709,873	250,180,762[a]
Population not covered in CPS	–3,998,221	–3,989,762
Institutionalized	*–3,334,018*	— [b]
Armed Forces in barracks	*–589,700*	— [b]
Unrelated individuals under 15 in group quarters	*–74,503*	— [b]
Unrelated individuals under 15 in households	–780,235[c]	–199,000
College dormitory residents	–1,953,558	0[d]
Poverty universe	241,977,859	245,992,000

[a]Reflects 1980 census-based population estimates, which estimated a higher resident population in 1990 than the 1990 census.

[b]Intercensal estimates used to derive CPS population controls for survey weighting are not available by component. The sum of these three components was 3,989,762.

[c]The large difference between census and CPS estimates reflects the census practice of excluding children under age 15 in unrelated subfamilies from the poverty universe.

[d]The CPS includes household members away at school who are living in dormitories as family members and includes them in the poverty universe.

SOURCE: Data from Housing and Household Economic Statistics Division, Bureau of the Census.

estimates of related children in poverty (who are children in families but not children in unrelated subfamilies), the CPS and census poverty universes differ less than the CPS and census poverty universes for the total population. Essentially, the only difference is the treatment of college students in dormitories.

Overall imputation rates for income nonresponse are about the same in the census and the CPS, and the amount of income imputed is similar. Table B-2 presents aggregate income estimates for different income components from the two sources. The more detailed set of income questions in the CPS—28 compared with 8 in the census—and the direct interviewing methodology in the CPS would be expected to provide more comprehensive and accurate income data. It is believed that overreporting of some components of income, such as wages and salaries, occurs in the self-reported census data. (Another reason for the overestimate of wages and salaries in the census, compared with the independent benchmark, may be the editing procedures that were applied to responses to the total income question.)

The net effect of the differences between the CPS and the census in data collection, processing, and other aspects of the two systems is that there are dif-

TABLE B-2 Household Income by Type, 1989: 1990 Census and March 1990 CPS

	Aggregate Income (in $ billions)			Percentage of Benchmark	
Source of Income	1990 Census	March 1990 CPS	Independent Benchmark	1990 Census	March 1990 CPS
Total money income	3,537.4	3,460.4	—	—	—
Amounts for which benchmarks can be computed	3,499.2	3,393.9	3,819.7	91.6	88.7
Wages and salaries	2,652.7	2,545.9	2,625.2	101.0	96.8
Nonfarm self-employment	218.6	207.1	290.0	75.4	71.4
Farm self-employment	20.3	18.6	49.9	40.7	37.2
Interest, dividends, and rent	258.8	247.7	471.5	54.9	52.5
Social Security and Railroad Retirement	188.2	201.4	207.9	90.5	96.9
Public assistance	28.3	25.9	32.8	86.3	79.0
Retirement, disability, and survivor income	132.3	147.3	142.4	92.9	103.4
Other income sources	38.2	66.5	—	—	—

NOTE: The independent benchmarks shown here for 1989 were extrapolated from 1990 independent estimates. For a detailed discussion of development of independent benchmarks, see Bureau of the Census (1993b:C1-C3).

ferences between them in estimates of income and poverty. Aggregate estimates of total income and income by type when compared with independent benchmarks differ between the CPS and census. For example, total income from the 1990 census (for income types for which independent estimates can be constructed) is 91.6 percent of the benchmark; the corresponding figure from the March 1990 CPS is 88.7 percent. For income from such sources as Social Security and Railroad Retirement and retirement, disability, and survivor income, the CPS is closer to the benchmark than the census.

Estimates of median household income in 1989 by state differ between the March 1990 CPS and the 1990 census by amounts that are statistically significant for 18 states; see Table B-3. For 15 of the 18 states, the census estimates are higher than the CPS estimates; for some states, the differences are as much as 10-15 percent. For three states, the census estimates of median household income are significantly lower than the CPS estimates, in the range of 6-8 percent. Estimates of poverty rates by state also differ between the CPS and the census: statistically significant differences are observed for seven states; for six of them, the census poverty rates are significantly higher; see Table B-4.

Estimates of poverty rates for related children aged 5-17 differ between the CPS and the census. The CPS estimate that 18 percent of all related children aged 5-17 were poor in 1989 (based on a 3-year average of data from the March 1989, 1990, and 1991 CPS) exceeds the census estimate by 1 percentage point (5.9% of the census estimate), a difference that is statistically significant at the 10 percent significance level; see Table B-5. (The CPS estimate of the number of poor school-age children also exceeds the census estimate by a statistically significant amount; see Table 2-4.)

A question is whether CPS and census estimates of poverty rates and numbers of poor related children aged 5-17 differ in terms of geographic distribution, perhaps because of differences in completeness of income reporting that reflect the income mix in different areas or other reasons. In preliminary research conducted by the panel, no statistically significant differences were found in the ratios of CPS to census estimates of the number of poor school-age children among various geographic groupings of counties and states. For example, when counties were grouped by size, there were no significant differences among the groups in their ratios of CPS to census estimates of the number of poor school-age children.

In summary, there are many factors reflecting differences between the census and CPS concepts and procedures that may account for variations in their estimates of poverty levels, rates, and distributions. It is important to keep these factors in mind, particularly when attempting to measure changes in poverty at specific levels of geography since 1990.

TABLE B-3 Median Household Income in 1989 by State: 1990 Census and March 1990 CPS

State and National	Income (in $) 1990 Census Median	1990 Census Standard Error	March 1990 CPS Median	March 1990 CPS Standard Error	Percentage Difference: Census – CPS as Percent of CPS
U.S. Total	30,056	7	28,906	159	4.0*
Alabama	23,597	58	21,284	1,070	10.9*
Alaska	41,408	165	36,006	1,378	15.0*
Arizona	27,540	72	28,552	1,210	–3.5
Arkansas	21,147	50	21,433	915	–1.3
California	35,798	27	33,009	618	8.4*
Colorado	30,140	55	26,806	1,398	12.4*
Connecticut	41,721	77	42,321	1,592	–1.4
Delaware	34,875	151	32,068	1,133	8.8*
D.C.	30,727	156	26,752	1,015	14.9*
Florida	27,483	35	26,085	475	5.4*
Georgia	29,021	53	27,542	1,021	5.4
Hawaii	38,829	173	35,035	1,328	10.8*
Idaho	25,257	79	24,654	953	2.4
Illinois	32,252	32	31,300	623	3.0
Indiana	28,797	50	25,898	1,022	11.2*
Iowa	26,229	46	26,265	792	–0.1
Kansas	27,291	56	26,862	908	1.6
Kentucky	22,534	55	23,283	1,206	–3.2
Louisiana	21,949	49	22,861	1,857	–4.0
Maine	27,854	82	28,221	1,389	–1.3
Maryland	39,386	73	36,016	1,187	9.4*
Massachusetts	36,952	58	36,086	704	2.4
Michigan	31,020	32	30,775	790	0.8
Minnesota	30,909	37	30,185	1,278	2.4
Mississippi	20,136	59	19,917	947	1.1
Missouri	26,362	42	26,497	746	–0.5
Montana	22,988	98	23,692	1,311	–3.0
Nebraska	26,016	60	26,319	1,521	–1.2
Nevada	31,011	92	29,340	8,455	5.7*
New Hampshire	36,329	97	37,532	1,371	–3.2
New Jersey	40,927	51	39,120	948	4.6*
New Mexico	24,087	89	22,602	1,028	6.6
New York	32,965	41	31,496	453	4.7*
North Carolina	26,647	38	26,406	517	0.9
North Dakota	23,213	94	25,229	903	–8.0*
Ohio	28,706	37	29,021	655	–1.1
Oklahoma	23,577	57	23,667	1,236	–0.4

continued on next page

TABLE B-3 Continued

State and National	Income (in $) 1990 Census Median	1990 Census Standard Error	March 1990 CPS Median	March 1990 CPS Standard Error	Percentage Difference: Census – CPS as Percent of CPS
Oregon	27,250	55	28,529	1,435	–4.5
Pennsylvania	29,069	32	28,690	684	1.3
Rhode Island	32,181	120	30,124	1,354	6.8
South Carolina	26,256	56	23,798	1,059	10.3*
South Dakota	22,503	89	24,108	999	–6.7*
Tennessee	24,807	50	22,611	1,305	9.7
Texas	27,016	27	25,886	559	4.4*
Utah	29,470	86	30,717	1,014	–4.1
Vermont	29,792	110	31,295	1,136	–4.8
Virginia	33,328	60	34,118	1,205	–2.3
Washington	31,183	44	31,961	1,472	–2.4
West Virginia	20,795	61	21,677	843	–4.1
Wisconsin	29,422	53	29,123	1,240	1.0
Wyoming	27,096	133	29,521	1,289	–8.2*

*Statistically significant difference from 0 at the 10 percent significance level.

SOURCE: Data from Bureau of the Census.

TABLE B-4 Poverty Rates for All Persons by State, 1989: 1990 Census and March 1990 CPS

State and National	1990 Census Population (000s)	1990 Census Total Poverty Rate	1990 Census Standard Error[a]	March 1990 CPS Population (000s)	March 1990 CPS Total Poverty Rate	March 1990 CPS Standard Error	Difference Between Rates
U.S. Total	241,978	13.12	0.012	245,992	12.82	0.2	0.30
Alabama	3,946	18.34	0.096	4,074	18.90	2.0	–0.56
Alaska	532	9.00	0.140	488	10.45	1.5	–1.45
Arizona	3,584	15.74	0.086	3,556	14.12	1.9	1.63
Arkansas	2,292	19.07	0.116	2,419	18.27	2.0	0.80
California	29,003	12.51	0.027	29,346	12.85	0.7	–0.35
Colorado	3,213	11.68	0.080	3,258	12.06	1.8	–0.38
Connecticut	3,188	6.82	0.057	3,136	2.87	1.0	3.95*
Delaware	645	6.71	0.086	676	10.06	1.7	–1.35
D.C.	571	16.87	0.233	569	17.93	2.4	–1.06
Florida	12,641	12.69	0.042	12,762	12.46	0.9	0.23
Georgia	6,300	14.65	0.069	6,197	14.99	1.9	–0.34
Hawaii	1,071	8.25	0.119	1,036	11.23	1.8	–2.98*
Idaho	966	13.25	0.153	1,014	12.43	1.6	0.82
Illinois	11,144	11.91	0.033	11,559	12.72	0.9	–0.81
Indiana	5,372	10.68	0.051	5,453	13.74	1.9	–3.06
Iowa	2,677	11.48	0.074	2,835	10.26	1.5	1.22
Kansas	2,392	11.48	0.074	2,835	10.26	1.5	1.22

continued on next page

TABLE B-4 Continued

State and National	1990 Census Population (000s)	1990 Census Total Poverty Rate	1990 Census Standard Error[a]	March 1990 CPS Population (000s)	March 1990 CPS Total Poverty Rate	March 1990 CPS Standard Error	Difference Between Rates
Kentucky	3,582	19.03	0.093	3,576	16.11	2.0	2.92
Louisiana	4,101	23.58	0.108	4,080	23.28	2.3	0.29
Maine	1,190	10.80	0.108	1,233	10.38	1.6	0.72
Maryland	4,661	8.27	0.054	4,567	9.00	1.6	–0.73
Massachusetts	5,812	8.93	0.050	5,831	8.80	0.8	0.14
Michigan	9,077	13.12	0.040	9,297	13.21	0.9	–0.09
Minnesota	4,259	10.22	0.053	4,268	11.20	1.7	–0.98
Mississippi	2,503	25.21	0.147	2,574	21.99	2.0	3.22
Missouri	4,971	13.34	0.055	5,193	12.59	1.8	0.75
Montana	777	16.07	0.186	816	15.56	1.8	0.51
Nebraska	1,531	11.14	0.097	1,602	12.73	1.6	–1.59
Nevada	1,178	10.15	0.118	1,127	10.74	1.7	–0.58
New Hampshire	1,076	6.42	0.095	1,100	7.64	1.6	–1.21
New Jersey	7,563	7.58	0.030	7,623	8.16	0.8	–0.58
New Mexico	1,584	20.61	0.141	1,519	19.55	2.0	1.05
New York	17,482	13.03	0.040	17,938	12.57	0.7	0.46
North Carolina	6,397	12.97	0.059	6,301	12.22	0.9	0.75
North Dakota	614	14.38	0.210	642	12.31	1.6	2.07
Ohio	10,574	12.54	0.036	10,754	10.63	0.8	1.91*
Oklahoma	3,052	16.71	0.091	3,126	14.72	1.8	1.99
Oregon	2,776	12.42	0.080	2,915	11.22	1.8	1.21
Pennsylvania	11,536	11.13	0.033	12,141	10.39	0.8	0.84

Rhode Island	964	9.61	0.121	963	6.65	1.5	2.96*
South Carolina	3,368	15.37	0.101	3,446	16.98	1.8	–1.60
South Dakota	670	15.86	0.160	696	13.22	1.6	2.64
Tennessee	4,744	15.70	0.045	4,833	16.44	1.9	–2.73
Texas	16,580	18.10	0.044	16,886	17.06	1.0	1.04
Utah	1,694	11.36	0.104	1,683	8.20	1.4	3.16*
Vermont	541	9.86	0.145	556	7.91	1.6	1.94
Virginia	5,969	10.25	0.055	6,159	10.89	1.4	–0.65
Washington	4,741	10.92	0.058	4,729	9.62	1.6	1.30
West Virginia	1,755	19.66	0.134	1,799	15.73	1.9	3.93*
Wisconsin	4,754	10.70	0.054	4,694	8.37	1.4	2.32*
Wyoming	442	11.86	0.196	462	10.82	1.9	1.04

*Statistically significant difference from 0 at the 10 percent significance level.

[a]Standard errors of estimates from the 1990 census are calculated using 1980 census design factors.

SOURCE: Data from Bureau of the Census.

TABLE B-5 Poverty Rates for Related Children Aged 5-17 in 1989, by Selected Categories of Counties: 1990 Census and March CPS

County Category	Percent Poor Related Children Aged 5-17		Difference Between Rates
	1990 Census	March CPS[a]	
U.S. Total	17.0	18.0	1.0*
Metropolitan			
Central	16.4	17.9	1.5*
Other	11.4	12.5	1.1
Nonmetropolitan	20.4	19.9	–0.5
Region[b]			
Northeast	14.3	15.5	1.2*
Northcentral	14.9	15.8	0.9*
South	20.5	21.3	0.8
West	16.2	17.3	1.1*
Population Size			
Under 2,500	22.9	22.1	–0.8
2,500-4,999	22.2	14.6	–7.6
5,000-9,999	23.1	24.7	1.6
10,000-49,999	20.6	20.9	0.3
50,000-99,999	16.6	15.7	–0.9
100,000-499,999	14.7	15.7	1.0*
500,000-999,999	14.6	15.6	1.0
1,000,000 and over	19.1	21.5	2.4*

*Statistically significant difference from 0 at the 10 percent significance level.

[a]The CPS estimates are 3-year centered averages of data from the 1989, 1990, and 1991 March CPS (reported income in 1988, 1989, and 1990, with population controls derived from the 1980 census).

[b]The Census Bureau's regions are defined in Note c of Table 2-3.

SOURCE: Data from Bureau of the Census.

APPENDIX C
Census Bureau's Methodology for Model-Based Estimates

The Census Bureau's estimation methodology for producing county estimates of the number and percentage of related children aged 5-17 in poverty (poor school-age children) can be separated into four distinct steps: (1) the production of county estimates of the number of poor school-age children; (2) the production of state estimates of the number of poor school-age children; (3) the modification of the county estimates so that they add to the state estimates; and (4) the use of the estimated number of related children aged 5-17 as a denominator to produce estimates of the percentage of those children in poverty. Steps 1, 2, and 3 are described below; Appendix D describes the development of the denominators used in step 4.

Two time periods must be differentiated in this discussion. The March 1994 CPS supports model-based estimates of the numbers of school-age children who lived in each county in 1994 and were in poverty in 1993 (the reference year for the income questions). Estimates that refer to the March 1994 CPS (or 1994 and surrounding years) are therefore referred to as 1993 estimates, although strictly speaking they involve information for both 1993 and 1994. Similarly, the 1990 decennial census produced estimates of school-age children who lived in each county in 1990 and were in poverty in 1989. The March 1990 CPS (or 1990 and surrounding years) supports estimates that are for the same income reference year (1989) as the census; we refer to these estimates as the 1989 estimates. The 1993 estimates are the current objective of the small-area estimation program; the 1989 estimates are important for evaluation purposes because they can be compared to the census. This appendix considers both the 1993 and 1989 estimates.

COUNTY-LEVEL ESTIMATION[1]

The county-level model uses regression to produce the estimates, with (3-year average) CPS measures as the dependent variable and administrative data and population estimates for the independent variables. In this model:

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + u_i + e_i \ ,$$

where:

y_i = log(3-year weighted average of the number of poor school-age children in county i),[2]

x_{1i} = log(number of child exemptions [assumed to be under age 21] reported by families in poverty on tax returns in county i),

x_{2i} = log(number of people receiving food stamps in county i),

x_{3i} = log(estimated noninstitutionalized population under age 21 in county i),[3]

x_{4i} = log(number of child exemptions on tax returns in county i),

x_{5i} = log(number of poor school-age children in county i in the previous census),

u_i = model error for county i, and

e_i = sampling error for county i.

Variables are transformed using logarithms for two reasons. First, it is more plausible that the model is homoscedastic on the log scale (corresponding to a constant coefficient of variation or equal model variances of share in poverty) than on the original scale (equal model variances of number in poverty) over the extremely wide range of county sizes. Second, the transformed variables have a much more symmetric distribution, and the scatterplots of various covariates with the dependent variable are more linear.

Only CPS sample counties that have some poor school-age children in at least one of the 3 years contributing to the 3-year average are used in the regres-

---

[1]The following section draws heavily from the Census Bureau's documentation (Coder et al., 1996).

[2]The estimated number of poor school-age children is the product of the weighted 3-year average CPS county poverty rate for related children aged 5-17 and the weighted 3-year average CPS county number of related children aged 5-17. The weights for this average are the fractions of the 3-year total of CPS interviewed housing units containing children aged 5-17 in each year. For estimates from a given year, stratum-level weights ordinarily used have been removed. These stratum-level weights result from an over- or undersampling of counties to account for certain demographic or other characteristics. As a result, for this analysis, counties receive a weight depending directly on their population size and not on other characteristics.

[3]For the 1989 model, estimates of this variable are from the 1990 census; for the 1993 model, estimates are from the Census Bureau's population estimates program.

sion. For the 1989 model, 1,028 of 3,141 counties were included in the regression; for the 1993 model, 1,184 of 3,143 counties were included (see Coder et al., 1996:Table 3).

As represented above, the variability of y_i , after the effects of the predictor variables are accounted for, is due to model error and sampling error. Since the sum of these vary substantially among counties, resulting in heterogeneous variances, a weighted least-squares regression is used. The weights are developed as follows. A mean square error is computed from the unweighted regression of log(1990 census estimates of the number of poor school-age children in 1989), using the covariates appropriate for an estimate of the dependent variable for 1989 (e.g., x_{5i} would pertain to the 1980 census) and including only counties that have sample households with poor school-age children in the March 1993, 1994, or 1995 CPS. The mean square error or variance of total error for this regression is the sum of sampling variance and variance due to model error, that is, var(u_i) + var(e_i). The variance due to model error for this regression can be estimated by subtracting the contribution to mean square error due to the (estimated) sampling variances of log(census poverty estimates) that are derived from published generalized variance function estimates for each county. Since the census sampling variances are relatively small, variance due to model error is about 88 percent of the mean square error in the census regression model. This estimated model variance is assumed to closely approximate the model variance for an (unweighted) regression with the dependent variable of log(3-year average CPS estimates of the number of poor school-age children). Therefore, when this estimate of model error is subtracted from the mean square error for the CPS regression, the remainder is an estimate of the total county-level CPS sampling variability.

The individual county-level sampling variance (for the log dependent variable) is then estimated by assuming that it is inversely proportional to sample size. To obtain the individual county-level contribution from model error, the model error is assumed to be homogeneous (i.e., the variance of the model error is assumed to be equal for each county). The mean square error for county *i* is then the sum of the variance due to model error and the estimated sampling variance (which depends on the county sample size). Most of the CPS mean square error (about 90 percent) is derived from sampling variance. The reciprocals of the mean square errors are then used as weights to recompute the regression using weighted least squares, which provides new weights since the mean square error has changed. Only one iteration is performed.

The weights for the 1989 and 1993 CPS regressions differ because of their different data sets and because each year's model uses the counties in the CPS sample for that year. Together, these differences cause the estimated sampling variances to differ. However, the procedure used to develop the weights for the 1989 and 1993 CPS regressions assumes that the CPS regressions have the same model error as the 1989 census regression. Implicit in this assumption are the

assumptions that the CPS and census regression models are very similar and that the time from the last census (the 1980 census for the 1989 model and the 1990 census for the 1993 model) is not an important source of differences in mean square error for these models. These assumptions have not been fully validated.

For the counties that do not appear in the 3-year CPS sample, estimates of log(number of poor school-age children) are calculated by substituting the covariates for that county in the estimated regression model and computing the model prediction. For the 1,028 (1989 model) or 1,184 (1993 model) counties for which direct CPS estimates are available, the direct 3-year average CPS estimates and the model predictions are combined, using a weighted average (referred to as empirical Bayes or shrinkage estimation) in which the weight for the model prediction is the ratio of the estimated sampling variance to the sum of the estimated sampling variance and the model error variance for that county. It is important to note that for almost all counties, the great majority of the weight is given to the model predictions; for only 13 counties is the weight for the model prediction less than 0.5.

The numbers of poor related children aged 5-17 in each county estimated from the county-level model are then controlled to the state poverty estimates.

STATE-LEVEL ESTIMATION[4]

For most states, direct estimates of the number of poor school-age children from the March CPS are insufficiently reliable to be used alone. A model-based approach that borrows strength from administrative records (IRS tax files, food stamp files, etc.), the decennial census, and other states is therefore used. The methodology for development of the 1989 state estimates is described below; similar methods were used for 1993 estimates, following the specifications that were found to work well for 1989.

The regression model for producing state estimates of the proportion of school-age children in poverty has the following form (for details, see Fay, 1996):

$$y_{it} = (\Sigma \beta_{tj} x_{itj} + z_{it}) + e_{it} ,$$

where:

i = the state of interest,
t = the year of estimation,
j = the covariate,

[4]This section draws heavily from the Census Bureau's documentation (Fay, 1996).

y_{it} = direct estimate of the percentage of poor school-age children from the CPS in year t,[5]

z_{it} = a random effect that represents differences between the model-based estimates and the direct estimates from the CPS, and

e_{it} = sampling variance for the dependent variable for state i in year t.

The regression coefficients, β_{tj}, have the subscript t to indicate that they are reestimated for each year, and $\Sigma\beta_{tj}x_{itj}$ represents the portion of poverty that is linearly related to the covariates described below. The z_{it} are assumed to be independent and identically distributed in any given year. The e_{it} are normal disturbances resulting from sampling variance. The quantity $\Sigma\beta_{tj}x_{itj} + z_{it}$ represents the true poverty count for state i, which is the goal of the estimation procedure.

The Census Bureau first performed a cross-sectional (linear) regression of the 1980 census estimates of poverty rates (1979 income) for school-age children on a variety of covariates. A cross-sectional regression was also fit for the 1990 census estimates of poverty rates (1989 income) for school-age children. (These regressions used ordinary least-squares estimation.) The residuals for the 1980 and 1990 census models were observed to be correlated (Fay, 1996), indicating that counties that had more poverty than predicted by the cross-sectional model for 1979 also tended to have more poverty than predicted by the cross-sectional model for 1989. This fact can be used to improve the 1989 predictions.

Next, a regression model was built for the CPS estimates of school-age children's poverty rates in 1989. The covariates that were predictive in the regression models with census estimates of school-age children's poverty rates as the dependent variable were selected for inclusion in this model, along with the residuals from the regression of the 1980 census estimates of children's poverty rate on the same covariates. These covariates were (1) the percentage of child exemptions reported by families in poverty on tax returns, (2) the percentage of the noninstitutionalized population under age 65 that do not file income tax returns, (3) the percentage of the population that receives food stamps, and (4) the residuals from the regression fit on 1980 census poverty rates (discussed above). The CPS model of the poverty rates for school-age children was not used to select covariates because of the large sampling variability in the dependent variable.

During the exploratory phase of model development, various transformations of both the dependent and independent variables were examined. The untransformed versions seemed to fit best, justifying the use of a model that is linear in percentages.

[5]The percentage is calculated through the following ratio: the numerator is the number of poor related children aged 5-17 from the CPS, and the denominator is the estimated total population of noninstitutionalized children aged 5-17 (whether related or not) from the CPS.

In the basic regression model, the β_{tj} were estimated by weighted least squares, the weights being the inverse of the sum of the estimated sampling variance and the estimated random effects variance. Estimation of the sampling variances of the direct CPS estimates of school-age children's poverty rates was done in several steps. The computer program developed to produce variances for complex samples (VPLX)—with successive difference replication (related to balanced-half sample replication)—was used to provide the original variance estimates for the CPS estimated state poverty rates. To reduce the instability of these variance estimates, they were modeled using a generalized variance function, which is a function of the poverty rate (e.g., $\beta y + \gamma y^2$, where y is the poverty rate) divided by the state's sample size for each year. The years 1989-1993 were used to estimate the generalized variance function. The estimated variances for the random effects were calculated using maximum likelihood estimation.

One complication of this approach is that the mean and the variance of the estimated poverty rates are linked, in the sense that the variance of an estimated proportion (p) is proportional to $p(1 - p)$. Therefore, an iteration was performed, in which the estimated variance for the sampling errors was updated to reflect new values for the model predictions. The iteration was repeated six times.

Finally, the CPS direct estimates of school-age children's poverty rates were combined with fitted values from the regression, using an empirical Bayes approach similar to that applied in county estimation. These procedures produced CPS estimates of 1989 poverty rates; the same methods were used to produce CPS estimates of 1993 poverty rates.

The estimated rates were then multiplied by either census counts (for the 1989 model) or population estimates (for the 1993 model) to arrive at estimates of the number of poor school-age children in each state. The state estimates were then benchmarked to sum to the CPS national estimate of the number of related school-age children in poverty. This adjustment was a minor one, involving multiplying the state estimates from the 1989 model by 1.0168 and those from the 1993 model by 1.0091.

APPENDIX D

Population Estimates

The Census Bureau has long had an active program of using demographic analysis to develop updated estimates of total population and population by age for various levels of geography, such as states, counties, and cities. The Census Bureau's state- and county-level models of school-age children who were poor in 1993 use county and state postcensal population estimates for age groups as of July 1994; these estimates were developed within the framework of the Census Bureau's population estimates program (Long, 1993).

TOTAL POPULATION ESTIMATES

Total population estimates are developed by the component method of demographic analysis. In general, the component method starts from an area's population in the previous census. That number is then updated by the net demographic change—adding births and international immigration and subtracting deaths and emigration. The final component, internal migration or migration to and from other parts of the United States, is currently estimated from administrative records. No adjustments are made for the estimated net population undercount in the census.

Postcensal county estimates of total population are produced by the component method, with three elements: (1) the numbers of births and deaths are based on reported vital statistics for each county; (2) reports of the Immigration and Naturalization Service are used to estimate net immigration from abroad; and (3) administrative records are used to estimate net migration between counties. Net migration of people under 65 years of age is estimated for each county from a year-to-year match of federal income tax returns; for people aged 65 and over, net

migration is estimated for each county from the change in Medicare enrollment figures (Bureau of the Census, 1995b).

The county population totals are summed for each state to provide estimates of the total population of each state. All county and state population totals are then adjusted to sum to independently derived estimates of the total U.S. population. The county estimates are also reviewed locally under the Census Bureau's Federal State Cooperative Program for Local Population Estimates.

ESTIMATES BY AGE

Estimates by age group are prepared separately, but within the framework of the total population estimates for states and counties.

Specifically, county age estimates are prepared in a two-step procedure. In the first step, estimates of total county population are developed as described above. In addition, estimates of state population by single years of age, sex, race, and Hispanic origin are developed. The state age estimates (which are controlled to the state total population estimates) use a component method in which migration rates by age are derived from school enrollment data (Bureau of the Census, 1987). In the second step, the county age estimates are developed by using a raking ratio adjustment of the estimates from the previous census. In this approach, the beginning matrix of counts for each county by age, sex, race, and Hispanic origin from the previous census is simultaneously adjusted to agree with the postcensal estimate of the total county population and the postcensal estimates for the applicable state by age, sex, race, and Hispanic origin (Sink, 1996).

This procedure means that the errors in the county estimates of an age group reflect errors in the assumption that the age distribution of each county within a state changes in the same manner as that state's age distribution. The errors in the county estimates of an age group also reflect errors in the derivation of the state estimates of age groups and errors in the derivation of the county estimates of total population.

As more data that reflect population change have become available and accessible, the methods for developing updated population estimates for geographic areas have changed, and the estimates have improved. The Census Bureau evaluates its estimates at regular 10-year intervals by comparisons with the decennial census figures. Also, the continuing research and evaluation program helps to determine the best approaches for improved performance (Davis, 1994).

ESTIMATES OF RELATED CHILDREN

For estimating numbers of poor school-age children, the Census Bureau's state- and county-level regression models use population estimates for the noninstitutionalized population under age 21 (county-level model) and the noninstitutionalized population under age 65 and the noninstitutionalized population

aged 5-17 (state-level model). These estimates are developed by subtracting administrative records counts of institutionalized people in the relevant age groups from the demographic estimates developed as described above.

Finally, the Census Bureau provides to the Department of Education estimates of *related* children aged 5-17 for counties as of July 1994 to use as denominators in calculating county poverty rates to use in the Title I allocations. (The numerators are the Census Bureau's estimates of the number of related school-age children in each county who were poor in 1993, developed as described in Appendix C.) These estimates of related children are developed by adjusting the population estimates of noninstitutionalized children aged 5-17 on the basis of the ratio of related children aged 5-17 to noninstitutionalized children aged 5-17 for each county in the 1990 census.

APPENDIX E

Future Research

This appendix briefly discusses some of the analyses conducted by the panel of the Census Bureau's county-level model, notes questions raised by these analyses, and suggests possible future investigations. Five topics are covered: questions about the specification of the county-level model in comparison with the state-level model; the possible problems from the high degree of correlation (multicollinearity) among the predictor variables in the county-level model; the effects of implementing a constrained model; the possible advantages of using other models, such as a model of change in poverty over time or a model of poverty rates or ratios; and two issues raised by the use of a logarithmic transformation of the predictor and dependent variables.

MODEL SPECIFICATION

The model used by the Census Bureau to produce updated estimates of poor school-age children for states is of a form that appears in the literature, and the estimation procedures are those that have been used previously. Sampling variances are estimated directly, and the model error component of variance is estimated as a part of the model estimation. The model uses the poverty rate as the dependent variable and rates or ratios appear as predictor variables (covariates). The estimated rate is then applied to a population estimate (obtained from demographic analyses) to obtain the estimated number of poor school-age children.

The county-level model differs from the state-level model in two notable aspects: the county-level model is expressed in terms of logarithms of counts and the state-level model is in terms of rates; the county-level model uses data from

the March 1993, 1994, and 1995 CPS in constructing the dependent variable, and the state level model uses data only from the March 1994 CPS.

The logarithm of the number of school-age children in poverty is the dependent variable of the county-level model, and the covariates are all logarithms of estimated population counts. The use of counts rather than rates in the county-level model is justified on the basis that the chosen model permits estimation of standard errors of model predictions. The Census Bureau has argued that a model for rates would not permit estimation of standard errors since there are no standard errors for the demographic population estimates. This argument applies to the estimated numbers of school-age children in poverty, which are the main consideration in Title I allocations. However, the allocations also depend on county poverty rates for school-age children, which the Department of Education produces by dividing the estimated counts by the demographic population estimates. Hence, the problems resulting from any lack of understanding of the variability in the population estimates are unavoidable. Furthermore, the Census Bureau has performed considerable evaluation of its population estimates that should make it possible to develop adequate estimates of standard errors for these purposes.

The county-level model uses a component of variance for counties that is estimated from census data (see Appendix C); this use of census data is potentially valuable, but it also raises questions. It requires the assumption that the variance of model error for the 1990 census data for poverty in 1989 is the same as the variance of model error for CPS data for poverty in 1993: this assumption can be investigated. For example, it is possible to construct direct estimates of the sampling variance of the CPS poverty estimates for a number of the larger counties in the sample. These directly computed CPS sampling variance estimates can then be compared to those obtained from the county-level model. In addition, a model for the county-level CPS sampling variances could be constructed. Given CPS sampling variances estimated from such a model, procedures analogous to those used for the state-level model could be used to estimate the variance of model error for the county-level model. The estimated variance of model error can then be compared to that estimated from the census data. The panel believes such comparisons should be conducted. It would also be possible to develop a model to combine the estimated variance of model error derived from the census with the estimate obtained from a standard small-area analysis.

The Census Bureau controls the county estimates to sum to the state estimates by means of a ratio adjustment procedure. As Table E-1 shows, this procedure produced some rather large adjustments of the county estimates. The reasons for these sizable adjustments need investigation. An alternative approach for aligning the state and county estimates would be to include a state component of variance in the county-level model. Such a model can be written as

$$y_i = \Sigma\beta_j x_{ij} + z_s + u_i + e_i \ ,$$

where z_s is the state component. Although the forms of the state- and county-level models may contribute to the sizable variability in the ratio adjustments, that variability is consistent with the existence of a state (or similar) component of variance. Smoothed estimates that are constructed with a two-component county-level model would differ from those constructed under the present procedure. To the panel's knowledge, the magnitude of the differences and the degree to which the magnitude of the state adjustments are consistent with the size of a directly estimated state component of variance have not been investigated.

The state- and county-level models are inconsistent in that if one derives a model for states from the county-level model by aggregating counties to states, it differs from the state-level model that was used. It is not a requirement that the state- and county-level models be consistent, but inconsistent models should be used only if there is a good reason to do so. In this case, the decision to use distinct models may have been primarily driven by the administrative organiza-

TABLE E-1 Ratios of State Estimates of the Number of School-Age Children in Poverty in 1993 to the Sum of Uncontrolled County Estimates for 1993, Selected States

State	Ratio of State Estimate to Sum of County Estimates
Alaska	1.33
Connecticut	1.24
Michigan	1.22
Massachusetts	1.21
West Virginia	1.16
New Jersey	1.12
Arizona	1.11
New York	1.11
Florida	1.05
California	1.01
Wyoming	1.01
Texas	0.98
Mississippi	0.98
Alabama	0.97
Illinois	0.97
Nebraska	0.94
Idaho	0.89

SOURCE: Calculated by the panel from data that were made available to the panel in January 1997. Subsequently, the Census Bureau discovered errors in the input data for a few counties that changed somewhat the 1993 estimates from the county-level model; however, the general patterns reported above hold true. (The state estimates were unchanged.)

tion of the estimation process, time pressures that made it necessary to divide up the estimation process into distinct stages, and the use of different data sources. Future research should consider a more integrated approach.

MULTICOLLINEARITY

The variables used in the county-level model have two important statistical properties. First, they are very strongly correlated; Table E-2 presents the correlation matrix of the explanatory variables. Second, all the variables are measured with error. These two properties mean that individual regression coefficients have large variances and that there is a potential for bias in the model's predictions (see below). Predictions for individual counties will not be seriously affected by variation in individual coefficients if the counties used for estimation are representative of all counties. However, the counties used for parameter estimation are not a random sample of all counties, since counties without CPS sample households with poor school-age children are not used in estimating the regression model. Therefore, the conditions under which predictions based on error-prone observations are unbiased are not satisfied.

TABLE E-2 Correlation Matrix of Independent Variables Used in 1993 County-Level Model

Variable	x_1	x_2	x_3	x_4	x_5
x_1 (tax returns, poor, < 21)	1.000	0.959	0.948	0.950	0.971
x_2 (food stamp recipients)		1.000	0.917	0.908	0.973
x_3 (population <21)			1.000	0.996	0.914
x_4 (tax returns, total, <21)				1.000	0.907
x_5 (1990 census, poor, 5-17)					1.000

SOURCE: Data from Bureau of the Census.

CONSTRAINING THE REGRESSION COEFFICIENTS TO SUM TO ONE

The estimated county-level model for poverty among school-age children in 1993 is (the standard errors of the regression coefficients are shown in parentheses)

$$\hat{y}_i = \underset{}{-0.883} + \underset{(0.150)}{0.295}\, x_{1i} + \underset{(0.081)}{0.249}\, x_{2i} + \underset{(0.068)}{0.042}\, x_{3i} + \underset{(0.218)}{0.030}\, x_{4i} + \underset{(0.219)}{0.429}\, x_{5i} \quad (0.087),$$

where, at the county level, the dependent variable is

y_i = log(3-year weighted average of the CPS number of poor school-age children in county i),

and the independent (predictor) variables are

x_{1i} = log(number of child exemptions [assumed to be under age 21] reported by families in poverty on tax returns in county i),
x_{2i} = log(number of people receiving food stamps in county i),
x_{3i} = log(estimated noninstitutionalized population under age 21 in county i),
x_{4i} = log(number of child exemptions on tax returns in county i), and
x_{5i} = log(number of poor school-age children in county i in the previous [1990] census).

The sum of the regression coefficients for the predictor variables in this model is 1.045. The fact that this sum exceeds 1 implies that the model estimates a higher poverty rate for large counties than for small counties, as is shown in Table E-3 for three hypothetical counties of different sizes, where each of the predictor variables increases across the counties directly in the proportions 1:5:25. As is shown in the final row of the table, the Census Bureau's county-level model estimates a poverty rate that is 15 percent higher for the large county than for the small county.

At the request of the panel, the Census Bureau estimated a restricted model that imposed the condition that the sum of the coefficients of the predictor variables be 1.0. The model was estimated with CPS data for the period corresponding to the 1990 census as the dependent variable. The Census Bureau's unconstrained county-level model was estimated from the same data. The estimated coefficients for the two models (constrained and unconstrained) are given in Table E-4. They are similar for three variables: number of child exemptions reported

TABLE E-3 1993 County-Level Model Estimates for Three Hypothetical Counties

Variable	County 1	County 2	County 3
x_1 (tax returns, poor, <21)	480	2,400	12,000
x_2 (food stamp recipients)	760	3,800	19,000
x_3 (population <21)	3,880	19,400	97,000
x_4 (tax returns, total, <21)	3,200	16,000	80,000
x_5 (1990 census poor 5-17)	320	1,600	8,000
$\hat{y}$ (CPS 3-year average, poor school-age children)	285	1,534	8,246
$\hat{y} / x_3$	0.0735	0.0791	0.0850

TABLE E-4 Estimated Coefficients, Unrestricted and Restricted, for County-Level Model of Poverty Among School-Age Children in 1989

Variable	Unrestricted Coefficient	Restricted Coefficient
Intercept	–0.841	–0.439
x_1 (tax returns, poor, <21)	0.505	0.436
x_2 (food stamp recipients)	0.307	0.354
x_3 (population < 21)	0.713	0.393
x_4 (tax returns, total, <21)	–0.754	–0.436
x_5 (1980 census poor 5-17)	0.270	0.253

SOURCE: Data from Bureau of the Census.

by families in poverty on tax returns, the 1980 census estimate of poor school-age children in 1979, and food stamp recipients. They differ considerably for the two size variables—number of child exemptions on tax returns and population under age 21. However, the sum of the coefficients for the unrestricted model is 1.041, close to the 1 for the restricted model.

The fitted models were used to estimate the number of school-age children in poverty in 1989 for every county; the averages of the relative differences (model estimate minus census estimate, divided by census estimate) for counties grouped by population size are reported in Table E-5. Not surprisingly, all the differences are positive since poverty rates based on CPS data are 6 percent higher than rates based on census data. It is clear that imposing a restriction on the coefficients makes a difference in the behavior of the estimates with respect to size.

The sum of the coefficients for the predictor variables is significantly different from 1.0 in the unrestricted county model. There are several possible reasons that large counties may tend to have higher poverty rates than small counties. For example, there is an urban factor: that is, counties with very large populations always include urban areas, and very small counties are always rural, although the relationship of urban areas and county size is not consistent. (As examples, some small cities constitute only a part of a populous surrounding county while the nation's largest city, New York, is divided among five counties.) County size may also be a proxy for other variables that are less obvious. However, county size may enter into the model for artifactual reasons because the county model is nonlinear and may have some biases when the population is very small (such as the removal of counties with no poor school-age children from the model). Since it is difficult to interpret and evaluate the role of size in the county model, it would be desirable, if possible, to identify the relevant characteristics for which size is a proxy and enter them into the model directly.

TABLE E-5 Percentage Differences Between 1990 Census and Model Estimates of Poor School-Age Children in 1989, for Unrestricted and Restricted County-Level Models, for Counties Grouped by Population Size

County Size Category	Unrestricted Model	Restricted Model
0 to 9,999	2.1	12.7
10,000 to 19,999	2.8	10.6
20,000 to 49,999	3.9	9.2
50,000 to 99,999	4.4	7.6
100,000 to 499,999	7.4	9.3
500,000 and over	7.4	5.9

SOURCE: Calculated by the panel from data provided by Bureau of the Census.

OTHER KINDS OF MODELS

It is possible that administrative records are not consistent for all counties. Thus, the relationship between administrative indicators and characteristics of interest may not be constant across counties. One way to reduce the effect of inconsistency in the administrative data is to model changes. A closely related procedure, which is used in the state-level model, is to include residuals from a model fit during an earlier "control" period as an explanatory variable. A change model assumes that administrative procedures have been relatively constant within any given county over the study period. The decision to use only current administrative data and not data on changes is based on the judgment that administrative procedures are more similar over areas than over time. The fact that residuals were useful in the state-level model, however, suggests that such variables should be considered for the county-level model.

An alternative to the use of a numbers model for counties is the use of a rate model. At the request of the panel, the Census Bureau estimated a county-level model with the dependent variable equal to the ratio of the CPS estimated number of poor school-age children (related children aged 5-17) to the CPS estimated population of noninstitutionalized children aged 5-17 (i.e., the same kind of dependent variable used in the state-level model). The results showed that the residual mean square error for this model was considerably less than that for the Census Bureau's county-level model. The panel believes that models for rates should be fully investigated.

USE OF THE LOGARITHMIC TRANSFORMATION

The county-level model is estimated in logarithms, and, in transforming back to the original scale of poverty counts, a correction is made to the exponentiated

values. A different correction for bias due to this nonlinearity that could be explored is to regress the original observations on the exponentiated log predicted values. (This regression format can also be used for other model checks.)

The use of a logarithmic transformation leads to a problem in the treatment of counties that contain no school-age children in poverty. All counties with no school-age children in poverty in the CPS 3-year sample were dropped from the estimation of the model's coefficients since the logarithm of 0 cannot be computed. Generally, these were counties with extremely small numbers of sampled households. Because of the large variance of poverty estimates for such counties, it is conjectured that the omission of such data has little impact on the estimates. However, it would be desirable to have analyses supporting this conjecture. Also, it would be desirable to investigate the use of a generalized linear model as an alternative modeling approach that does not require removing counties with no school-age children in poverty from the estimation of the regression coefficients.

APPENDIX
F

Special Case: Estimates for Puerto Rico

Puerto Rico is included in the Title I fund allocations. Since the commonwealth has no administrative subdivisions, the Department of Education treats it as a single unit (equivalent to a U.S. county) for the allocation of these funds. In order to incorporate Puerto Rico in the fiscal 1997 fund allocation, estimates of its number and proportion of related children aged 5-17 living in poverty are needed for 1993.

If the fiscal 1997 allocations were based on 1990 census estimates (which the panel does not recommend), the estimates for Puerto Rico could be obtained straightforwardly from the commonwealth's 1990 decennial census. From that census it is estimated that Puerto Rico had about 558,000 poor related children aged 5-17 in 1989, 66.4 percent of all related children in this age range. However, the panel recommends that the fiscal 1997 allocations be based in part on estimates of the number and proportion of school-age children in poverty in 1993, and it is not straightforward to develop such estimates for Puerto Rico.

The Puerto Rico Bureau of Labor Statistics conducts a periodic labor force survey, but that survey does not collect CPS-type income information on a regular basis. In addition, the specific model-based estimation procedures developed by the Census Bureau for U.S. states and counties cannot be applied to Puerto Rico since they are based on tax return and food stamp participation data for which there are no precise equivalents for Puerto Rico.

The only data source that appears to be available for updating estimates of poor school-age children in Puerto Rico is an experimental March 1995 income survey modeled after the CPS March Income Supplement.[1] The Census Bureau

[1]The survey was repeated in March 1997, and it is planned to repeat it at 2-year intervals.

has based its 1993 estimates of poor school-age children on data from this survey, together with data for Puerto Rico from the decennial census and updated population estimates.

The derivation of the estimates of poor school-age children in Puerto Rico in 1993 from these data sources required a number of adjustments, for several reasons: (1) the March 1995 experimental survey did not collect information on the ages of family members under 18 (so that related children aged 5-17 could not be identified among those aged under 18); (2) the updated Puerto Rico population estimates are for all children in the resident population, not for related children only; and (3) the survey, which was conducted in 1995, obtained information on 1994 income, not 1993 income. In making the adjustments, the Census Bureau assumed that certain relationships observed in 1990 census data still applied and that the change in the number of Puerto Rico school-age children in poverty between 1989 and 1994 was linear.

The panel does not have any data with which to test the validity of these assumptions. It has only limited information about the sample design, sampling and nonsampling errors, response rates, and other features of the experimental survey. The sample size of about 3,200 households should be large enough to provide a direct estimate of the number of poor school-age children with adequate precision. However, only limited information is available about other key aspects of data quality, including response rates for households to the income questions and the editing or imputation procedures used.[2]

The approach adopted by the Census Bureau for producing updated estimates of poor school-age children in Puerto Rico seems appropriate, given the data available. However, at this time the panel is unable to make a firm recommendation on how estimates of the number and proportion of children in poverty in 1993 in Puerto Rico should be made. If, on further examination, the assumptions seem reasonable and the data quality appears adequate, the panel would endorse the Census Bureau's approach. Under those conditions, the panel would then recommend that the Census Bureau's proposed estimates for Puerto Rico be treated as equivalent to 1993 U.S. county estimates in determining the fiscal 1997 Title I allocations, as discussed in Section 5 of the report.

[2]At the time of writing, we understand that the Census Bureau is obtaining additional information about the quality of the income data from the March 1995 Puerto Rico survey.

References

Bureau of the Census

1980 *Population and Per Capita Income Estimates for Local Area: Detailed Methodology and Evaluation*. Current Population Report, Series P25-699. Washington, D.C.: U.S. Department of Commerce.

1987 *State Population and Household Estimates, with Age, Sex, and Components of Change: 1981-86*. Current Population Reports, Series P-25, No. 1010. Washington, D.C.: U.S. Department of Commerce.

1990 *Money Income and Poverty Status in the United States, 1989*. Current Population Reports, Series P60-168. Washington, D.C.: U.S. Department of Commerce.

1993a Census of Population and Housing, 1990: Summary Tape File 4 Technical Documentation. U.S. Department of Commerce, Washington, D.C.

1993b *Poverty in the United States: 1992*. Current Population Reports, Series P60-185. Washington, D.C.: U.S. Department of Commerce.

1995a *Income, Poverty, and Validation of Noncash Benefits: 1993*. Current Population Reports, Series P60-188. Washington, D.C.: U.S. Department of Commerce.

1995b *Subnational Estimates of Total Population by the Tax Return Method*. Population Estimates Branch, Population Division. Washington, D.C.: U.S. Department of Commerce.

1996 *Poverty in the United States: 1995.* Current Population Reports, Series P60-194. Washington, D.C.: U.S. Department of Commerce.

Citro, C.F., and G. Kalton, eds.

1993 *The Future of the Survey of Income and Program Participation*. Panel to Evaluate the Survey of Income and Program Participation, Committee on National Statistics, National Research Council. Washington, D.C.: National Academy Press.

Citro, C.F., and R.T. Michael, eds.

1995 *Measuring Poverty: A New Approach.* Panel on Poverty and Family Assistance: Concepts, Information Needs, and Measurement Methods, Committee on National Statistics, National Research Council. Washington, D.C.: National Academy Press.

Coder, J.F., R.C. Fisher, and P.M. Siegel

1996 Making Model-Based Estimates of the Number of Related Persons Age 5-17 in Poverty for all U.S. Counties, 1994. Housing and Household Economic Statistics Division, Bureau of the Census, U.S. Department of Commerce (November).

Davis, S.
1994 *Evaluation of Post Censal County Estimates for the 1980's*. Population Division Working Paper Series, Number 5. Bureau of the Census. Washington, D.C.: U.S. Department of Commerce.

Edmonston, B. and C. Schultze, eds.
1995 *Modernizing the U.S. Census*. Panel on Census Requirements in the Year 2000 and Beyond, Committee on National Statistics, National Research Council. Washington, D.C.: National Academy Press.

Fay, R.E.
1996 Methodology for poverty estimation for 1993. Section 2, Estimates of Poverty and Income by State for Income Year 1993: Preliminary Report. Bureau of the Census, U.S. Department of Commerce.

Fay, R.E., C.T. Nelson, and L. Litow
1993 Estimation of median income for 4-person families by state. Pp. 9-1 to 9-17 in *Indirect Estimators in Federal Programs*. Statistical Policy Working Paper 21, Statistical Policy Office, Office of Information and Regulatory Affairs, Office of Management and Budget. Washington, D.C.: U.S. Government Printing Office.

Fisher, G.M.
1992 Poverty guidelines for 1992. *Social Security Bulletin* 55(1)(Spring):43-46.

Ghosh, M. and J.N.K. Rao
1994 Small-area estimation: An appraisal. *Statistical Science* 8 (1):55-93.

Long, J.
1993 *Post Censal Population Estimates: States, Counties, and Places*. Population Division Working Paper Series, Number 3. Bureau of the Census. Washington, D.C.: U.S. Department of Commerce.

Moskowitz, J., S. Stullich, and B. Deng
1993 *Targeting, Formula, and Resource Allocation Issues: Focusing Federal Funds Where the Needs are Greatest*. Washington, D.C.: U.S. Department of Education.

Platek, R., J.N.K. Rao, C.E. Särndal, and M.P. Singh, eds.
1987 *Small Area Statistics*. New York: John Wiley and Sons

Robinson, J.G., B. Ahmed, P. Das Gupta, and K. Woodrow
1993 Estimation of population coverage in the 1990 U.S. census based on demographic analysis. *Journal of the American Statistical Association* 88(423):1061-1079.

Sink, L.
1996 *Estimates of the Population of Counties by Age, Sex, Race, and Hispanic Origin: 1990-1994*. Release PE-48 (Methodology), U.S. Bureau of the Census. Washington, D.C.: U.S. Department of Commerce.

Spencer, B.D. and C.-F. Lee
1980 Postcensal population estimation methods of the Census Bureau. Pp. 131-187 in *Estimating Population and Income of Small Areas*. Panel on Small-Area Estimates of Population and Income, Committee on National Statistics, National Research Council. Washington, D.C.: National Academy Press.

U.S. Office of Management and Budget
1993 *Indirect Estimators in Federal Programs*. Statistical Policy Working Paper 21. Subcommittee on Small-Area Estimation, Federal Committee on Statistical Methodology, Statistical Policy Office, Office of Information and Regulatory Affairs. Washington, D.C.: U.S. Government Printing Office.

Zitter, M. and H.S. Shryock, Jr.
1964 Accuracy of methods of preparing postcensal estimates for states and local areas. *Demography* 1(1):227-241.

Biographical Sketches, Panel Members and Staff

GRAHAM KALTON (*Chair*) is a senior statistician and senior vice president of Westat, Inc. He is also a research professor in the Joint Program in Survey Methodology at the University of Maryland. Previously he was a research scientist in the Survey Research Center and a professor of biostatistics and statistics at the University of Michigan, professor of social statistics at the University of Southampton, and reader in social statistics at the London School of Economics. His research interests are in survey sampling and general survey methodology. He is a fellow of the American Statistical Association and of the American Association for the Advancement of Science. He has served as president of the International Association of Survey Statisticians and is the president-elect of the Washington Statistical Society. He is a past member of the Committee on National Statistics and has served as chair or a member of several of its panels. He received a B.Sc. in economics and an M.Sc. in statistics from the University of London and a Ph.D. in survey methodology from the University of Southampton.

DAVID M. BETSON is an associate professor of economics at the University of Notre and a visiting scholar at the Joint Center for Poverty Research of the University of Chicago and Northwestern University. His previous positions have been as a research associate at the Institute for Research on Poverty at the University of Wisconsin and an economist in the Office of the Assistant Secretary for Planning and Evaluation in the U.S. Department of Health, Education, and Welfare. His research examines the effects of governments on the distribution of economic well-being with special reference to the measurement of poverty and the analysis of child support policy. He received a Ph.D. degree in economics from the University of Wisconsin-Madison.

CONSTANCE F. CITRO is a member of the staff of the Committee on National Statistics. She is a former vice president and deputy director of Mathematica Policy Research, Inc., and was an American Statistical Association/National Science Foundation research fellow at the Bureau of the Census. For the committee, she has served as study director for numerous panels, including the Panel on Poverty and Family Assistance, the Panel to Evaluate the Survey of Income and Program Participation, the Panel to Evaluate Microsimulation Models for Social Welfare Programs, and the Panel on Decennial Census Methodology. Her research has focused on the quality and accessibility of large, complex microdata files, as well as analysis related to income and poverty measurement. She is a fellow of the American Statistical Association. She received a B.A. degree from the University of Rochester and M.A. and Ph.D. degrees in political science from Yale University.

MICHAEL L. COHEN is a senior program officer for the Committee on National Statistics, currently serving as study director for the Panel on Statistical Methods for Testing and Evaluating Defense Systems. Previously, he was a mathematical statistician at the Energy Information Administration, an assistant professor in the School of Public Affairs at the University of Maryland, a research associate at the Committee on National Statistics, and a visiting lecturer at the Department of Statistics at Princeton University. His general area of research is the use of statistics in public policy, with particular interest in census undercount and model validation, and in robust estimation. He received a B.S. degree in mathematics from the University of Michigan and M.S. and Ph.D. degrees in statistics from Stanford University.

NANCY E. DUNTON is a principal social scientist at the Midwest Research Institute. Formerly, she was a senior research scientist with the New York State Department of Social Services and the New York State Council on Children and Families. Her work focuses on outcome indicators and social demography, with a special emphasis on children's policy issues. She received a Ph.D. degree in sociology from the University of Wisconsin-Madison.

WAYNE A. FULLER is a distinguished professor in the Department of Statistics and Economics at Iowa State University. He is a fellow of the American Statistical Association, the Econometric Society, and the Institute of Mathematical Statistics and is the author of *Introduction to Statistical Time Series* and *Measurement Error Models*. He also has an active research program in survey sampling. He has held offices in national and international statistical organizations and has previously served on National Research Council panels.

THOMAS B. JABINE is a statistical consultant who specializes in the areas of sampling, survey research methods and statistical policy. He was formerly a

statistical policy expert for the Energy Information Administration, chief mathematical statistician for the Social Security Administration, and chief of the Statistical Research Division of the Bureau of the Census. He is a fellow of the American Statistical Association and a member of the International Statistical Institute. He has a B.S. degree in mathematics and an M.S. degree in economics and science from the Massachusetts Institute of Technology.

SYLVIA T. JOHNSON is professor of research methodology and statistics in education at Howard University, where she is also editor-in-chief of the *Journal of Negro Education*. She has served on the faculties of Roosevelt, Trenton (NJ) State, and Western Illinois universities and Augustana and Chicago City Colleges and as a visiting scholar at the Educational Testing Service. She is currently a principal investigator at the Center for Research on the Education of Students Placed at-Risk (CRESPAR), a joint activity of Howard University and Johns Hopkins University, and of the Design and Analysis Committee of the National Assessment of Educational Progress (NAEP). Formerly, she was a member of the Technical Advisory Committee for the National Adult Literacy Survey and the Graduate Record Examination and served as a trustee of the College Board. She is a fellow of the American Psychological Association and chairs its Division 15 Committee on Members and Fellows. Dr. Johnson received a Ph.D. degree in educational measurement and statistics from the University of Iowa.

THOMAS A. LOUIS is professor and head of the division of biostatistics at the University of Minnesota School of Public Health and professor of statistics in the School of Statistics. His research interests include Bayes and empirical Bayes methods, research synthesis, risk assessment, analysis of longitudinal and spatial data and sequential design of experiments. He is codirector of the statistical center for Community Programs for Clinical Research on AIDS and for 6 years directed the statistical center for the school's Cancer Prevention Research Unit. He is a senior associate editor of CHANCE, on the editorial board of *Statistics Neerlandica*, and on the board of the Medical Follow-up Agency of the Institute of Medicine. He is a member of the International Statistics Institute, a fellow of the American Statistical Association and the American Association for the Advancement of Science, and a trustee of the National Institute for Statistical Sciences. He received a B.A. from Dartmouth College and a Ph.D. in mathematical statistics from Columbia University.

SALLY C. MORTON is head of the Statistics Group at RAND in Santa Monica, California. She is on the faculty of the RAND Graduate School of Public Policy Studies and is a lecturer in the School of Public Health at the University of California-Los Angeles. Previously she taught at the School of Business of the University of Southern California, visited the Centre for Mathematical Analysis at

Australian National University and the University of Southampton, and worked at the Stanford Linear Accelerator Center. She serves as an associate editor for the *Journal of the American Statistical Association*, is chair of the association's Section on Statistical Graphics, and is a member of the Caucus for Women in Statistics and the Institute of Mathematical Statistics. Her health policy research concerns homelessness, severe mental illness, and outcomes research and quality of care in the areas of childbirth and AIDS. Her methodological research concentrates on meta-analysis, nonparametric regression, and the sampling of vulnerable populations. She received a Ph.D. in statistics from Stanford University.

JEFFREY S. PASSEL is a principal research associate at the Urban Institute and the director of the institute's Program for Research on Immigration Policy. Previously, he was assistant division chief for estimates and projections in the Population Division of the Bureau of the Census, and he also directed the research on demographic methods for measuring census undercount. His research interests include the demography of immigration, particularly the measurement of illegal immigration; the effects and integration of immigrants into American society; and measuring and defining racial and ethnic groups in the United States. He is a member of a number of professional societies and has served in various capacities in the Population Association of America, the American Statistical Association, and the American Association for the Advancement of Science. He is a fellow of the American Association for the Advancement of Science. He received a B.S. in mathematics from the Massachusetts Institute of Technology, an M.A. in sociology from the University of Texas at Austin, and a Ph.D. in social relations from the Johns Hopkins University.

J.N.K. RAO is professor of statistics at Carleton University in Ottawa, Canada, and a consultant to Statistics Canada. Formerly, he was a professor at University of Manitoba and Texas A&M University. His research interests include survey sampling theory and methods, particularly small area estimation. He is a fellow of the Royal Society of Canada, the American Statistical Association, and the Institute of Mathematical Statistics. He received a Ph.D. degree in statistics from Iowa State University.

ALLEN L. SCHIRM is senior researcher at Mathematica Policy Research, Inc. Formerly, he was Andrew W. Mellon assistant research scientist and assistant professor at the University of Michigan. His principal research interests include small-area estimation and sample and evaluation design, with application to welfare, food and nutrition, and education policy. He is a member of the American Statistical Association, the American Economic Association, and the Population Association of America. He received an A.B. in statistics from Princeton University and a Ph.D. in economics from the University of Pennsylvania.

PAUL R. VOSS is professor of rural sociology and chair of the Department of Rural Sociology at the University of Wisconsin-Madison. For the past 21 years he has been affiliated with the Wisconsin Applied Population Laboratory and is currently its director. He also is affiliated with the Wisconsin Institute for Research on Poverty. His research involves modeling small-area population change for purposes of population estimation and projection, and he also has studied and written about the demographic composition of small-area migration streams. He is a member of the U.S. Department of Commerce's Advisory Committee for the 2000 Census as well as the Census Bureau's Advisory Committee of Professional Associations. He received a Ph.D. degree in sociology (demography) from the University of Michigan.

KIRSTEN K. WEST is a mathematical statistician at the Bureau of the Census who is working under an Intergovernmental Personnel Act agreement as a research associate to the Committee on National Statistics. Her research has focused on coverage errors in the decennial census and census data quality. She has a Ph.D. degree in sociology from the University of North Carolina at Chapel Hill.

JAMES WYCKOFF is associate professor of public administration and policy at the University at Albany, of the State University of New York. His research involves applied public economics and public policy, with particular focus on the economics of education. He was an American Statistical Association Fellow at the Census Bureau. He received a Ph.D. degree from the University of North Carolina at Chapel Hill.

ALAN M. ZASLAVSKY is associate professor of statistics in the Department of Health Care Policy at Harvard Medical School. He was formerly on the faculty of the Department of Statistics at Harvard. His research interests include measurement of quality in health care, census methodology, estimation and correction of census undercount, small-area estimation, microsimulation, design and analysis of surveys, and Bayesian methods. He has served on two other panels of the Committee on National Statistics concerned with planning for the 2000 census. He received a Ph.D. degree in applied mathematics from the Massachusetts Institute of Technology.

MEYER ZITTER is an independent demographic consultant. Formerly, he was chief of the Census Bureau's Population Division and also served as assistant director for international programs. He is a fellow of the American Statistical Association and a member of the International Statistical Institute and the International Union for the Scientific Study of Population. He has a B.B.A. degree from City College of New York.