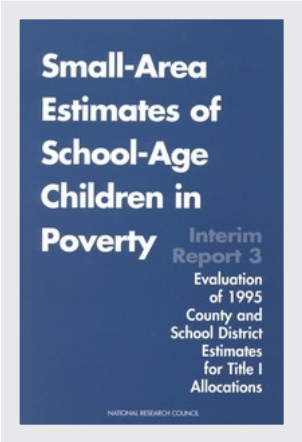

Small-Area Estimates of School-Age Children in Poverty: Interim Report 3

DETAILS

134 pages | 6 x 9 | PAPERBACK
ISBN 978-0-309-06395-1 | DOI 10.17226/6427

AUTHORS

Constance F. Citro and Graham Kalton, Editors; Committee on National Statistics, National Research Council

Small-Area Estimates of School-Age Children in Poverty

Interim Report 3: Evaluation of 1995 County and School District Estimates for Title I Allocations

Constance F. Citro and Graham Kalton, *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. 1999

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. Support of the work of the Committee on National Statistics is provided by a consortium of federal agencies through a grant from the National Science Foundation (Number SBR-9709489). 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.

International Standard Book Number 0-309-06395-7

Printed in the United States of America.

PANEL ON ESTIMATES OF POVERTY FOR SMALL GEOGRAPHIC AREAS

GRAHAM KALTON (*Chair*), Westat, Rockville, Maryland
DAVID M. BETSON, Department of Economics, University of Notre Dame
NANCY E. DUNTON, Midwest Research Institute, Kansas City, Missouri
WAYNE A. FULLER, Department of Statistics, Iowa State University
THOMAS B. JABINE, Consultant, Washington, D.C.
SYLVIA T. JOHNSON, School of Education, Howard University
THOMAS A. LOUIS, School of Public Health, University of Minnesota
SALLY C. MORTON, RAND, Santa Monica, California
JEFFREY S. PASSEL, Urban Institute, Washington, D.C.
J.N.K. RAO, Department of Mathematics and Statistics, Carleton University
ALLEN L. SCHIRM, Mathematica Policy Research, Inc., Washington, D.C.
PAUL R. VOSS, Department of Rural Sociology, University of Wisconsin–Madison
JAMES H. WYCKOFF, Graduate School of Public Affairs, State University of New York, Albany
ALAN M. ZASLAVSKY, Department of Health Care Policy, Harvard Medical School

CONSTANCE F. CITRO, *Study Director*
MICHAEL L. COHEN, *Senior Staff Officer*
MICHELE VER PLOEG, *Research Associate*
MEYER ZITTER, *Consultant*
TELISSIA M. THOMPSON, *Senior Project Assistant*

Acknowledgments

The Panel on Estimates of Poverty for Small Geographic Areas wishes to thank the many people who contributed to the preparation of this, the panel's third interim report.

We thank, first, the staff of the Census Bureau who prepared the state, county, and school district estimates of poor school-age children in 1995, many of whom also worked on the evaluations of those estimates: David Aultman, William Bell, Patrick Cardiff, Robert Fay, Robin Fisher, Matthew Kramer, Esther Miller, Ronald Prevost, Douglas Sater, Paul Siegel, Cotty Armstrong Smith, Alexander Strand, Jess Thompson, George Train, David Waddington, and Signe Wetrogan. We also thank the Census Bureau staff who facilitated the arrangements for the work: Cynthia Clark, Nancy Gordon, Charles Nelson, and Daniel Weinberg.

Daniel Kasprzyk of the National Center for Education Statistics, who serves as project officer for the study for the U.S. Department of Education, was helpful as always in facilitating this phase of the panel's work. The panel also appreciates the continued help of other Department of Education staff–in particular, Sandy Brown, Kay Rigling, William Sonnenberg, and Stephanie Stullich–in providing information and educating us about the Title I allocation process.

I thank my panel colleagues for their continued commitment to the work of the panel and to its third interim report, which was prepared, like the first two reports, under a very demanding time schedule. I particularly thank James Wyckoff, who, with his student Frank Papa, evaluated the use of school lunch data in comparison with the Census Bureau's method for developing school

district estimates of poor school-age children in New York State; that work appears in the appendix.

The panel was assisted by a very able staff. As always, Constance Citro did an outstanding job as the study director. She had primary responsibility for drafting this report, and without her exceptional writing skills and dedication, the report could not have been produced in the time available. Michael Cohen made important contributions to the evaluation of the models and to many sections of the report. Michele Ver Ploeg and James Sexton (a National Research Council summer intern) prepared tabulations for the panel of school lunch and other data. Meyer Zitter ably assisted the panel's working group on school district estimates. Telissia Thompson provided excellent administrative support for the study and for the preparation of the report. Eugenia Grohman, associate director for reports of the Commission on Behavioral and Social Sciences and Education, improved the report through her fine technical editing. To all we are grateful.

Our report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the NRC's Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making the published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. We wish to thank the following individuals for their participation in the review of this report: Johnny Blair, Survey Research Center, University of Maryland; James R. Chromy, Research Triangle Institute, Research Triangle Park, N.C.; Emerson J. Elliott, National Council for the Accreditation of Teacher Education, Washington, D.C.; Lyle V. Jones, L.L. Thurstone Psychometric Laboratory, University of North Carolina, Chapel Hill; Roderick J.A. Little, School of Public Health, University of Michigan; Lincoln E. Moses, Department of Biostatistics, Stanford University Medical Center (emeritus); William O'Hare, The Annie E. Casey Foundation, Baltimore, Md.; John W. Pratt, Graduate School of Business, Harvard University (emeritus); and Franklin D. Wilson, Department of Sociology, University of Wisconsin–Madison.

While the individuals listed above have provided constructive comments and suggestions, it must be emphasized that responsibility for the final content of this report rests entirely with the authoring panel and the institution.

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

Contents

Small-Area Estimates of School-Age Children in Poverty

Executive Summary

BACKGROUND

The U.S. Department of Education uses estimates of school-age children in poverty to allocate federal funds under Title I of the Elementary and Secondary Education Act for education programs to aid disadvantaged children. Historically, the allocations have been made by a two-stage process: the department's role has been to allocate Title I funds to counties; the states have then distributed these funds to school districts. Until recently, the department has based the county allocations on the numbers and proportions of poor school-age children in each county from the most recent decennial census. States have used several different data sources, such as the decennial census and the National School Lunch Program, to distribute the department's county allocations to districts.

In 1994 Congress authorized the Bureau of the Census to provide updated estimates of poor school-age children every 2 years, to begin in 1996 with estimates for counties and in 1998 with estimates for school districts. The Department of Education is to use the school district estimates to allocate Title I basic and concentration grants directly to districts for the 1999-2000 and later school years, unless the Secretaries of Education and Commerce determine that they are "inappropriate or unreliable" on the basis of a study by the National Research Council. That study is being carried out by the Committee on National Statistics' Panel on Estimates of Poverty for Small Geographic Areas.

Under a direct allocation procedure, there would be no allocations to counties and, hence, no need for states to distribute them to school districts. However, a provision in the 1994 legislation permits states to aggregate the department's

allocation amounts for all districts in a state that have fewer than 20,000 people and to redistribute the aggregate amount among those districts by using some other method that the department approves.

UPDATED SCHOOL DISTRICT ESTIMATES

The Census Bureau's procedure for producing updated school district estimates of poor school-age children is a basic synthetic shares approach, in which the proportions of poor school-age children in school districts within each county in 1989 (as measured by the 1990 census) are applied to updated estimates of numbers of poor school-age children from a statistical model for counties. The Census Bureau decided that the most recent school district estimates it could produce by the end of 1998 (for the Title I allocations in spring 1999) were for school-age children in 1996 who were living in and related to a family in poverty in 1995. Reasons for this decision included the time required to ascertain the changes in school district boundaries since the 1990 census and the 1-2 year lag in the availability of the data sources used in the county statistical model.

The synthetic shares method assumes that the shares of poor school-age children among school districts in each county in 1995 are the same as they were in 1989. Consequently, the synthetic estimates reflect only the changes in school-age poverty from 1989 to 1995 that occurred in each county as a whole. The estimates do not capture any variation in school-age poverty among the districts within each county that occurred since the 1990 census.

The synthetic shares method was used because no administrative records data are available for a model for school districts (which would be similar to the Census Bureau's county model) that could capture changes in poverty for school districts within counties. There are several reasons for the lack of data and the difficulties of developing estimates for school districts: most districts are small in size, many district boundaries do not coincide with the boundaries for counties or other governmental units, district boundaries can and often do change, and some districts do not serve all elementary and secondary grades.

ASSESSMENT AND CONCLUSION

In assessing the Census Bureau's 1995 school district estimates of the numbers of poor school-age children for use in Title I allocations for the 1999-2000 school year, the panel first examined the 1995 county estimates that were produced by the Census Bureau's statistical model. Although the Department of Education would not use the county estimates for Title I allocations if it were to make allocations directly to school districts, the county estimates are central to the synthetic shares method for district estimates.

The model that was used to produce the 1995 county estimates is essentially the same model that was used to produce county estimates of poor school-age

children for 1993. On the basis of internal and external evaluations that were conducted of alternative 1993 county models, which resulted in some changes in the Census Bureau's original 1993 county model, the panel supported the use of revised 1993 county estimates for Title I allocations for the 1998-1999 school year (see National Research Council, 1998). Additional evaluations of the 1995 county model, which focused on its behavior when estimated for several time periods, confirmed that the county model is performing well. A separate estimation procedure for Puerto Rico, which is treated as a single county and school district for Title I allocations, also appears to be reasonable, given the available data.

Evaluations of the Census Bureau's synthetic procedure for school districts over the 1980-1990 period revealed large differences for many districts between the synthetic estimates of poor school-age children and the comparison estimates from the 1990 census; the large differences occur mainly for small districts. In contrast, the estimates for school districts with 40,000 or more people in 1990 are not markedly worse than the county model estimates. Also, a number of districts are coterminous with counties, so that their estimates come from the county model. Together, these two groups of districts comprise only 13 percent of the districts (as of 1990), but they contain 62 percent of all school-age children.

Although the Census Bureau's 1995 estimates of poor school-age children have potentially large errors for many school districts, the panel nonetheless concludes that they are not inappropriate or unreliable to use for direct Title I allocations to districts as intended by the 1994 legislation. In reaching this conclusion, the panel interprets "inappropriate and unreliable" in a relative sense. Some set of estimates must be used to distribute Title I funds to school districts. The panel concludes that the Census Bureau's 1995 estimates are generally as good as–and, in some instances, better than–estimates that are currently being used. Also, while further research is needed, a limited evaluation suggests that school lunch data are not appreciably better than the 1990 census for constructing within-county school district shares of poor school-age children.

A benefit of using the synthetic shares estimates is that the department would be able to determine eligibility of school districts for both basic and concentration Title I grants on the basis of a consistent set of estimates nationwide. Also, use of the synthetic shares estimates for direct allocation of concentration grants would respond to the intent of the 1994 legislation that eligible districts be able to receive concentration grants even when they are in counties that would not be eligible under the current two-stage allocation process.

The Census Bureau's updated estimates of poor school-age children for counties are the only postcensal small-area estimates of poverty that have been thoroughly evaluated. It is important that they be considered in the direct allocations to school districts, as is done when the allocations are based on the synthetic estimates. If a state chooses to reallocate the amounts for school districts with less than 20,000 population, the county estimates can be reflected in the alloca-

tions by grouping the allocations for small size districts by county and redistributing the county totals to those districts.

RECOMMENDATIONS

(1) The panel recommends to the Secretaries of Education and Commerce that the Census Bureau's 1995 school district estimates of poor school-age children be used to make direct Title I allocations to school districts for the 1999-2000 school year.

(2) The panel recommends that any state plan approved by the Department of Education for redistributing the sum of the department's allocations for school districts with under 20,000 population maintain the county total amounts for such districts to the extent possible.

The Department of Education should undertake a thorough study of the direct allocation of Title I funds to school districts, which will be a new procedure for the 1999-2000 school year. The study should examine the allocation methods used and assess the results.

FUTURE RESEARCH AND DEVELOPMENT

It is important to continue an active program of research and development for methods of estimating poverty for school-age children at the county and school district levels. The county model is performing well, but, like other models, it can probably be improved. Work should also be pursued to improve the current synthetic shares method for school district estimates. Research on ways to produce the estimates with data that are closer in time to the year for which the allocations are to be made should also be pursued.

Improving school district estimates so that they reflect within-county, as well as between-county, changes in school-age poverty over time will require a substantial research and development effort. It is particularly important to obtain relevant administrative records data for districts, such as income tax return data coded to the district level. Such administrative data, together with data from the 2000 census and the planned American Community Survey, could provide the means to develop a much improved model-based approach for estimating school-age poverty at the district level.

For its work in small-area poverty estimation, the Census Bureau needs to provide adequate staff and other resources on a continuing basis. Because small-area estimates of poverty support a range of important public policy needs for federal, state, and local governments, the Bureau's program should include not only data and model development and production, but also thorough evaluation and detailed documentation of each set of estimates produced.

1

Introduction

Small-area estimates of poverty for school-age children are used by the U.S. Department of Education to allocate funds under Title I of the Elementary and Secondary Education Act, which supports compensatory education programs to meet the needs of educationally disadvantaged children (see Moskowitz et al., 1993). Title I allocations for the 1998-1999 school year totaled over $7 billion. Until now, the department's role has been to allocate Title I funds to the nation's more than 3,000 counties (including Puerto Rico as a county equivalent), and the states have then distributed the county funds to school districts. For the 1999-2000 school year, the intent of legislation passed in 1994 is for the department to make allocations directly to almost 15,000 school districts (formally known as local educational agencies, LEAs).

Historically, the Title I allocations made by the Department of Education to counties used poverty estimates from the most recent decennial census for which data were available. The estimates from one census were used for a decade or more until estimates from the next census became available. Since the proportions and numbers of children in poverty can change significantly over time, Congress in 1994 authorized the Bureau of the Census to provide updated estimates of school-age children in poverty every 2 years, beginning in 1996 for the Title I allocations for counties for the 1997-1998 and 1998-1999 school years and in 1998 for the Title I allocations for school districts for the 1999-2000 and later school years. Having the most up-to-date estimates possible is important so that resources can be directed toward areas that are most in need.[1]

[1]See National Research Council (1997:Ch. 2; App. B) for data on the significant changes that occurred in the numbers and proportions of poor school-age children between the 1980 and 1990 censuses and following the 1990 census.

The Title I allocations are based on estimates of eligible children: predominantly, children aged 5-17 in families with incomes below the poverty level,[2] but also children in foster homes, children in families above the poverty level that receive Aid to Families with Dependent Children (AFDC),[3] and children in local institutions for neglected and delinquent children. At present, funds are provided for two different types of allocations—basic grants and concentration grants:

—Basic grants allocate funds to all counties and to school districts that have at least 10 formula-eligible children *and* whose percentage of formula-eligible children exceeds 2 percent of the district's total school-age children.

—Concentration grants allocate funds only to counties and school districts with high numbers (6,500 or more) or high proportions (more than 15%) of formula-eligible children.

The allocation amounts for both basic and concentration grants depend primarily on the number of eligible children in a county or school district; they also take into consideration the state's average per-pupil expenditure.[4] Currently, the formulas for basic and concentration grants include a 100 percent hold-harmless provision so that no county or school district may receive less than its previous year's allocation.

Congress also authorized a study—through the Department of Education—by a panel of the National Research Council's Committee on National Statistics to review the Census Bureau's small-area poverty estimates for school-age children. The statute requires that the Department of Education use the Census Bureau's updated estimates for Title I allocations unless the Secretaries of Commerce and Education determine that "some or all of the data" are "inappropriate or unreliable" on the basis of the panel'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 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. The panel began its work in June 1996 and is

[2]The poverty status of individuals is determined by comparing the before-tax money income of their families 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 National Research Council (1995) for an 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.

[3]The Personal Responsibility and Work Opportunity Reconciliation Act of 1996 abolished AFDC and replaced it with Temporary Assistance to Needy Families (TANF).

[4]For details of the allocation process, see National Research Council (1997:App. A). With direct allocation by the Department of Education to school districts, the provisions for county eligibility and grant amounts would no longer apply.

scheduled to work through 1999, producing a final report at that time and such interim reports as are needed.

UPDATED ESTIMATES

County Estimates

The Census Bureau was initially charged to produce updated estimates of poor school-age children at the county level for use in Title I allocations for the 1997-1998 school year. For this purpose, the Census Bureau provided county estimates of the number of children aged 5-17 in1994 from families with incomes below the poverty level in 1993.[5] The estimates were developed from a statistical model that used administrative data from Internal Revenue Service and Food Stamp Program records for 1993, estimates of poor school-age children in 1989 from the 1990 census, and 1994 population estimates to predict county numbers of poor school-age children in 1993 as measured in the March Income Supplement to the Current Population Survey (CPS). To increase the reliability of the predictions, the model used a weighted average of 3 years of data from the March 1993, 1994, and 1995 CPS, covering income in 1992, 1993, and 1994. The estimates from the county model were calibrated to estimates from a similar statistical model for states.

The data used in the county model are obtained from several sources, and most data are not available until 2 years after the period to which they refer. When the developmental work began in 1994, the Census Bureau decided that it could not expect to produce estimates in time for the 1997-1998 allocations for a later year than 1993, given the time required for acquiring, processing, and applying the data for a new statistical model.

In its first interim report (National Research Council, 1997), the panel reviewed the Census Bureau's modeling approach favorably but concluded that there had not been sufficient time to thoroughly evaluate the updated estimates produced by the specific model that the Bureau developed. As an interim solution for Title I allocations for the 1997-1998 school year, the panel recommended that the 1993 county estimates be averaged with 1990 census estimates. This recommendation was adopted. Subsequently, the Census Bureau completed an extensive evaluation of the county model, modified it in several respects, and

[5]More precisely, the Census Bureau's estimates pertain to related children aged 5-17 in poor families, termed "poor school-age children" in this report. Related children in families include all members of a household who are under 18 years of age and related to the householder by birth, marriage, or adoption, except the spouse of the householder. Foster children are not included since they are not related to the householder, who is the person in whose name the house is owned or rented (see Bureau of the Census, 1993).

produced a revised set of 1993 county estimates of poor school-age children. In its second interim report (National Research Council, 1998), the panel recommended that the revised 1993 county estimates be used for Title I allocations for the 1998-1999 school year, which was done.

For both the 1997-1998 and 1998-1999 school years, the Department of Education used the Census Bureau's poverty estimates to make allocations to counties. As in the past, the states then allocated the county amounts to school districts. The states used a variety of data sources for these allocations: many states used 1990 census data wholly or in part; some states used such data sources as estimates of children enrolled in the National School Lunch Program or children in families receiving AFDC in each district. In some states in which the boundaries of school districts bear little relationship to county boundaries, the department has permitted the state to ignore the county allocations in dividing up the total allocation amount for the state among school districts. The Department of Education must approve a state's allocation plan but not the specific estimates used by a state or the allocation amounts.

School District Estimates

For Title I allocations for the 1999-2000 school year, the 1994 legislation charges the Census Bureau to provide the Department of Education with updated estimates of poor school-age children for school districts. The legislation charges the department, in turn, to make direct allocations to school districts rather than to counties unless the Secretaries of Education and Commerce determine that the school district estimates are inappropriate or unreliable for this purpose, taking into account the panel's recommendations.

Under this procedure, the Department of Education would use the Census Bureau's estimates together with district-level estimates of the other groups of formula-eligible children (e.g., children in foster homes) to make allocations to districts according to the provisions of the formula. The states would not be involved. However, a provision in the 1994 legislation permits a state to aggregate the Department of Education's allocation amounts for all school districts in the state that are estimated to have fewer than 20,000 people. The state may then, using a method and data source approved by the department (e.g., school lunch data), divide up the total allocation for these districts in a manner different from the department's original allocations. This provision is a significant one because about four-fifths of school districts contain fewer than 20,000 people, although these districts contain only 27 percent of total school-age children in the United States.

There appear to be several reasons that Congress in the 1994 legislation deemed it desirable for the Department of Education to make direct allocations to school districts. First, direct allocations by the department impose a measure of consistency on the allocation process (leaving aside the proviso for states to

reallocate the amounts for districts with fewer than 20,000 people). Second, direct allocations to school districts solve a problem with the concentration grant formula in which a county may not be eligible for a concentration grant, but one or more of the school districts in the county may meet the eligibility criteria. (This can happen when a poor school district is located within a county that, on average, is not poor enough to qualify.) Under the current two-stage allocation process, poor school districts in counties that do not qualify for a concentration grant receive less funds than they would receive with direct allocations.[6] Finally, if adequate data were available for estimation, the use of updated school district-level estimates in the allocations would take account of changes that have occurred since the previous census in the distribution of poverty among school districts within counties.

In fall 1998 the Census Bureau provided estimates for school districts of the numbers of children aged 5-17 in 1996 who were living in families with incomes below the poverty level in 1995. It has not been easy to develop reliable updated estimates for counties, and the task is much more difficult for school districts. Some school districts are the same as counties; however, most school districts are smaller than counties, the boundaries of many of them cross county lines, and the boundaries can and often do change over time. Also, some school districts provide education for specific grade levels, such as kindergarten-8 or 9-12. Largely because of these complicating factors, there is a paucity of data for developing updated poverty estimates at the school-district level: there are currently no school district equivalents of federal income tax return or Food Stamp Program data that are used in the Census Bureau's state and county estimation models.

Because of the lack of data at the school district level, the Census Bureau's procedure for developing 1995 school district poverty estimates uses a simple model that assumes that the proportions or shares of poor school-age children in school districts within each county in 1995 are the same as they were in 1989 (as measured by the 1990 census). The estimation procedure involves the following steps: 1990 census data are retabulated to match 1995-1996 school district boundaries (determined from a special survey); the proportion of the county total of poor school-age children in the 1990 census is determined for each school district (or part of a school district) in the county; and the 1990-based proportions are then applied to updated 1995 county estimates from the Census Bureau's county model to produce 1995 school district estimates.[7]

Because of the time required to complete the survey of 1995-1996 school district boundaries and the time lags in the availability of data for the county

[6]States may reserve up to 2 percent of their concentration grant funds to allocate to such districts.

[7]The 1995 school district estimates and the 1993 and 1995 county estimates are available on the Census Bureau's web site: http://www.census.gov/hhes/www/saipe.html.

model, the Census Bureau was not able to produce school district estimates for later than 1995. Moreover, the Census Bureau's "synthetic" estimation procedure does not capture intracounty variation in the extent to which school-age poverty has increased or decreased among school districts between 1989 and 1995. However, the estimation procedure does produce estimates more recent than the census, it is consistent across the nation, and it responds to the concern that concentration grants be directed to all eligible school districts, including those in counties that are not eligible.

PLAN OF THE REPORT

This, the panel's third interim report, assesses the Census Bureau's 1995 school district estimates for use in Title I allocations for the 1999-2000 school year. The report contains five chapters and an appendix.

Chapter 2 briefly describes the Census Bureau's procedure for obtaining updated county estimates of the numbers and proportions of poor school-age children in 1995 and summarizes the evaluations of those estimates. Although the Department of Education does not use county estimates in Title I allocations when the allocations are made directly to school districts, the county estimates are central to the method used by the Census Bureau to derive updated school district estimates and, therefore, to an evaluation of those estimates.

Chapter 3 describes and evaluates, as best as can be done, the Census Bureau's procedure for obtaining 1995 school district estimates of poor school-age children. Given the scarcity of data with which to implement alternative estimation procedures for school districts, the opportunities for evaluation are equally very limited. Chapter 3 also describes the Census Bureau's procedure for obtaining, from its population estimates program, 1996 school district estimates of the total number of school-age children and the total population in each district.

Chapter 4 provides the panel's assessment of the 1995 school district estimates and its recommendations for 1999-2000 Title I allocations. Chapter 5 outlines research and development activities for further work on developing updated county and school district estimates of poor school-age children.

The appendix uses data for the state of New York to illustrate the results of an alternative procedure for developing updated school district estimates that is based on counts of participants in the National School Lunch Program.

2

County Estimates

Reliance on the most recent decennial census to allocate federal funds to counties and other small areas has primarily reflected the absence of alternative data sources with comparable or superior reliability. Mindful of the need for small-area estimates that are more up to date than census estimates, the Census Bureau organized the Small Area Income and Poverty Estimates (SAIPE) Program to develop methods for producing postcensal income and poverty estimates for states and counties by using multiple data sources and innovative statistical methods. The program began in late 1992 with financial support from a consortium of five federal agencies. Congress made this work more urgent by passing legislation in 1994 that charged the Census Bureau to produce updated estimates of poor school-age children for counties and school districts every 2 years, to begin in 1996 with estimates for counties, discussed in this chapter, and in 1998 with estimates for school districts, discussed in Chapter 3.

The SAIPE Program faces a challenging task to produce county-level estimates. For Title I allocations, there is no single administrative or survey data source that provides sufficient information with which to develop reliable direct estimates of the number and proportion of school-age children in families in poverty by county. The March Income Supplement to the Current Population Survey (CPS) can provide reasonably reliable annual direct estimates of such population characteristics as the number and proportion of poor children at the national level and possibly for the largest states. However, the CPS cannot provide direct estimates for the majority of counties because the sample does not include any households in them. And for almost all of the counties with households in the CPS sample (about 1,250 of a total of 3,143 counties in 1995), the

estimates have a high degree of sampling variability.[1] Nonetheless, the CPS data may serve as the basis for creating usable estimates for counties through the application of statistical estimation techniques to develop "model-based" or "indirect" estimates.

Model-based or indirect estimators use data from several areas, time periods, or data sources (which could include the previous census) to "borrow strength" and improve the precision of estimates for small areas. A model-based approach is needed when there is no single data source for the area and time period in question that can provide direct estimates that are sufficiently reliable for the intended purpose. The Census Bureau has 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).

This chapter provides a summary description and evaluation of the model-based approach used by the Census Bureau to develop estimates by county of the number and proportion of school-age children in families in 1996 who were poor in 1995 (referred to as the 1995 county estimates). A document prepared by the Census Bureau describes the estimation procedure and evaluations of the 1995 estimates in detail (Bureau of the Census, 1998; see also National Research Council, 1998:Chs. 3, 4, Apps. C, D on the evaluations of the 1993 estimates).

If the Department of Education uses the Census Bureau's 1995 school district estimates of poor school-age children for direct allocation of Title I funds to districts, the 1995 county estimates will not be used directly. However, the 1995 county estimates are critical to the development of 1995 school district estimates. As a result of the lack of data at the school-district level, the Census Bureau has been constrained to use for school districts a very simple model-based method referred to as synthetic estimation, which applies the shares of poor school-age children for the school districts in a county according to the 1990 census to the updated 1995 county estimates to obtain updated school district estimates (see Chapter 3).[2] Therefore, in order to evaluate the 1995 school district estimates, it is essential to understand and evaluate the 1995 county estimates.

[1]For a description of the March CPS and differences between income and poverty data from the CPS and the 1990 census long-form sample, see National Research Council (1997:Ch. 2; App. B). The 1990 census sample includes households in all counties and covers 15 million households, 300 times more than the 50,000 households in the CPS, yet even the 1990 census estimates are relatively variable for some small counties (National Research Council, 1997:Table 2-1).

[2]We use the term "synthetic estimation" for the Census Bureau's shares procedure for school district estimates and distinguish it from the statistical regression modeling that was done for the state and county estimates. However, synthetic estimation is sometimes used more broadly in the small-area literature.

ESTIMATION PROCEDURE

The Census Bureau's estimation procedure for counties uses two regression models that predict poor school-age children—a county model and a separate state model. The estimation procedure was first used to develop the 1993 county estimates. It includes the following steps summarized below: (1) a regression model is developed to provide initial estimates of the number of poor school-age children at the county level; (2) a state model is developed to produce estimates of the number of poor school-age children by state; and (3) the initial county-level estimates are adjusted so that the final estimates for counties within each state sum to the state-level estimates. In addition, the Census Bureau produces county population estimates of the total number of school-age children, which the Department of Education has used to calculate estimated proportions of poor school-age children for counties. Finally, the Census Bureau produces separate estimates of poor school-age children for Puerto Rico.

Step 1: County Model

The first step in the estimation process is to develop and apply the Census Bureau's county model to produce initial estimates of the numbers of poor school-age children. This step involves:

—obtaining data from the March CPS for three consecutive years to construct a dependent variable in a county model regression equation that is the estimated log number of poor school-age children for counties with households in the CPS sample;

—obtaining data from administrative records and other sources that are available for all counties to construct predictor variables for the regression equation;

—specifying and estimating the regression equation to relate the predictor variables to the dependent variable; and

—using the estimated regression coefficients from the equation and the predictor variables to develop estimates of poor school-age children for all counties.

For counties with households in the CPS sample, the predictions from the model are then combined by a "shrinkage" procedure with the CPS direct estimates (on a logarithmic scale) for those counties. (The shrinkage procedure weights the two sets of estimates according to their relative precision; see Fay and Herriot [1979], Ghosh and Rao [1994], and Platek et al. [1987] on shrinkage methods.) The initial county estimates are then obtained by transforming the predictions from the logarithmic to the numeric scale.

The county model equation takes the following form:

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

where:

y_i = log(3-year weighted average of poor school-age children in county i),
x_{1i} = log(number of child exemptions 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 population under age 18 in county i),
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 of the dependent variable for county i.

The predictor variables in the county equation for the 1995 estimates are based on data from Internal Revenue Service (IRS) records for 1995 (x_{1i}, x_{4i}), Food Stamp Program records for 1995 (x_{2i}), the Census Bureau's population estimates program for 1996 (x_{3i}), and the 1990 census (x_{5i}).[3] As the dependent or outcome variable, the county equation uses county estimates of the number of poor school-age children averaged over 3 years of the March CPS (data from the March 1995, 1996, and 1997 CPS, covering income in 1994, 1995, and 1996).[4]

The relationships between the predictor variables and the dependent variable in equation (1) are estimated solely from the subset of counties that have households in the March CPS sample. This subset includes proportionately more large counties and proportionately fewer small counties than the distribution of all counties. Because values of zero cannot be transformed into logarithms, a number of counties whose sampled households contain no poor school-age children are excluded from the estimation. In all, 985 of the country's 3,143 counties were included in the 1995 model estimation.

Step 2: State Model

The second step in the estimation process is to develop and apply the Census Bureau's state model to produce estimates of the number of poor school-age children by state. The state estimation is similar to that for counties, although the state model differs from the county model in several respects.[5]

[3]Variables x_{3i} and x_{4i} are included in the model in order to cover children not reported on tax returns (i.e., in nonfiling families), who are assumed to be poorer on average than other children.

[4]See Bureau of the Census (1998) and National Research Council (1998:Ch. 2) for the derivation of the 3-year weighted average of poor school-age children from the CPS and of the last two terms in the equation (u_i and e_i).

[5]See National Research Council (1998:Ch. 2) for a detailed review of the forms of the state and county models and the differences between them.

The state model equation takes the following form:

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

where:

y_i = proportion of school-age children in state i that are poor, estimated from one year of the CPS (March 1996 CPS for the 1995 model),

x_{1i} = proportion of child exemptions reported by families in poverty on tax returns in state i,

x_{2i} = proportion of people receiving food stamps in state i,

x_{3i} = proportion of people under age 65 who did not file an income tax return in state i,[6]

x_{4i} = residual for state i from a regression of the proportion of poor school-age children from the most recent decennial census on the other three predictor variables,[7]

u_i = model error for state i, and

e_i = sampling error of the dependent variable for state i.

All states have sampled households with poor school-age children in the CPS; however, the variability associated with estimates from the CPS is large for some states. As is done for the initial county estimates, the predictions from the state model and the CPS direct estimates are combined in a shrinkage procedure 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 are multiplied by estimates of the total number of noninstitutionalized school-age children from the Census Bureau's program of population estimates. Finally, the state estimates of the numbers of poor school-age children are adjusted to sum to the CPS national estimate of related school-age children in poverty. This adjustment is a minor one; for 1995 it changed the state estimates by less than one-half of 1 percent.

Step 3: Combining the County and State Estimates

The last step in the estimation process is to adjust the initial estimates of poor school-age children from the county model (step 1) for consistency by state with the estimates from the state model (step 2) to produce final estimates of the

[6]This percentage is obtained by subtracting the estimated number of exemptions on income tax returns for people under age 65 from the estimated total population under age 65 that is derived from demographic analysis (see National Research Council, 1998:App. B).

[7]For the 1995 state model, x_{4i} is the residual from a regression of poor school-age children from the 1990 census on the other three predictor variables for 1989.

numbers of related children aged 5-17 in poverty by county. The estimate for each state from the state model is divided by the sum of the estimates for each county in that state to form a state raking factor. Each of the county estimates in a state is multiplied by the state raking factor so that the sum of the adjusted county estimates equals the state estimate. For the final county estimates of poor school-age children in 1995, the average state raking factor was 0.97; two-thirds of the factors were between 0.88 and 1.06.

Differences Between 1995 and 1993 Estimation Procedures

The procedure summarized above to produce the 1995 county estimates differs in a few respects from the procedure that was used to produce the revised 1993 estimates described in the panel's second interim report (National Research Council, 1998). The changes involved the input data for the state and county models:

- An error in processing the 1989 IRS data was discovered and corrected. The corrected data were used to reestimate the decennial census equation that provides the residual predictor variable in the 1995 state model (x_{4i} in equation (2)). The corrected data were also used to reestimate the 1989 state and county models for evaluation purposes.
- Several changes were made to the food stamp data for input to the state model: instead of using data for July, the number of food stamp recipients was changed to a 12-month average centered on January 1 of the following year; counts by state of the numbers of people who received food stamps due to specific natural disasters were obtained from the Department of Agriculture and subtracted from the counts of the total number of recipients; time-series analysis of monthly state food stamp data from October 1979 through September 1997 was used to smooth outliers; and food stamp recipient data for Alaska and Hawaii were adjusted downward to reflect the higher eligibility thresholds for those states.
- The food stamp numbers for the county model were raked to the adjusted state food stamp numbers.
- In both the state and county models, child exemptions reported by families on tax returns were redefined to include children away from home in addition to children at home. This change may increase the number of IRS poor child exemptions in households with children away from home both because of the additional children and because poverty thresholds are higher for larger size families.

Population Estimates

To accompany county estimates of school-age children in 1996 who were in poor families in 1995, the Census Bureau produced county-level estimates of the

total number of children aged 5-17 for 1996 from its demographic population estimates program. The estimates from step 3 above and the population estimates can then be used to calculate estimated proportions of poor school-age children for counties. The Census Bureau also produced county-level estimates of total population for 1996. The population estimates pertain to July of the year following the one for which poverty status is estimated. A detailed description and evaluation of the Census Bureau's population estimates procedures for counties is provided in National Research Council (1998:App. B).

Puerto Rico

Estimates of poor school-age children for Puerto Rico, which is treated as a county equivalent in the allocation formula, are developed separately. The county model cannot be used for them because there are no precise equivalents for Puerto Rico of tax return and food stamp data to form predictor variables for the model.

The original estimates for Puerto Rico of school-age children in 1994 who were poor in 1993 were developed with data from an experimental March 1995 income survey modeled after the CPS March Income Supplement, together with data from the decennial census and updated population estimates. 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 16 (so that related children aged 5-17 could not be identified among those aged under 18); (2) the updated Puerto Rico population estimates were 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, 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 sample size of the experimental survey of about 3,200 households appeared large enough to provide a direct estimate of the number of poor school-age children with adequate precision. However, only limited information was available about other key aspects of data quality, including household response rates on the income questions and the editing or imputation procedures used. Hence, it was difficult to evaluate the quality of the 1993 estimates for Puerto Rico, although the estimation procedures seemed appropriate given the data available.

The Puerto Rican Family Income Survey is now an ongoing survey, conducted at 2-year intervals. The Census Bureau used income data from the 1996 survey, in which about 2,300 households were interviewed in February-March 1997, together with decennial census data and updated population estimates for Puerto Rico, to construct estimates of school-age children in 1996 who were poor in 1995. The three adjustments that were made for the 1993 estimates were also

required. The change in the number of children in poor families between 1994 and 1996 was assumed to be linear. Additional information was obtained from Puerto Rico about the quality of the income survey, which in general, supported the use of the survey data to develop estimates of poor school-age children in the commonwealth (see Santos and Waddington, 1999).

EVALUATION

The development of model-based estimates for small areas is a major, continuing research and development effort for which extensive evaluation is required. For updated estimates of poor school-age children for counties, a thorough assessment of all aspects of the estimation procedure is necessary to have confidence in the estimates—whether the estimates are used by the Department of Education to allocate Title I funds to counties (as has been the practice up to now) or to develop estimates for school districts.

Since there are no absolute criteria for what are acceptable evaluation results, one method for determining if the performance of a model can be improved is to examine alternative models. Such comparisons may indicate changes that would be helpful for a model; they may also suggest that an alternative model is preferable. As summarized above, the Census Bureau's county estimates of poor school-age children are produced by using a county regression model, a state regression model, and county population estimates developed with demographic analysis techniques. A comprehensive evaluation for each of these components of the estimation procedure should include "internal" and "external" evaluations.

An internal evaluation is primarily an investigation of the validity of the underlying assumptions and features of a model. For a regression model, an internal validation is typically based on an examination of the residuals from the regression—the differences between the predicted and reported values of the dependent variable for each observation. In an external evaluation, the estimates from a model are compared with target or "true" values that were not used to develop the model. Ideally, an internal evaluation of regression model output should precede external evaluation. Changes made to the model to address concerns raised by the internal evaluation would likely improve its performance in the external evaluation. Both internal and external evaluations should be carried out for alternative models.

In its second interim report, the panel reviewed a series of internal and external evaluations that were conducted for the revised 1993 county estimates of poor school-age children (National Research Council, 1998:Ch. 4, Apps. B, C, D). The state model and the county population estimates were examined as well, both directly and as they contributed to the county estimates of poor school-age children. The evaluation determined that the revised procedure for developing updated county estimates, which principally involved a change in one of the

predictor variables in the original county model,[8] produced estimates for 1993 that were appropriate for use in allocating Title I funds to counties.

Because the 1995 county estimates were developed by using a procedure similar to that used to develop the revised 1993 county estimates, the focus of the evaluation effort for the 1995 estimates shifted to how the state and county models behave over several time periods, and specifically, to determining whether there are persistent biases or other problems. The evaluations of the 1995 county estimates, which are described in this chapter, included:

(1) internal evaluation of the regression output for the 1995 county model estimated for 1995, 1993, and 1989 (using uncorrected and corrected tax return data);

(2) comparison of estimates of poor school-age children that were developed from the 1995 form of the county model for 1995, 1993, and 1989 with CPS estimates for groups of counties, a form of external evaluation; and

(3) evaluation of the state model, including examination of regression output for 1996, 1995, 1993, 1992, 1991, 1990, and 1989 and consideration of the state raking factors by which county model estimates are adjusted to make them consistent with the state model estimates.

County Model Internal Evaluations

The first test of a regression model is that it perform well when evaluated internally, that is, for the set of observations for which it is estimated. The evaluation of the county regression output pertains to the regression model itself, that is, before the predictions are combined with the direct CPS estimates in a shrinkage procedure or raked to the estimates from the state model. The regression output comprises the model predictions for counties that have at least one household with poor school-age children in the CPS sample. We first summarize the evaluation work done on the 1993 county model predictions and then detail the work on the 1995 county model predictions.

1993 Evaluation

As part of the evaluation of the revised 1993 county estimates (National Research Council, 1998:Ch. 4 and App. C), the panel and the Census Bureau examined the underlying assumptions of 13 alternative county models through

[8]The predictor variable x_{3i} in equation (1) was changed from the estimated population under age 21 to the estimated population under age 18. This change improved the model predictions, particularly for groups of counties classified by the percentage of group quarters residents (see National Research Council, 1998:Ch.2).

evaluation of the regression model output for 1989 and 1993. The models varied on three dimensions: treatment of information from the previous census (bivariate or single-equation), form of the variables (poverty rates or numbers, transformed to logarithms or not transformed), and whether the model included fixed state effects. Although an evaluation of the regression output would not likely provide conclusive evidence with which to rank the performance of alternative models, particularly when they use different transformations of the dependent variable, such an examination could help determine which models perform reasonably well.

The assumptions examined included:

- linearity of the relationships between the dependent variable and the predictor variables, assessed by examining a variety of graphical plots;
- constancy of the assumed linear relationship over different time periods, assessed through comparison of the regression coefficients on the predictor variables for the years for which the model was estimated;
- whether any of the included predictor variables are *not* needed in the model, evaluated by looking for insignificant *t*-statistics for the estimated values of individual regression coefficients, and, conversely, whether other potential predictor variables *are* needed in the model, evaluated by looking for nonrandom patterns, indicative of possible model bias, in the distributions of standardized residuals displayed for categories of counties;[9]
- normality (primarily symmetry and moderate tail length) of the distribution of the standardized residuals;
- whether the standardized residuals have homogeneous variances, that is, whether the variability of the standardized residuals is constant across counties and does not depend on the values of the predictor variables; and
- absence of outliers.

The analysis for the most part supported the assumptions for the 13 models that were examined; it did not strongly support one model over another. A few problems characterized all or most of the models. First, most models tended to

[9]The standardization of the residuals involved estimating the predicted standard errors of the residuals, given the predictor variables, and dividing the observed residuals by the predicted standard errors. The predicted standard error of the residual for a county is a function of the estimated model error variance and the estimated sampling error variance (see Belsley et al., 1980).

The categories of counties were specified in terms of: census region, census geographic division, metropolitan status of county, population size in 1990, population growth from 1980 to 1990, percentage of poor school-age children in 1980, percentage of Hispanic population in 1990, percentage of black population in 1990, persistent poverty from 1960 to 1990 for rural counties, economic type for rural counties, percentage of group quarters residents in 1990, and number of households in the CPS sample.

overpredict the number of poor school-age children in larger urban counties, especially those with large percentages of Hispanics. Second, all models showed evidence of some variance heterogeneity, particularly with respect to CPS sample size and often with respect to the predicted value (number or proportion of poor school-age children). Some of the models exhibited more problems with skewness and outliers than others. Finally, according to the internal evaluation, none of the other models was clearly superior to the revised Census Bureau 1993 county model.

1995 Evaluation

The internal evaluation for the 1995 county model focused on comparisons of the properties of the model when estimated for different time periods. The analysis looked in particular at three characteristics: the constancy of the regression coefficients on the predictor variables over time; distributions (box plots) of the standardized residuals for categories of counties to determine if there were any nonrandom patterns that persisted over time; and the phenomenon observed in the 1993 evaluations by which the variance of the standardized residuals was related to CPS sample size and the predicted value of the dependent variable (variance heterogeneity).

Constancy of the Regression Coefficients Because the county model is refitted for each prediction year, constancy of the regression coefficients for the predictor variables over time is not as important as it would be if the estimated regression coefficients from the model were used for predictions for subsequent years. Also, major changes in economic conditions would be expected to cause some changes in the coefficients. Nonetheless, it is desirable for the coefficients to be in the same direction and not fluctuate wildly in size over time.

Table 2-1 shows the regression coefficients for the predictor variables for the 1995 county model estimated for 1995 and 1993 and for 1989 with corrected IRS data and with original (uncorrected) IRS data.[10] The coefficients for the three "poverty level" predictor variables—child exemptions reported by families in poverty on tax returns (column 1), food stamp recipients (column 2), and poor school-age children from the previous census (column 5)—are fairly similar in the equations for all three time periods. There are more substantial differences across the three time periods in the size of the estimated coefficients for the other two variables–population under age 18 (column 3) and total number of child exemptions on tax returns (column 4). However, the sum of these two coeffi-

[10]The regressions for 1995 and for 1989 with corrected IRS data also used modified food stamp data (i.e., the county food stamp data were raked to the adjusted state food stamp data, as described above).

TABLE 2-1 Estimates of Regression Coefficients for Census Bureau 1995 County Model, Estimated for 1989, 1993, and 1995

Year	No. of Counties	Predictor Variables[a] (1)	(2)	(3)	(4)	(5)
1989 (revised IRS data)	1,028	0.52 (.06)	0.29 (.06)	1.55 (.31)	–1.56 (.30)	0.26 (.06)
1989 (original IRS data)	1,028	0.50 (.06)	0.23 (.05)	1.79 (.27)	–1.80 (.27)	0.32 (.07)
1993	1,184	0.38 (.08)	0.27 (.07)	0.65 (.24)	–0.59 (.24)	0.34 (.09)
1995	985	0.31 (.10)	0.29 (.08)	0.88 (.25)	–0.80 (.25)	0.33 (.09)

NOTE: All predictor variables are on the logarithmic scale for numbers. Standard errors of the estimated regression coefficients are in parentheses.

[a]Predictor variables: (1) number of child exemptions reported by families in poverty on tax returns; (2) number of people receiving food stamps; (3) population under age 18; (4) total number of child exemptions on tax returns; (5) number of poor school-age children from previous (1980 or 1990) census.

cients is close to zero in each year. Because the two variables are highly positively correlated and close in magnitude, the predictions from equations with a similar sum for the two coefficients will be similar.

Finally, the sum of all the coefficients is close to 1 for all 3 estimation years: 1.01 for 1995, 1.05 for 1993, and 1.06 for 1989 with the revised IRS data. It is desirable for the coefficients in a model of this form to sum to 1, which indicates that the model predictions do not vary by the scale of the predictor variables. If the sum of the coefficients is much greater than or less than 1, the model should be examined to determine if additional predictor variables or other changes in the model may be needed.

Patterns of Residuals Given typical random variation, it is likely that the distributions of standardized residuals will display apparently nonrandom patterns for some categories of counties in a particular year. However, if the distributions display the same patterns across years, it is evidence of model bias. The persistence of the same patterns should be investigated to determine ways to eliminate or reduce the bias, for example, by adding a variable to the equation. (There are ample degrees of freedom in the county model to permit the inclusion of additional predictor variables.)

Investigation of the standardized residuals for categories of counties for the

county model estimated for 1995, 1993, and 1989 reveals little evidence of persistent bias. However, there is some suggestion that the model tends to consistently overpredict the number of poor school-age children in smaller size counties (i.e., the model estimates are somewhat higher than the CPS direct estimates for smaller counties). It also tends to overpredict the number of poor school-age children in counties that are in metropolitan areas but are not the central county in the area. These patterns, while not strong, are evident in the regression output for all 3 years

Variance Heterogeneity The regression output for the 1995 county model clearly demonstrates variability in the size of the absolute standardized residuals as a function of the predicted value (number of poor school-age children) and the CPS sample size. If the variance estimates for the model are correct, then the standardized residual variance should remain constant over the distribution of CPS sample size, but it increases with increasing CPS sample size. This phenomenon was evident in the evaluations conducted for the 1993 county model, and it is evident in all 3 years for which the 1995 county model was estimated.

Adjusting a model to remove this type of heterogeneity is likely to have only a small effect on the estimated regression coefficients or the model estimates (although it will affect the estimated confidence intervals around the model estimates). The effect on estimates of the number of poor school-age children would stem from two factors: a shift in the weights assigned to each county in fitting the regression model, which would very likely result in only a modest change in the estimated regression coefficients; and a change in the weight given to the direct estimates, which could have an appreciable effect on the estimates only for the few counties with large CPS sample sizes.

Nonetheless, it is clear that the current method for estimating the variance of the sampling errors (e_i in equation (1)) in the county model is incorrect. The current approach essentially obtains the total sampling error variance by estimating the total squared error for the model and subtracting from that estimate the estimated model error variance from a 1989 equation in which 1990 census data form the dependent variable. The total sampling error variance is then distributed to counties by assuming that the sampling error variance in a county is inversely proportional to the county's CPS sample size.

The Census Bureau is investigating an alternative approach that would estimate the CPS sampling variances for larger counties on the basis of direct calculations of these variances, which take account of the clustered sample design within these counties, and then develop a generalized variance function for modeling the sampling variances by using the directly estimated variances as a dependent variable. The variance of the model error (u_i in equation 1) would then be calculated by subtracting the sampling variance from the total squared error, thus avoiding the questionable assumption that the model variances for the 1989 census equation and the CPS equations are equal.

The Census Bureau should pursue its work on a generalized variance function for the county model. The Bureau should also investigate the use of some function other than proportionality to the inverse of sample size to distribute the total sampling error variance to counties to eliminate the pattern of an increase in the standardized residual variance with increases in CPS sample size. (The Census Bureau is currently examining the assumption that the sampling error variances at the county level are inversely proportional to the square root of the county's CPS sample size.) The effects of changes in the estimates of sampling error variance and model variance on the estimates of poor school-age children can then be assessed.

Summary The panel concludes that the analysis of the regression output for the 1995 county model estimated for 1989, 1993, and 1995 largely supports the assumptions of the model: there is little evidence of important problems with the assumptions. However, the model does exhibit a few minor problems that appear to persist over time. First, it tends to overpredict the number of poor school-age children in smaller counties and metropolitan counties that are not the central county compared with other counties. The differences are not marked, but research should be conducted to determine possible ways to modify the model to eliminate or reduce this problem. Second, the model shows evidence of variance heterogeneity with respect to both CPS sample size and poverty rate. The function that is used to distribute the total sampling error variance to counties should be changed to eliminate or reduce this problem, while the Census Bureau pursues longer term research on direct estimates of CPS county-level sampling variances (see Chapter 5).

County Model External Evaluations

Before using the estimates of a model for such important public policy purposes as allocating Title I funds, it is important to perform as much external evaluation of the estimates as is possible, with target values that were not used to develop the model. We first briefly review the external evaluations that were conducted for alternative 1993 county models, estimated for 1989 and 1993, and then summarize some additional external evaluations that were conducted for the 1995 county model estimated for three time periods.

1993 Evaluations

1990 Census Comparisons As part of the evaluation of the revised 1993 county estimates, the panel and the Census Bureau compared the estimated numbers and proportions of poor school-age children for 1989 for seven alternative models with 1990 census estimates (National Research Council, 1998:Ch. 4 and

App. D).[11] The 1990 census comparisons were also performed for four simpler procedures that relied much more heavily on 1980 census estimates, such as a procedure that assumed the same distribution of poverty among counties within a state in 1989 as was found in the 1980 census. The evaluation examined the overall difference between the estimates from a model or procedure and the census and the differences for groups of counties categorized by various characteristics.[12] It addressed the accuracy of model estimates for the prediction year, that is, for 1989; it did not address the issue that model-based estimates may be used for Title I allocations for a school year several years after the prediction year.[13]

The 1990 census estimates that were used in the model-census comparisons were ratio adjusted by a constant factor to make the census-based national estimate of poor school-age children equal the 1989 CPS national estimate. This adjustment removed the difference of about 5 percent between the CPS and census estimates of total poor school-age children for 1989. Consequently, the differences between a model and the 1990 census in estimating poor school-age children for groups of counties can be interpreted as differences in shares. This feature is useful because the Title I allocation formula distributes funding as shares (percentages) of a fixed dollar amount.

External evaluation by comparison with the 1990 census is not ideal because the census estimates are not true values: they are affected by sampling variability and population undercount, and the census measurement of poverty differs from the CPS measurement in ways that are not fully understood (see National Research Council, 1997:Ch. 2, App. B; see also the Census Bureau's web site: http://www.census.gov/hhes/www/saipe93/inputs/cencpsdf.html). In addition, there

[11]The county estimates reflect the effects of the state model and the county population estimates, as well as the county regression model, but the differences in model performance vis-a-vis the census in the evaluation are due to the particular form of the county model. Fewer models were evaluated externally by comparison with the 1990 census than were included in the internal evaluation of regression diagnostics (7 versus 13 models); lack of data prevented estimating the bivariate model formulations for 1989.

The models for which the 1990 census comparisons were performed were estimated with the method of moments. Maximum likelihood is used to estimate the 1995 county model. The differences in the estimates from the two techniques are small.

[12]The categories were specified in terms of: census geographic division, metropolitan status of county, population size in 1990, population growth from 1980 to 1990, percentage of poor school-age children in 1980, percentage of Hispanic population in 1990, percentage of black population in 1990, persistent poverty from 1960 to 1990 for rural counties, economic type for rural counties, percentage of group quarters residents in 1990, whether the county had households in the CPS sample in 1989-1991, and percentage change in the poverty rate for school-age children from 1980 to 1990.

[13]Research should be conducted to reduce the time lag between the prediction year for model-based estimates and the year for Title I allocations to the extent possible (see Chapter 5).

is only one census-based validation opportunity, 1990. Because of the lack of IRS and Food Stamp Program data for counties for 1979, it is not possible to evaluate model-based estimates by comparison to the 1980 census. Reliance on a single validation using the 1990 census is a problem because a model may perform better or worse in any one validation than it would on average when used over multiple years. In the absence of other means of external validation, however, the panel and the Census Bureau relied heavily on the 1990 census comparisons to understand the performance of alternative models.

The 1990 census comparisons produced a large volume of statistics and assessments (see National Research Council, 1998:Ch. 4). From its examination, the panel concluded that the models that were tested performed better than the simpler procedures that were tested. The models exhibited smaller overall absolute differences of their estimates of poor school-age children from the census estimates than did the simpler procedures. Also, for most categories of counties, the algebraic differences between the model-based estimates and the census estimates were smaller and exhibited fewer obvious patterns across categories than did the differences for the simpler procedures.

Comparing alternative models, the panel found that there were some county characteristics for which some or all models exhibited poor performance in terms of the spread between the largest and smallest algebraic category differences, the pattern of the differences across categories, or the size of the differences. Even on these characteristics, the models generally performed better than the simpler procedures. There were also some characteristics for which all models performed well.

Of the alternative models, the panel concluded that, on balance, the revised 1993 county model performed somewhat better than the other models. The only problems evident for this model were that it tended to overpredict the number of poor school-age children in counties that experienced the greatest decline in the poverty rate for school-age children from 1980 to 1990, counties with large percentages of Hispanic residents, and counties in the Mountain and Pacific Divisions. Also, it tended to underpredict the number of poor school-age children in counties that experienced the greatest increase in the poverty rate for school-age children from 1980 to 1990.

One would not expect any model to perform particularly well for the counties that experienced the largest changes (increase or decrease) in the poverty rate for school-age children from 1980 to 1990. This variable is closely related to the variable that the models are trying to estimate, and any regression model can only partially predict which cases will have the most extreme values of the outcome variable. The overprediction for counties in the West Region (Pacific and Mountain Divisions), given that the county estimates are raked to the state estimates from the Census Bureau's state model, must be attributable to the state model. Yet the evaluations showed that the state raking procedure improved the esti-

mates for counties categorized by geographic division in comparison with a procedure that made no adjustments by state.

Local Assessment of 1993 County Estimates The panel performed another type of external evaluation of the county estimates of poor school-age children–the use of local knowledge. Using the original 1993 model estimates for all 3,143 counties in the United States, the analysis first sought to identify groups of counties for which the 1993 estimates seemed unusually high or low in relation to prior levels and trends (e.g., from 1980 to 1990) in the number and proportion of poor school-age children and known social and economic trends for these groups of counties. Then, local people–including staff and members of local councils of government, economic development authorities, welfare agencies, state demographic units, state data centers, and other agencies—were contacted to obtain their assessment of the reasonableness of the implied trends in poverty for school-age children given their knowledge of local socioeconomic conditions.[14]

Individuals with local knowledge expressed a great deal of concern about the statistical reliability of the original 1993 county estimates, which was mostly consistent with the Census Bureau's own cautions in this regard, coupled with specific county estimates that seemed on the basis of local knowledge to be doubtful. These concerns notwithstanding, no categories of counties were identified that experienced apparent trends in the number and proportion of poor school-age children between 1989 and 1993 that were not accepted by knowledgeable local people. The trends for a few counties were not accepted locally, but the analysis found no strong indicators of potential bias for groups of counties sharing common characteristics in the county model.

1995 Evaluations

For the 1995 county model external evaluations, the emphasis shifted to finding a way to look for persistent bias. An apparent bias identified in a single validation, such as the 1990 census comparisons summarized above, may be a one-time effect that will not occur in other years for which a model is estimated. For any particular year, it is almost inevitable that the differences between the model estimates and target values will be somewhat larger for some categories of counties than others. But if such differences persist for the same categories of counties over time, some areas may continually receive more funding than if the true values were known, and other areas may continually receive less funding.

[14]This evaluation was carried out at the University of Wisconsin–Madison by Dr. Paul Voss, a member of the panel, with the assistance of Richard Gibson and Kathleen Morgen (see Voss et al., 1997). The evaluation used the original 1993 county estimates because the revised estimates were not available at the time.

As a type of external validation by which the issue of persistent bias could be examined, the panel and the Census Bureau compared estimates of poor school-age children from the 1995 county model for categories of counties for 1989, 1993, and 1995 with CPS direct estimates for those categories for the three periods. Three years of CPS data were used to form the weighted estimates in each case in order to reduce the sampling variability.[15]

Table 2-2 shows the difference in the number of poor school-age children from the county model, estimated for 1989 (using corrected IRS data), 1993, and 1995, and the weighted 3-year CPS direct estimates centered on those years for categories of counties. The measure shown is the algebraic difference by category, which is the sum for all counties in a category of the algebraic (signed) difference between the model estimate of poor school-age children and the weighted CPS direct estimate, divided by the sum of the weighted CPS direct estimates for the category.

Comparisons with weighted CPS direct estimates have the advantage over comparisons with the census that they can be performed for multiple years. They have the disadvantage that the sample sizes for CPS estimates, even aggregated for 3 years, are small for many categories of counties, thus making the comparisons much more uncertain than the 1990 census comparisons because of the much greater variability in the standard of comparison. Also, in analyzing the CPS comparisons, one must keep in mind that the model estimates are raked to the state estimates, which are developed from a single year of the CPS.

The model-CPS aggregate differences in Table 2-2 differ widely among categories of counties, in large part because of the small sample sizes for the CPS estimates, even when aggregated for 3 years. Some of the differences are very large, larger than any of the differences seen in the model-1990 census comparisons (see National Research Council, 1998:Table 4-3, column b). Generally, the larger model-CPS aggregate differences are for categories of counties with smaller numbers of CPS sample households. For example, the model-CPS aggregate differences often exceed 5 percent for counties grouped into the nine geographic divisions, but they are all less than 5 percent for counties grouped into the four geographic regions.[16]

In addition, the model-CPS aggregate differences for 1989 frequently differ from the model-1990 census differences. This finding is expected, given that the measurement of poverty differs between the census and the CPS because of the many differences in data collection procedures.

---

[15]This analysis is not the same as the analysis of regression output described above, in which the standardized residuals from the model for counties with sampled households in the CPS–representing the standardized differences between the model estimates and the direct estimates on the log scale–were examined for categories of counties.

[16]For future evaluations of this type, the Census Bureau should develop estimates of the standard errors of the differences so that significant differences between the model estimates and the CPS 3-year aggregate estimates can be identified.

TABLE 2-2 Comparison of County Model Estimates with CPS Aggregate Estimates of the Number of Poor School-Age Children, 1995, 1993, and 1989: Algebraic Difference by Category of County (in percent)

Category	No. of Counties[a] (1)	Model-CPS, 1995[b] (2)	Model-CPS, 1993[b] (3)	Model-CPS, 1989[b] (4)	Sample Size, CPS 1996[c] (5)
Census Region[d]					
Northeast	217	–2.87	0.81	–4.36	10,708
Midwest	1,055	–0.49	0.61	–4.31	11,393
South	1,425	4.05	–0.13	4.48	15,440
West	444	–4.16	–0.95	–0.43	12,141
Census Division[d]					
New England	67	–13.51	1.87	27.07	3,696
Middle Atlantic	150	0.05	0.54	–9.79	7,012
East North Central	437	–6.10	–0.64	–3.04	6,841
West North Central	618	18.31	4.25	–7.44	4,552
South Atlantic	591	1.82	0.83	4.12	8,150
East South Central	364	–5.53	–5.85	9.32	2,529
West South Central	470	12.00	1.90	2.44	4,761
Mountain	281	–3.91	19.87	0.84	5,543
Pacific	163	–4.24	–6.48	–0.92	6,598
Metropolitan Status					
Central county of metropolitan area	493	–2.75	–0.91	–3.53	34,343
Other metropolitan	254	53.75	–3.64	8.44	2,801
Nonmetropolitan	2,394	1.24	3.50	8.32	12,538
1990 Population Size					
Under 7,500	525	–17.21	57.03	0.74	933
7,500-14,999	630	19.82	–23.67	–0.19	1,550
15,000-24,999	524	2.94	6.24	17.02	2,289
25,000-49,999	620	30.46	–0.23	–4.46	4,204
50,000-99,999	384	–2.52	4.99	22.47	5,979
100,000-249,999	259	17.27	12.12	–3.88	8,263
250,000 or more	199	–7.24	–2.49	–3.10	26,464
1980 to 1990 Population Growth					
Decrease of more than 10.0%	444	–2.71	–22.03	–4.29	2,170
Decrease of 0.1-10.0%	972	–4.31	2.44	–1.32	10,655
0.0-4.9%	547	6.04	3.41	3.18	8,015
5.0-14.9%	620	1.12	5.97	4.61	11,590
15.0-24.9%	260	–0.07	–4.11	–10.44	9,305
25.0% or more	292	–0.52	–2.27	10.31	7,947

continued

TABLE 2-2 Continued

Category	No. of Counties[a] (1)	Model-CPS, 1995[b] (2)	Model-CPS, 1993[b] (3)	Model-CPS, 1989[b] (4)	Sample Size, CPS 1996[c] (5)
Percentage of Poor School-Age Children, 1980					
Less than 9.4%	516	2.74	7.22	–1.07	14,980
9.4-11.6%	524	1.39	5.28	4.35	12,291
11.7-14.1%	530	–10.01	–6.49	–6.72	9,837
14.2-17.2%	523	1.28	–5.82	0.44	5,217
17.3-22.3%	519	9.32	17.41	0.23	4,623
22.4-53.0%	523	1.05	–14.81	4.11	2,734
Percentage Hispanic, 1990					
0.0-0.9%	1,770	1.26	–0.75	3.13	12,848
1.0-4.9%	847	9.33	1.45	4.32	16,966
5.0-9.9%	193	–2.81	17.24	6.38	6,999
10.0-24.9%	181	–4.02	–5.14	–8.29	7,236
25.0-98.0%	150	–7.90	–3.29	–5.26	5,633
Percentage Black, 1990					
0.0-0.9%	1,446	8.32	8.02	5.09	10,929
1.0-4.9%	615	7.41	1.04	–1.83	10,630
5.0-9.9%	294	5.41	–2.07	0.95	8,646
10.0-24.9%	381	–4.89	–0.75	3.51	13,437
25.0-87.0%	405	–6.85	–2.82	–6.30	6,040
Persistent Rural Poverty, 1960-1990[e]					
Rural, not poor	1,740	–2.62	1.53	5.47	9,734
Rural, poor	535	22.45	–0.15	14.81	1,698
Not classified	866	–1.28	–0.28	–2.68	38,250
Economic Type, Rural Counties[e]					
Farming	556	–24.56	–29.31	–12.41	1,634
Mining	146	46.97	27.59	40.67	901
Manufacturing	506	–7.10	–3.58	–1.51	2,369
Government	243	120.13	27.59	59.39	1,661
Services	323	–12.18	–12.42	–11.86	2,760
Nonspecialized	484	6.99	18.35	23.89	2,018
Not classified	883	–1.18	–0.20	–2.59	38,339
Percentage of Group Quarters Residents, 1990					
Less than 1.0%	545	3.32	22.03	16.60	3,494
1.0-4.9%	2,187	–1.58	–1.27	–1.84	41,648
5.0-9.9%	299	11.90	–1.22	4.51	3,980
10.0-41.0%	110	49.44	–6.28	17.02	560

TABLE 2-2 Continued

Category	No. of Counties[a] (1)	Model-CPS, 1995[b] (2)	Model-CPS, 1993[b] (3)	Model-CPS, 1989[b] (4)	Sample Size, CPS 1996[c] (5)
Change in Poverty Rate for School-Age Children, 1980-1990					
Decrease of more than 3.0%	536	–3.88	–11.16	–10.04	4,038
Decrease of 0.1-3.0%	649	–4.57	2.63	4.44	12,658
0.0-0.9%	272	2.16	–2.75	9.66	5,102
1.0-3.4%	621	–1.07	0.11	–5.06	14,660
3.5-6.4%	532	9.09	–2.60	–0.66	7,507
6.5-38.0%	523	–1.07	5.17	3.98	5,719

[a]3,141 counties are assigned to a category for most characteristics; 3,135 counties are assigned to a category for 1980-1990 population growth and 1980 percentage of poor school-age children; 3,133 counties are assigned to a category for 1980-1990 percent change in poverty rate for school-age children.

[b]The formula, where there are n counties (i) in category (j), Y_{model} is the estimated number of poor school-age children from the county model, and Y_{CPS} is the estimated number of poor school-age children from a 3-year weighted average of the CPS, is

$$\Sigma_i \, (Y_{\text{model } ij} - Y_{\text{CPS } ij}) \, / \, \Sigma_i Y_{\text{CPS } ij} \, .$$

[c]Number of households (unweighted) in the sample for the March 1996 CPS is shown to give an idea of the relative sample sizes for each category. The 3-year weighted averages are based on 3 years' worth of sample, although some sample cases are the same for 2 years because of the rotational design.

[d]Census region and division states:

Northeast
- New England: Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut
- Middle Atlantic: New York, New Jersey, Pennsylvania

Midwest
- East North Central: Ohio, Indiana, Illinois, Michigan, Wisconsin
- West North Central: Missouri, Minnesota, Iowa, North Dakota, South Dakota, Nebraska, Kansas

South
- South Atlantic: Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida
- East South Central: Kentucky, Tennessee, Alabama, Mississippi
- West South Central: Arkansas, Louisiana, Oklahoma, Texas

West
- Mountain: Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah, Nevada
- Pacific: Washington, Oregon, California, Alaska, Hawaii

[e]The Economic Research Service, U.S. Department of Agriculture, classifies rural counties by 1960-1990 poverty status and economic type. Counties not classified are urban counties and rural counties for which a classification could not be made.

SOURCE: Data from Bureau of the Census.

Despite the sample size limitations, Table 2-2 can inform an assessment of the performance of the county model if the results are used with caution. Of particular interest are instances in which the model-CPS aggregate differences are both large and in the same direction (plus or minus) for all 3 years for which the county model is estimated. Such findings suggest a possible systematic bias in the model that should be investigated to determine the nature of the bias and what steps could be taken to eliminate or reduce it (e.g., by adding a predictor variable to the model). Several persistent patterns are evident in the model-CPS aggregate differences:

- The model shows a tendency to underpredict the number of poor school-age children in the largest counties, those with 250,000 or more population. This finding is consistent with the results from analyzing the distribution of the standardized residuals from the regression output. The extent of the underprediction is not large, but it appears to be significant given the large number of CPS households in the largest counties.
- The model shows a tendency to underpredict the number of poor school-age children in counties with large percentages of Hispanic residents (10% or more). There is a similar, although less pronounced, tendency for the model to underpredict the number of poor school-age children in counties with large percentages of blacks. It is likely that counties with large percentages of Hispanics or blacks are not homogeneous (e.g., large-percentage black counties include both inner-city and rural areas). Hence, further research is needed to determine whether the underprediction is more or less pronounced for particular subgroups of these counties and, consequently, what steps are appropriate to ameliorate the bias in the model.
- The model estimates are consistently very different from the weighted CPS estimates for some categories of rural counties classified by economic type. In particular, the model estimates for rural counties characterized as government are much higher than the corresponding weighted CPS estimates. Although the comparisons by economic type are based on small CPS sample sizes, it seems worthwhile to examine some of these counties to see if a reason for these large differences can be found.
- Finally, the model shows a tendency to underpredict the number of poor school-age children in counties that experienced the largest declines in the poverty rate for school-age children from 1980 to 1990. As was noted above, this finding is consistent with the knowledge that any regression model can only partially predict which cases will have the most extreme values of the outcome variable.

Summary

Considering both the external evaluations of alternative models that were conducted for the revised 1993 county model and the external evaluations of 3

years of estimates that were conducted for the 1995 county model, the panel concludes that the county model is working reasonably well. However, further investigation is needed of categories of counties for which the model appears to overpredict or underpredict the number of poor school-age children, particularly when that phenomenon is evident for several periods.

State Model Evaluation

The state model plays an important role in the production of county estimates of poor school-age children. Evaluations conducted of the state model for the assessment of the revised 1993 county estimates included an internal evaluation of the regression output for 1989 and 1993 and an external evaluation that compared 1989 estimates from the model with 1990 census estimates of proportions of poor school-age children. The results in each case supported the use of the model. However, the state model evaluations were more limited than the county model evaluations, as alternative state model formulations were not evaluated explicitly.

For the assessment of the 1995 county estimates, further evaluations were conducted of the state model. In particular, the model was estimated for 7 years—1989, 1990, 1991, 1992, 1993, 1995, and 1996—and the regression output for those years was examined to determine if there were any systematic biases in the model estimates. (The model was not estimated for 1994 because the redesign of the CPS sample, consequent to the 1990 census, was partly but not completely phased in for the March 1995 CPS.) Also, there was an evaluation of the state raking factors for 1993 and 1995.

State Model Regression Output

The state regression model is a poverty rate model with the variables not transformed (see equation (2)). The analysis of the regression output for the state model for 1989-1993 and 1995-1996 examined the same assumptions that were examined for the 1995 county model estimated for 1989, 1993, and 1995. The analysis is somewhat less informative for the state model than for the county model because there are about 1,000 counties with poor school-age children in the CPS, but only 51 states (including the District of Columbia), and states are collectively much more homogeneous than counties with respect to poverty rates and other characteristics. In addition, with respect to both internal and external evaluation, some categories of states do not contain enough states for analysis, thereby reducing the utility of evaluation.

Nonetheless, examination of the regression output for the state model helps assess the validity of its assumptions. With a few exceptions, the analysis supports the assumptions underlying the state model (see below); there is little evi-

dence of significant problems with the model formulation (although there may be other models that fit just as well).

Linearity Plots of standardized residuals against the four predictor variables in the state model–the proportion of child exemptions reported by families in poverty on tax returns, the proportion of people receiving food stamps, the proportion of people under age 65 who did not file a tax return, and a residual from the analogous regression equation using the previous census estimate as the dependent variable–support the assumption of linearity. Furthermore, the standardized residuals, when plotted against the model's predicted values, provide no evidence of the need for any transformation of the variables. This result helps justify the decision not to use the log transformation of the proportion poor as the dependent variable.

Constancy over Time Table 2-3 shows the regression coefficients for the predictor variables for the state model for each of the years from 1989 to1996, excluding 1994. The coefficients for all four poverty-rate predictor variables are positive in all 7 years and generally similar across all years. All of the coefficients are significant at the 5 percent level except that the coefficient of the proportion of people under age 65 who did not file a tax return (column 3) is not significant in 1989.

Inclusion or Exclusion of Predictor Variables The standardized residuals for the state regression model were grouped into four categories for each of the following characteristics: census region, population size in 1990, 1980 to 1990 population growth, percentage of black population in 1990, percentage of Hispanic population in 1990, percentage of group quarters residents in 1990, and percentage of poor school-age children in 1979 (from the 1980 census). The distributions of the standardized residuals for each category were then displayed using box plots. For none of these box plots is there an obvious pattern to the standardized residuals across categories, with one exception: in 1989, 1990, 1991, and 1993 the model underpredicts the proportion poor of school-age children in the West Region (i.e., the model estimates are lower than the CPS direct estimates for this group of states). The Census Bureau experimented with adding a West Region indicator predictor variable to the model. The coefficient of this variable has a negative sign for all 7 years; however, it is significant for only 1991, 1992, and 1993. For those 3 years, the model with the West Region variable performs better for states in the West Region. A further examination of the residuals from the state model without the West Region predictor variable for individual Western states reveals that the model fairly consistently underpredicts the proportion poor of school-age children in some Western states but just as consistently overpredicts the proportion poor of school-age children in other Western states. Further investigation is needed to explain these patterns.

TABLE 2-3 Estimates of Regression Coefficients for the 1995 State Model, Estimated for 1989-1993, and 1995-1996

	Predictor Variables[a]			
Year	(1)	(2)	(3)	(4)
1989	0.52	0.71	0.23	0.71
	(.09)	(.20)	(.13)	(.34)
1990	0.46	0.65	0.42	1.07
	(.09)	(.20)	(.15)	(.36)
1991	0.46	0.52	0.59	0.84
	(.10)	(.21)	(.14)	(.37)
1992	0.41	0.71	0.42	1.38
	(.10)	(.21)	(.13)	(.37)
1993	0.28	1.14	0.51	1.24
	(.12)	(.25)	(.14)	(.39)
1995	0.57	0.79	0.32	1.54
	(.12)	(.25)	(.13)	(.36)
1996	0.37	0.97	0.59	1.02
	(.12)	(.26)	(.14)	(.36)

NOTES: All predictor variables are in terms of rates. Standard errors of the estimated regression coefficients are in parentheses.

[a]Predictor variables: (1) ratio of child exemptions reported by families in poverty on tax returns to total child exemptions; (2) ratio of people receiving food stamps to total population; (3) ratio of people under age 65 who did not file an income tax return to total population under age 65; (4) residual from a regression of poverty rates for school-age children from the prior decennial census (1980 or 1990) on the other three predictor variables.

Normality, Homogeneous Variances, and Outliers The distribution of the standardized residuals from the state regression model shows some small degree of skewness, especially in the 1992 equation. However, the skewness does not appear sufficiently marked to be a problem. Also, the residual plots and the box plots of the distributions of the standardized residuals against the categories of states show little evidence of any heterogenous variance. Finally, there is no evidence of outliers from examination of the residual plots or displays of the distributions of the standardized residuals from the state regression model.

Model Error Variance One problem in the state model concerns the variance of the model error (u_i in equation (2)). In the state model, the variances of

the sampling errors (e_i in equation (2)) are estimated directly from the CPS data using a generalized variance function. The total model error variance is calculated using maximum likelihood estimation. The result of this calculation is an estimate of zero for the model error variance in the equation for every year except 1993. This result, which implies (absent sampling variability) that the model gives perfect predictions of state poverty rates for school-age children, is not credible. It produces zero weight for the direct estimates even when those estimates are quite precise, as is the case for several large states in the CPS sample. Even a small model error variance can substantially change the weight on the relatively high-precision direct estimates when they are combined in a shrinkage procedure with the model estimates.

To evaluate the effects of using zero model error variance in the estimation, the panel examined tables that compared the model estimates of the proportion poor of school-age children to the CPS direct estimates by state for 1989-1993 and 1995-1996; as an illustration, Table 2-4 shows this comparison for 1995. This examination demonstrated two important points. First, there are some appreciable differences between the model estimates and the direct estimates. For example, for Mississippi in 1995, the difference is over 7 percentage points. Therefore, if a non-zero estimate for model error variance is produced, it might have important consequences for the state estimates of poor school-age children. Second, while there are some appreciable differences, the model estimates were within two standard errors of the direct estimates for almost all states in each year. The range of model estimates that exceeded that limit in either a positive or negative direction was from one state in 1992 to six states in 1996. (Mississippi's difference in 1995 was not statistically significant at the 5 percent level.) For no single state did the model estimates exceed two standard errors of the direct estimates for more than 3 of the 7 years for which the state model was estimated. (And this analysis ignores the variance of the model estimates, which means that a yet smaller number of differences are statistically significant.) These results suggest that the state model is performing reasonably well: differences between model and direct estimates are neither unusually large nor strongly persistent. However, more work should be conducted to evaluate the current procedures for estimating the sampling error variance of the state model and the effects on the model estimates (see Chapter 5).

State Raking Factors

The final stage in producing updated estimates of the number of poor school-age children for counties is to rake the estimates from the county model for consistency with the estimates from the state model. The model-1990 census comparisons found that the raking procedure was beneficial to the county estimates. The raking factors vary considerably across states. For 1995, the raking

TABLE 2-4 CPS Direct Estimate and Regression Model Estimate of Percentage of School-Age Children in Poverty by State, 1995

State	CPS Direct Estimate (1)	Lower Confidence Bound on Direct Estimate (2)	Upper Confidence Bound on Direct Estimate (3)	State Model Regression Estimate (4)	Regression Estimate Minus Direct Estimate (4) – (1) (5)
Alabama	22.2	16.5	27.9	23.4	1.2
Alaska	6.3	1.6	11.1	10.9	4.5
Arizona	23.0	16.8	29.2	21.1	–1.9
Arkansas	21.4	14.0	28.7	24.0	2.6
California	22.5	19.4	25.7	21.5	–1.0
Colorado	9.4	5.1	13.8	11.8	2.3
Connecticut	15.6	7.3	24.0	12.6	–3.0
Delaware	15.6	8.3	23.0	12.8	–2.8
District of Columbia	30.2	17.9	42.4	33.8	3.7
Florida	21.1	16.8	25.4	20.7	–0.4
Georgia	14.8	8.2	21.3	21.4	6.7
Hawaii	14.1	7.9	20.3	11.9	–2.2
Idaho	15.4	9.9	20.9	12.7	–2.7
Illinois	19.4	14.6	24.2	15.7	–3.7
Indiana	12.9	9.0	16.8	12.6	–0.4
Iowa	15.2	8.9	21.4	11.2	–3.9
Kansas	10.6	4.8	16.4	12.7	2.1
Kentucky	18.9	13.4	24.4	22.9	4.0
Louisiana	24.2	15.6	32.9	28.0	3.8
Maine	10.7	4.1	17.4	13.8	3.1
Maryland	12.8	5.0	20.5	11.5	–1.3
Massachusetts	16.5	11.5	21.5	13.3	–3.2
Michigan	14.2	10.0	18.3	17.2	3.0
Minnesota	9.5	5.5	13.4	10.0	0.6
Mississippi	34.9	25.6	44.3	27.4	–7.6
Missouri	9.4	3.5	15.2	17.0	7.7
Montana	17.4	9.4	25.3	18.4	1.0
Nebraska	11.4	7.1	15.7	10.0	–1.4
Nevada	9.8	4.0	15.6	11.8	2.0
New Hampshire	4.2	0.6	7.8	6.5	2.3
New Jersey	9.3	6.5	12.0	12.3	3.0
New Mexico	34.0	27.8	40.3	28.6	–5.5
New York	22.7	19.1	26.3	23.1	0.4
North Carolina	19.7	13.8	25.5	17.1	–2.6
North Dakota	10.3	5.3	15.2	14.1	3.8
Ohio	16.6	11.1	22.2	15.1	–1.5
Oklahoma	22.6	13.1	32.1	22.5	–0.1
Oregon	12.5	7.1	17.9	12.4	–0.1
Pennsylvania	16.1	12.5	19.7	15.3	–0.9

continued

TABLE 2-4 Continued

State	CPS Direct Estimate (1)	Lower Confidence Bound on Direct Estimate (2)	Upper Confidence Bound on Direct Estimate (3)	State Model Regression Estimate (4)	Regression Estimate Minus Direct Estimate (4) – (1) (5)
Rhode Island	16.4	10.7	22.2	15.1	–1.3
South Carolina	30.8	21.9	39.7	21.9	–8.9
South Dakota	16.7	8.7	24.8	17.3	0.6
Tennessee	18.4	9.1	27.7	18.7	0.3
Texas	22.4	19.3	25.5	24.3	1.9
Utah	7.3	3.9	10.8	7.5	0.2
Vermont	11.3	3.2	19.4	11.6	0.3
Virginia	14.3	7.6	21.1	14.5	0.1
Washington	15.8	7.9	23.7	12.4	–3.4
West Virginia	23.0	13.2	32.9	25.7	2.7
Wisconsin	11.1	4.0	18.1	12.2	1.2
Wyoming	10.5	6.3	14.7	12.2	1.7

NOTE: Confidence bounds are plus or minus two standard errors on the direct estimate (95% confidence interval, obtained using direct estimates of the CPS standard errors).

SOURCE: Data from Bureau of the Census.

factors range from 0.71 to 1.14 (two-thirds fall between 0.88 and 1.06); for 1993, the raking factors range from 0.91 to 1.31 (two-thirds fall between 0.98 and 1.16).

The Census Bureau determined that the correlation between the raking factors for states in 1993 and 1995 is low, which implies that there is little systematic variation by state across these years. Also, some variation in the raking factors is expected given the form of the county model and the need to transform the predicted log values of poor school-age children to estimated numbers before the raking is performed. Nonetheless, the degree of variation in the raking factors suggests (though there are better ways to diagnose this) that there may be state effects not captured in the county model, which, in turn, could affect the behavior of the model in estimating the number of poor school-age children for counties within states. Preliminary work conducted by the panel suggests that such state effects may be present (see Chapter 5).

The panel urges the Census Bureau to estimate the variance of the state raking factors to determine if the variability that they exhibit for 1993 and 1995 is consistent with random error. If it is not, the panel urges the Census Bureau to further investigate the state raking factors, including consideration of whether there is any feature of the state model that might explain the variation. More generally, the Census Bureau should conduct research on how to account for state effects in the county model.

3

School District Estimates

For Title I fund allocations to be made in spring 1999 for the 1999-2000 school year, the Census Bureau was charged to produce updated estimates of the number of poor school-age children at the school district level. Three sets of school district estimates are required: (1) estimates of school-age children (aged 5-17) who were living in and related to a family in poverty in the preceding calendar year;[1] (2) estimates of all school-age children; and (3) estimates of the total population of the district. The first two sets of estimates are needed to implement the allocation formulas for both basic and concentration grants; the third set of estimates is needed to determine which school districts have fewer than 20,000 people.[2]

This chapter first considers estimates of poor school-age children for school districts. It reviews the difficulties that confront attempts to develop such estimates; describes the procedure that the Census Bureau used to develop district-level estimates of school-age children in July 1996 who were in poor families in 1995; and assesses the limited evaluations that are possible of these estimates. The chapter then describes the procedure and evaluations for estimates of the number of all school-age children and of the total population in July 1996 for school districts (see also Bureau of the Census, 1999, which describes the estima-

[1]See Chapter 1, footnote 5, for the definition of "related children."

[2]States, at their discretion, may aggregate the fund allocations for districts with less than 20,000 population and redistribute the funds by using another method that is approved by the Department of Education.

tion procedures and evaluations for the 1995 school district estimates). Finally, the chapter discusses the implications of the evaluations for the use of updated school district estimates for Title I allocations.

SCHOOL-AGE CHILDREN IN POVERTY

Issues in Estimating Poverty for School Districts

Developing estimates of the number of poor school-age children (or other characteristics) for school districts presents a number of difficult problems. These problems include the small population size of most districts and several other features of their boundaries and scope: school district boundaries in many instances cross county lines; they can and often do change over time; and some school districts cover specific grade levels, such as kindergarten-8 or 9-12. Because of these problems, there are no data sources now available for developing updated school district estimates of poor school-age children by using the type of model-based approach that was used for county estimates. These problems also compromise the quality of the estimates for school districts that are available by aggregating data for blocks from the decennial census. We briefly review each of these issues in turn.

Size

Table 3-1 shows the distribution of total school districts, school districts coterminous with counties, and total counties by population size from the 1990 census. Of 15,226 districts, 49 percent had fewer than 5,000 people, and fully 82 percent had fewer than 20,000 people, while only 9 percent had 40,000 or more people; the median population size was about 5,250. By comparison, of 3,141 counties, 10 percent had fewer than 5,000 people, and 32 percent had 40,000 or more people; the median population size was about 23,000. Small districts, while numerous, accounted for small proportions of school-age children: districts with fewer than 5,000 people included only 6 percent of all school-age children, and districts with fewer than 20,000 people included only 27 percent of all school-age children; in contrast, districts with more than 40,000 people included 58 percent of all school-age children. Such uses as Title I fund allocations, however, require estimates for all school districts, no matter how small. Yet it is not possible to obtain direct estimates for school districts from national surveys, such as the March CPS. Many school districts will have no sampled households in national surveys, and the estimates for all but the largest districts with sampled households will be very unreliable (i.e., exhibit high sampling variability). Even census data, as discussed below, are unreliable for many school districts.

TABLE 3-1 Percentage Distribution of School Districts, School Districts Coterminous with Counties, and Counties by Population Size, 1990 Census

	All School Districts		School Districts Coterminous with Counties		
Total Population	Districts (1)	School-Age Children (2)	Districts (3)	School-Age Children (4)	Counties (5)
Under 5,000	49.2	6.0	9.3	0.4	9.5
5,000-9,999	17.0	7.7	17.4	2.4	14.5
10,000-19,999	15.6	13.4	27.3	7.1	22.5
20,000-39,999	9.7	15.4	22.4	11.3	21.7
40,000 or more	8.5	57.6	23.7	78.8	31.7
Total (Number)	15,226	45.3 million	928	10.1 million	3,141

NOTE: School districts are defined as of 1989-1990.

SOURCE: Data from Bureau of the Census.

Boundaries

School district boundaries are, in general, determined by state regulations and practices. In seven states and the District of Columbia, school districts are coterminous with counties; these states included 370 districts in 1990 (2% of the total).[3] In another 17 states, school district boundaries coincide with other political units, such as townships. The boundaries of most, but not all, of the school districts in these states respect county lines. These states included 3,344 districts in 1990 (22% of the total), of which 190 crossed county lines. In the remaining 26 states, school district boundaries are unique to districts and often cross county lines. These states included 11,563 districts in 1990 (76% of the total), of which 3,931 crossed county lines. In all, 4,121 school districts (27% of the total) crossed county lines.

It is relatively easy to develop updated estimates of poor school-age children for districts that are coterminous with counties because county boundaries are generally stable over time, counties are relatively large areas, and data sources are available for counties (e.g., the data used to estimate the county model). Overall, in 1990, there were 928 districts that comprised an entire county or, in

[3]In some other states, some school districts are coterminous with counties; see below. Puerto Rico is treated as a single county and (coterminous) school district for purposes of Title I allocations.

the case of a few districts (e.g., New York City), more than one entire county. (The 928 districts include the districts in the seven states and the District of Columbia in which all school districts are counties together with selected districts in other states.) These districts accounted for 6 percent of all districts and 22 percent of all school-age children in 1990. Their median population size in 1990 was about 18,500 (Table 3-1, col. 3)—not far from the median population size for all counties (Table 3-1, col. 5).

Most of the remaining districts, whether or not they cross county lines, present more or less serious problems for updating: they are small, with a median population size of less than 5,000; their boundaries can and often do change; and few data are available for estimating poverty. These districts accounted for 94 percent of districts and 78 percent of all school-age children in 1990.

Grade Levels

In 1990, 11,284 school districts (74% of the total) served all grades—pre-kindergarten, kindergarten, or 1st grade through 12th grade. The remaining 3,942 districts (26% of the total) served a subset of grades, such as elementary grades, high school grades, or middle school grades. Developing updated estimates of poor school-age children for districts that serve specific grades is difficult because a method must be devised to allocate the limited available data on school-age poverty to the age range that is appropriate to the grade range of the school district.

Data Sources

The Census Bureau's county model can readily provide updated estimates of the number of poor school-age children for the small subset of school districts that comprise entire counties. However, as noted above, a model similar to the county model cannot be developed for the remaining 94 percent of school districts, principally because of the lack of administrative data with which to form the predictor variables in a regression model. For example, states do not generally geocode the addresses of Food Stamp Program participants to school districts, so there are no counts of food stamp participants for school districts. Similarly, a substantial proportion of addresses on federal income tax returns cannot be geocoded to census blocks, so it is not possible to estimate the number of poor children reported by families on tax returns for school districts. Finally, data from school districts on participation in the National School Lunch Program (requested from the states by the National Center for Education Statistics in its Common Core of Data Program) are far from complete, and they are of uncertain quality and applicability (see below, "School Lunch Data"). In the future, it may be possible to develop appropriate data sources for a model-based approach to

estimating poor school-age children for school districts (see Chapter 5), but such data are not now available.

Estimation Procedure

In the absence of data with which to develop a school district model similar to the county model, the Census Bureau used a simple synthetic approach to estimate poor school-age children by school districts for 1995. The approach involved seven steps:

(1) A survey was conducted to ascertain school district boundaries for the 1995-1996 school year.

(2) Each 1990 census block was assigned to a school district, as defined for 1995-1996.[4]

(3) The 1990 census data were aggregated for the blocks (or fractions of blocks) in each school district or part of a school district that lay wholly within a county.

(4) The 1990 census data for each school district or school district part were tabulated to form a ratio estimate of the number of poor school-age children: the ratio estimate was obtained by applying the proportion poor of school-age children from the census long-form sample data to the short-form complete-count estimate of all school-age children. The ratio estimate was used because it reduced somewhat the high sampling variability in the census estimates for school districts in comparison with estimates formed by simply inflating the long-form number of poor school-age children by the sampling weight.

(5) For the school districts or school district parts in a county, the share (proportion) for each school district or school district part of the 1990 census county total of poor school-age children was calculated from the ratio estimates. (For districts that are coterminous with a county, the share was 100%.)

(6) The 1990 census shares from step (5) were applied to the updated 1995 county estimates of poor school-age children produced by the county model (see Chapter 2) to obtain 1995 estimates of poor school-age children for school districts or school district parts.

(7) The 1995 school district estimates of poor school-age children were the estimates from step (6) for school districts wholly within a county and the sum of the estimates of school district parts for school districts that crossed county lines.

[4]When school district boundaries crossed census block boundaries, the poor school-age children in such a block were assigned to the appropriate school districts in proportion to the area of each district included in the block. When two or more school districts included a block because the districts covered selected grades (e.g., kindergarten-8 and 9-12), the poor children in the block in the relevant age ranges were assigned to the appropriate district on the basis of an analysis of the relationship of age to grade.

As an example of the procedure, take a county with 1,600 poor school-age children in 1989 (1990 census data) of whom 1,200 (75%) resided in school district A, 240 in school district B (15%), and 160 in school district C (10%). If the 1995 county model estimated that the county had only 1,200 poor school-age children, then the estimates of poor school-age children in 1995 for school districts A, B, and C are 900, 180, and 120, respectively. The estimation method assumes that all three school districts in the county experienced the same proportionate decrease in the number of poor school-age children—25 percent—as the county as a whole. If this assumption is incorrect (e.g., because the decrease in poverty in the county was concentrated in one of the districts, perhaps because of changes in the housing stock), then the estimates for the three school districts will be incorrect.

At present, 18 states use a similar procedure for allocating their Title I county funds to school districts, in that they make within-county allocations on the basis of 1990 census school district shares of poor school-age children, either solely or in combination with estimates of the other categories of formula-eligible children (e.g., foster children). Another nine states use 1990 census data together with other data sources, such as school lunch data, to allocate Title I county funds to school districts (according to the U.S. Department of Education).

The Census Bureau's 1995 school district estimates are not the only input to the Title I allocation formula. To make direct allocations to school districts for the 1999-2000 school year, the Department of Education must also obtain several other data elements for school districts, most of which have not been previously available at the district level: counts of the other categories of formula-eligible children (children in foster homes, in local institutions for neglected and delinquent children, and in families with income above the poverty line who receive welfare assistance);[5] and the dollar amounts of Title I allocations that school districts received for the 1998-1999 school year (to use in the hold-harmless computations). The Census Bureau's estimates of poor school-age children must also be adjusted to reflect school district boundary changes between 1995-1996 and 1998-1999 (although the department may leave it to the states to make appropriate adjustments).

Evaluations

To evaluate the Census Bureau's 1995 estimates of poor school-age children for school districts, the panel and the Census Bureau first assessed the 1990

[5]Poor school-age children as estimated by the Census Bureau were 96.2 percent of the total number of formula-eligible children counted in the Title I allocations for the 1998-1999 school year. Foster children, children in local institutions for neglected and delinquent children, and children in families with income above the poverty line receiving welfare assistance were 2.6 percent, 1.1 percent, and 0.1 percent, respectively, of the total number of formula-eligible children.

census estimates that are used to form school district shares of poor school-age children within counties. The 1990 census estimates are subject to high sampling variability, which is a problem for the Bureau's synthetic shares model. This high variability is also a problem for evaluations that use the 1990 census estimates as the standard of comparison.

Opportunities to evaluate the school district estimates are constrained by the limitations of available data. The panel and the Census Bureau used a 1980-1990 school district census file to evaluate a few variations of the Bureau's synthetic model for a subset of districts. The panel also evaluated the use of National School Lunch Program data as an alternative method for constructing updated school district estimates of poor school-age children in New York State.

Variability in Census Estimates

The two inputs to the Census Bureau's synthetic model for school district estimates of the number of poor school-age children are the county model estimates for the target year, which have been extensively evaluated (see Chapter 2), and the 1990 census estimates for determining school district shares, which are discussed in this section. The income data that are used to determine poverty status in the census are collected on the long-form questionnaire, which was administered to an average of about one-sixth of households in 1990. The long-form sample size is orders of magnitude larger than the sample size of such household surveys as the CPS, but for small areas, the long-form estimates can exhibit high sampling variability.

Table 3-2 shows the mean and median coefficient of variation (in percent) for the estimated number of poor school-age children from the 1990 census long-form sample, obtained as a simple inflation estimate, for school districts distributed into groups categorized by number of school-age children, with each group containing approximately the same number of districts. The mean coefficient of variation is 32 percent for all school districts, varying from 64 percent for districts in the smallest size category (1-185 students) to 14 percent for districts in the largest size category (3,770 or more students).[6] This degree of variability is high. For example, if a typical school district has about 200 poor school-age children, the long-form sample might give estimates anywhere from about 70 to about 330 poor school-age children. (This range is from 200 minus twice the coefficient of variation of 32% for the typical district to 200 plus twice that coefficient of variation.) By comparison, a common design goal for estimates that are published from a survey is a coefficient of variation of 10 percent or less.

Table 3-2 also shows the mean and median coefficient of variation for school district estimates of poor school-age children that were constructed by ratio esti-

[6]The districts in the largest size category have about 20,000 or more total population.

TABLE 3-2 Average Coefficients of Variation (C.V.) for Two Estimates of Number of Poor School-Age Children for School Districts by Number of School-Age Children, 1990 Census

Number of School-Age Children in District	Number of Districts	Estimate from Long-Form Census Sample (in percent)		Estimate Ratio-Adjusted from Long Form and Short Form (in percent)	
		Mean C.V.	Median C.V.	Mean C.V.	Median C.V.
Total	14,328	32	23	30	22
1 to 185	1,858	64	54	57	47
186 to 462	2,446	39	30	36	28
463 to 946	2,480	32	24	30	22
947 to 1,811	2,505	28	22	26	21
1,812 to 3,769	2,519	23	19	22	18
3,770 or more	2,520	14	11	13	11

NOTES: Excludes school districts for which the estimated number of poor school-age children is zero. School districts are defined as of 1988-1990. The coefficient of variation is the standard error of the estimate divided by the estimate.

SOURCE: Data from Bureau of the Census.

mation. In this approach, the proportion poor of school-age children is computed from the long-form sample data and that proportion is then applied to the estimated total number of school-age children from the short-form or complete-count census data, which are not subject to sampling variability. This procedure somewhat reduces the variability of the estimates: the mean coefficient of variation of the ratio-adjusted estimates is 30 percent, compared with 32 percent for the long-form estimates, a reduction of 7 percent.

The Census Bureau used the ratio-adjusted 1990 census estimates of poor school-age children to construct the 1995 school district estimates but, given time constraints, did not conduct research on ways to further reduce the variability of the census estimates. Such research should be a high priority. One possible approach is to use other short-form data (such as race and ethnicity, tenure, family type) as auxiliary information in the estimation of poor school-age children. Another approach is to smooth the 1990 census school district estimates

with the 1990 census county estimates, which would reduce the variability for smaller size districts (see Chapter 5).

Census Data Evaluations

The Census Bureau constructed a file of 1980 and 1990 census data for selected school districts, which was used to compare three sets of estimates of poor school-age children in 1989 with estimates from the 1990 census. In each instance, the 1980 census data that are used in the estimation are solely from the long form, while the 1990 census data are ratio adjusted. Three methods were used for the estimates:

(1) One synthetic method used county model estimates to construct school district estimates: method (1) applied the 1980 census shares of poor school-age children for school districts (or parts of school districts) within counties in 1979 to the Census Bureau's 1989 estimates of poor school-age children from its county model, with the county estimates controlled to the national estimate of poor school-age children in 1989 (from the 1990 census). This procedure is analogous to that used by the Census Bureau to produce the 1995 school district estimates from 1990 census shares applied to 1995 county model estimates, except that the 1980 census data are not ratio adjusted. (Also, the 1980 census estimates for 1979 are 10 years out of date for the 1989 estimates, while the 1990 census estimates for 1989 are 6 years out of date for the 1995 estimates.)

(2) A second synthetic method used 1990 census county estimates to construct school district estimates: method (2) applied the 1980 census shares of poor school-age children for school districts (or parts of school districts) within counties to the 1990 census county estimates of poor school-age children. This procedure eliminates the error in method (1) that is due to the county model.

(3) The third method was a national stable shares procedure: method (3) applied the 1980 census shares of poor school-age children for school districts within the nation as a whole to the national estimate of poor school-age children in 1989 from the 1990 census. This procedure assumes no change whatsoever in the relative shares of poor school-age children among school districts from the previous census, not even the change that occurs in methods (1) and (2) because of changes in the relative shares of poor school-age children among counties.

For several reasons, these comparisons provide only limited information with which to evaluate the Census Bureau's synthetic model for school district estimates. First, the alternative models are not very different from the Census Bureau's model. Second, the 1990 census estimates that are the standard of comparison are subject to high sampling variability even after ratio estimation. Finally, the evaluation file, of necessity, contains only a subset of school districts.

Scope of Evaluation File The 1980-1990 evaluation file was constructed from school district data sets that were prepared after each census. It was not possible to retabulate the individual block records from the 1980 census to match the 1990 census school district boundaries; instead, the goal was to identify a set of school districts in the data set for each year that could reasonably be assumed to have retained the same boundaries and grade ranges. The 1980 and 1990 census school district files were matched, using their identification numbers and other characteristics, and the following kinds of 1990 districts were dropped from the evaluation file:

- 928 districts or district parts for which the district or part was coterminous with a county and, hence, for which the county model would provide estimates;
- 4,108 districts that were not "unified," that is, that covered a limited grade range, such as Kindergarten-8 or 9-12;
- 416 districts that were newly formed and had no counterpart in 1980;
- 12 districts in counties that changed boundaries between 1980 and 1990; and
- 609 districts that crossed county lines and for which one or more of the county pieces in one year had no counterpart in the other year.

The resulting evaluation file contains 9,243 districts, which represent 61 percent of the 15,226 school districts that were included in the 1990 census school district file and 56 percent of school-age children. The subset of school districts in the evaluation file closely resembles the entire set of 1990 school districts in terms of the distribution of total population and total number of school-age children in 1990. For example, the subset of districts in the evaluation file includes 47 percent with fewer than 5,000 people and 8 percent with more than 40,000 people; the corresponding figures for the entire set of 1990 school districts are 49 percent and 9 percent, respectively.

A key assumption for using the evaluation file is that the 9,243 districts in the file, which had the same identification numbers in both 1980 and 1990, are the same districts and that their boundaries have not changed.[7] This assumption could be incorrect in some instances. For example, if a school district follows township boundaries and the township annexed land from another town between 1980 and 1990, it is likely that the school district identification number was the same in both 1980 and 1990 even though the boundaries changed.

[7]Another assumption for using the evaluation file is that school districts for which the boundaries did not change from 1980 to 1990 represent the behavior of districts for which the boundaries did change. To the extent that changes in boundaries are associated with changes in population, the synthetic shares approach may work less well for districts for which boundary changes occurred. However, these districts were less than 7 percent of the districts in 1990.

To investigate this assumption, the Census Bureau looked at unified school districts, not coterminous with counties, that had the same identification numbers in 1990 and in the 1995-1996 school district boundary survey. For 6 percent of these districts, which accounted for 2 percent of school-age children, the total number of school-age children originally tabulated in the 1990 census differed by 5 percent or more from the number retabulated according to the 1995-1996 boundaries. For the remaining 94 percent of districts, the two tabulations were exactly the same or differed by less than 5 percent, indicating that the same identification number is a reasonably good indicator of stability in school district boundaries.

Summary of Evaluation Results: Absolute Differences Table 3-3 provides summary statistics for the three sets of school district estimates of poor school-age children in 1989 in comparison with the 1990 census estimates. The statistics provided are the average absolute difference between the estimates from a model or method and the census, as a percentage of the average number of poor school-age children in the census, and the average proportional absolute difference between each set of estimates and the 1990 census estimates. For comparison purposes, the last row of the table provides the same statistics for county estimates of poor school-age children in 1989 from the Census Bureau's county model.

The first measure in Table 3-3 assesses the absolute difference between estimates from a method and the 1990 census in terms of numbers of poor children, while the second measure assesses the absolute difference in terms of proportional errors for school districts. From a national perspective, it can be argued that the absolute differences in terms of numbers are more important for effective Title I allocations because, with direct allocation, Title I funds are primarily distributed in proportion to the number of children in a school district. Therefore, the amount of funds that are misallocated depends primarily on the number of children rather than on the percentages by district. For example, an error of 5 percent in the number of school-age children in poverty in a large district could correspond to many thousands of children and have more impact on the allocation of funds than errors of 5 percent (or greater) in several smaller districts. However, from the district perspective, the proportional error for a district's allocation is also important.

Ideally, a method will perform well on both types of measures, but, as discussed below, all three synthetic shares methods perform much worse on the average proportional absolute difference measure overall than on the average absolute difference measure. The reason for this consistent finding is that there are many small school districts that tend to have much larger-than-average proportional errors, which are reflected in the average proportional absolute difference measure. However, the much larger proportional errors for small districts do not represent many poor school-age children and so do not contribute as much to the absolute difference measure.

TABLE 3-3 Comparison of Synthetic Estimates and 1990 Census School District Estimates of the Number of Poor School-Age Children in 1989

Model	Average Absolute Difference, Relative to Average Poor School-Age Children (in percent)[a]	Average Proportional Absolute Difference (in percent)[b]
1989 School District Estimates		
(1) Synthetic method using 1980 census shares applied to 1989 county model estimates	22.2	60.0
(2) Synthetic method using 1980 census shares applied to 1990 census county estimates	18.0	55.4
(3) National stable shares method using 1980 census shares applied to 1990 census national estimate	28.7	71.7
1989 County Estimates from Census Bureau's County Model	10.7	16.4

NOTES: School district estimates are based on 8,810 districts (9,243 districts in the 1980-1990 evaluation file minus 66 districts with estimated sample population of 30 or less in 1980 or 1990 and an additional 367 school districts with estimates of no children in poverty). The 1990 census estimates used in the comparisons are the ratio-adjusted estimates (see text). All three sets of school district estimates are controlled to the 1990 census national estimate of poor school-age children in 1989 before comparison with the 1990 census school district estimates.

[a]The formula, where there are n school districts or counties (i), and Y is the estimated number of poor school-age children from a model or the census, is

$$\Sigma[(\,|Y_{\text{model } i} - Y_{\text{census } i}|\,) / n] \,/\, [\,\Sigma(\,Y_{\text{census } i}\,) / n] \,.$$

[b]The formula is $\Sigma\,[(\,|Y_{\text{model } i} - Y_{\text{census } i}|\,) / Y_{\text{census } i}\,] / n$.

SOURCE: Data from Bureau of the Census.

As seen in the last row of Table 3-3, the average absolute difference of the county model estimates from the 1990 census county estimates is 10.7 percent of the 1990 census county average number of poor school-age children; the average proportional absolute difference is 16.4 percent. The school district estimates show much larger differences. The average absolute difference for the Census Bureau's synthetic method (1), which applies 1980 census school district shares of poor school-age children within counties to the county model estimates for 1989, is 22.2 percent of the 1990 census school district average number of poor school-age children (2.1 times the corresponding figure for the county model estimates); the average proportional difference is 60 percent (3.7 times the corresponding figure for the county model estimates).

Method (1) reduces the average absolute difference measure by 23 percent (22.2/28.7) and the average proportional absolute difference measure by 16 percent (60.0/71.7) compared with the national stable shares method (3), which assumes no change in school district shares of all poor school-age children in the nation between the 1980 and 1990 censuses. Method (2), which applies 1980 census school district shares within counties to the 1990 census county estimates of poor school-age children, performs somewhat better: it reduces the average absolute difference measure by 37 percent (18.0/28.7) and the average proportional absolute difference measure by 23 percent (55.4/71.7) when compared with the national stable shares method (3). However, method (2) is of theoretical interest only. In a noncensus year, such as 1995, model-based county estimates have to be used for adjusting school district shares from the census, and there will be errors in these estimates.

The Census Bureau also explored a fourth method in which a set of estimates was constructed by applying the 1980 census shares of poor school-age children for school districts within each *state* to the 1990 census state estimates of poor school-age children. This method produced average absolute and average proportional absolute differences between those of methods (2) and (3). It also is of theoretical interest only because it cannot be used in a noncensus year. However, it illustrates that using state estimates to control school district shares (which could be done with the Census Bureau's state model estimates) is better than using a single national control, but worse than using county controls.

There are several reasons for the large differences between the synthetic estimates of poor school-age children for districts produced by method (1) and the comparison ratio-adjusted estimates from the 1990 census: the sampling variability in the 1980 census estimates of school district shares, which is high for many districts; the inability of the synthetic shares method to capture within-county changes in school district shares of poor school-age children from the 1980 census to the 1990 census; the errors in the county model itself (although these are not a large component); and the sampling variability that remains in the 1990 census comparison estimates even after ratio estimation. Because of the sizable sampling variability in the 1990 census estimates, the difference measures in Table 3-3 are overestimates of the differences from the true numbers of poor school-age children in 1989. It would be useful to remove this effect, and that should be done as part of future research (see Chapter 5).

Considering school districts by size, method (1) performs reasonably well on both the average absolute difference measure and the average proportional absolute difference measure for districts with 40,000 or more people in 1990 (data not shown). For these districts, the estimates are not markedly worse than the county estimates. Districts with 40,000 or more people are only 8 percent of the total number of school districts in the 1980-1990 evaluation file, but they contain 55 percent of the poor school-age children in the file.

Method (1) performs less well for school districts with 10,000 to 39,999 people in the 1990 census and performs very poorly for districts with fewer than 5,000 people in the 1990 census. Thus, while the average absolute difference measure for districts with 40,000 or more people in 1990 is 17 percent, it is 24 percent for districts with 20,000 to 39,999 people, 26 percent for districts with 10,000 to 19,999 people, 30 percent for districts with 5,000-9,999 people, and 43 percent for districts with 5,000 or fewer people. Districts with 5,000 or fewer people in 1990 contain only 8 percent of the poor school-age children in the 1980-1990 evaluation file, but they are 46 percent of total districts.

The much larger differences between the estimates from method (1) and the 1990 census estimates for smaller school districts relative to larger districts are due in part to the greater sampling variability in the 1990 census estimates for smaller districts. As noted above, the panel believes there are ways to further reduce the variability in the 1990 census estimates beyond the reduction achieved by using simple ratio estimates instead of simple inflation estimates. A reduction in the variability of the 1990 census estimates would permit not only a more accurate assessment of the synthetic shares approach, but also an improvement in the 1995 school district estimates that are formed by applying 1990 census within-county school district shares to the 1995 estimates from the county model.

Summary of Evaluation Results: Algebraic Differences The evaluation also examined the algebraic differences by category of school district. The following categories were used: geographic division, 1980 population, 1990 population, 1980-1990 population growth, percentage of poor school-age children in 1980, percentage of poor school-age children in 1990, change in the poverty rate for school-age children from 1980 to 1990, percentage of Hispanic population in 1980, percentage of black population in 1980, and percentage of group quarters residents in 1980. The results are summarized below for method (1); detailed results for all three methods are provided by Bureau of the Census (1999).

The category algebraic difference is the sum, for all school districts in a category, of the algebraic (signed) difference between the model estimate of poor school-age children and the 1990 census estimate for each district, divided by the sum of the census estimates for all districts in the category. This measure expresses model-census differences in terms of the numbers of poor children, similar to the overall absolute difference in the first column of Table 3-3. However, the category algebraic difference is expressed as an algebraic measure in which positive differences (overpredictions) within a category offset negative differences (underpredictions). The measure is intended to identify instances of potential bias in the predictions from a model or method. For example, the method may over(under)predict, on average, the number of poor school-age children in larger school districts relative to smaller districts.

The comparison of category algebraic differences for estimates from the

Census Bureau's synthetic method (1) with 1990 census estimates found no strong patterns of over(under)prediction for school districts categorized by percentage of black, percentage of Hispanic, or percentage of group quarters residents in 1980. However, method (1) did somewhat overpredict the number of poor school-age children in districts with no black or Hispanic residents or a very small proportion of group quarters residents in 1980 relative to other districts. Method (1) also somewhat overpredicted the number of poor school-age children in districts with fewer than 5,000 people in 1980 and 1990 relative to other districts. These findings may be related, in that districts with no black or Hispanic residents or very few group quarters residents are also districts that have very small populations.

For school districts categorized by population growth from 1980 to 1990, method (1) overpredicted the number of poor school-age children in districts that experienced a decline in population of more than 10 percent and underpredicted the number of poor school-age children in districts that experienced an increase in population of more than 10 percent. The same pattern was even greater for districts categorized by change in the poverty rate for school-age children from 1980 to 1990. These findings are not unexpected in that the synthetic shares method, by definition, will not reflect large increases or decreases in population or poverty for school districts except to the extent that the district increase or decrease parallels that of the county in which it is located.

For school districts categorized by percentage of poor school-age children, method (1) underpredicted the number of poor school-age children in districts that had a lower school-age poverty rate in 1980 relative to districts with a higher rate. In contrast, method (1) overpredicted the number of poor school-age children in districts that had a lower school-age poverty rate in 1990 relative to districts with a higher rate. These findings are also not unexpected. They are evidence of the so-called "regression to the mean" phenomenon, in which, due to sampling variability, school districts that have low estimates of school-age poverty rates in one year will tend to have higher rates in another year (other things being equal) and vice versa.

Finally, for school districts categorized by census geographic division, method (1) overpredicted the number of poor school-age children in districts in the Pacific Division and, to a lesser extent, in the Mountain Division relative to districts in other divisions. This finding is consistent with a similar finding for the 1989 county model estimates, which, in turn, was attributed to the state model.

School Lunch Data

As noted at the beginning of the chapter, there is a lack of administrative data with which to estimate school-age poverty for school districts. Food stamp data are not generally available for districts, and federal tax return data at present

cannot be reliably coded to school districts in many areas. Another possible source of information on poverty for school districts is data from the National School Lunch Program, which provides free and reduced-price meals to qualifying children.

The Census Bureau decided that it could not use school lunch data in developing updated estimates of poor school-age children for school districts for two major reasons. First, there is no complete and accurate set of school lunch data for all school districts. The National Center for Education Statistics (NCES) obtains school lunch counts as part of its Common Core of Data (CCD) system, in which state educational agencies report a large number of data items for public school systems.[8] The school lunch data are not published and have not been a priority of NCES. The center does not follow up with states when there is no information provided for a school district or to evaluate the accuracy of the reports. Hence, the quality of the data is not established, and they are far from complete.

Files of school lunch data for 1990-1995 that NCES provided to the panel contain large numbers of missing and zero values. In some cases, missing data may be due to the fact that a school district no longer exists (e.g., it may have been combined with another district); however, most instances of missing data appear to be due to nonreporting by school districts. Zero values may be valid in many instances, but NCES staff indicated that missing data are sometimes reported as zero, and analysis supported this assessment. Also, while states are asked to report counts of participants in the free school lunch program, it appears that many states report the combined total for the free and reduced-price programs, which have different income eligibility limits.

Only 18 states have reports that are more than 90 percent complete (fewer than 10% of school districts with missing or zero values) in all 6 years of the NCES files. At the other extreme, 10 states have reports that are less than 50 percent complete in all 6 years; most of these states do not report school lunch data at all. Clearly, if school lunch data are to be used to estimate the number of poor school-age children, it will be necessary to make school lunch reporting a priority in the CCD system for follow-up and evaluation.

The second reason for the Census Bureau not to use school lunch data in developing a consistent set of school district estimates nationwide is that counts of participants in the National School Lunch Program differ from poor school-

[8]NCES is the only federal agency that attempts to obtain school lunch data for school districts. The Department of Agriculture obtains aggregate counts each October at the state level of the number of children approved for free lunch and reduced-price lunch in both public and participating private schools. In addition, each month the department obtains aggregate counts at the state level of meals served for purposes of reimbursing the states for meal costs (the subsidy varies by whether the meal was free, reduced price, or full price).

age children in at least four respects, and the differences are probably not the same across jurisdictions:

- The eligibility standard to qualify for free lunches is family income that is less than 130 percent of the poverty threshold, which means that school lunch program participants include near-poor as well as poor children. (Children in families with incomes as high as 185% of poverty can receive reduced-price lunches.)
- Participation in the school lunch program is voluntary and may be affected by such factors as perceived stigma (it is believed that high school students are less likely to participate than elementary school students for this reason) and the extent of outreach by school officials to encourage families to sign up for the program.
- Not all private schools participate in the program, although most do.
- School lunch program participants include children enrolled in participating schools in the district, whereas the Census Bureau is charged to produce estimates of poor school-age children who reside in the district. The two populations differ to the extent that poor resident children attend nonparticipating private schools or schools outside their district (nonresident poor children may also attend schools in the district).

If the differences between school lunch participants and poor school-age children are inconsistent across jurisdictions, it will not be possible to develop a uniform and equitable estimation procedure for school districts by using school lunch data. Having a uniform procedure for estimates that are produced by the Census Bureau for use in direct allocations of Title I funds is important for at least two reasons. First, there are substantial practical difficulties for the Census Bureau to evaluate and develop different estimation procedures for different sets of school districts, even when it might be possible to improve the accuracy of the estimates in some cases. Second, if the use of different estimation procedures produces estimates of varying quality across school districts, there could be a problem of equity for concentration grants because, under direct allocations, the concentration grant allocations to one area can affect the allocations to other areas. Such effects cannot occur under the current two-stage allocation process, in which states that use school lunch data (or another data source) to allocate concentration grant funds to school districts are constrained by the county allocations determined by the Department of Education.

Yet school lunch participation is an indicator of low income, and if school lunch data were available and determined to be reasonably consistent across jurisdictions, the Census Bureau could consider using such data to modify its current estimation process. For example, it could follow the practice of the states that currently use counts of school lunch participants, solely or together or with census data, to distribute the Department of Education's Title I allocations for

counties to school districts. At present, eight states use free school lunch data as the only factor in their subcounty allocation formula, three states use free and reduced-price school lunch data as the only factor, and six states use free or free and reduced-price school lunch data together with 1990 census data to make subcounty allocations. In effect, these 17 states use a shares approach for school district estimates that is similar to the Census Bureau's method, except that the district shares within counties are computed on the basis of contemporaneous counts of school lunch participants instead of 1990 census estimates of poor school-age children.

The panel undertook a limited evaluation of a school lunch-based shares approach in one state—New York—for which it was able to obtain complete free and reduced-price school lunch data for almost all public schools for 1989-1990 and assign them to school districts and counties.[9] There are 623 New York State school districts in the 1980-1990 evaluation file, or 7 percent of the total number of districts in the file. The New York State districts in the evaluation file are somewhat larger than average, with a median population size in 1990 of about 9,000 compared with a median population size of about 5,250 for all districts.

The analysis compared three sets of estimates of poor school-age children in 1989 for school districts in New York State with estimates from the 1990 census. The methods used to develop the three sets differ only in the estimation of within-county school district shares: the Census Bureau's synthetic method (2), in which 1980 census within-county school district shares of poor school-age children were applied to 1989 county estimates from the 1990 census; a synthetic method in which 1989-1990 within-county school district shares of free lunch program participants were applied to 1989 county estimates from the 1990 census; and a synthetic method in which 1989-1990 within-county school district shares of combined free and reduced-price lunch program participants were applied to 1989 county estimates from the 1990 census.

Table 3-4 provides summary statistics for the three sets of school district estimates of poor school-age children in 1989 for New York State compared with the 1990 census estimates for these districts. The table includes the average absolute difference between the estimates from a method and the census, expressed as a percent of the average number of poor school-age children in the census, and the average proportional absolute difference between each set of estimates and the 1990 census estimates. For comparison purposes, the last row of the table provides the same statistics for estimates of poor school-age children for all U.S. school districts in the evaluation file from method (2), which applies 1980 census within-county school district shares to 1990 census county estimates.

[9]This evaluation was carried out at the State University of New York–Albany by Dr. James Wyckoff, a member of the panel, assisted by Frank Papa; see the appendix to this report, which includes overall and category comparisons.

TABLE 3-4 Comparison of Synthetic Estimates and 1990 Census School District Estimates of the Number of Poor School-Age Children in 1989, New York State

Model	Average Absolute Difference, Relative to Average Poor School-Age Children (in percent)[a]	Average Proportional Absolute Difference (in percent)[b]
New York State School District Estimates (N = 623)		
Synthetic method (2) using 1980 census shares applied to 1990 census county estimates	23.9	53.4
Synthetic method using 1989-1990 free lunch participants applied to 1990 census county estimates	22.3	48.7
Synthetic method using 1989-1990 free and reduced-price lunch participants applied to 1990 census county estimates	24.2	52.1
U.S. School District Estimates (N = 8,810) from synthetic method (2) using 1980 census shares applied to 1990 census county estimates	18.0	55.4

[a]The formula, where there are n school districts (i), and Y is the estimated number of poor school-age children from a model or the census, is

$$\Sigma[(|Y_{\text{model } i} - Y_{\text{census } i}|) / n] / [\Sigma(Y_{\text{census } i}) / n] .$$

[b]The formula is $\Sigma [(|Y_{\text{model } i} - Y_{\text{census } i}|) / Y_{\text{census } i}] / n .$

SOURCE: Wyckoff and Papa (in appendix); see also Table 3-3.

The average absolute difference of the estimates for all school districts from the 1990 census estimates using method (2) is 18 percent; the average proportional absolute difference is 55 percent. The corresponding figures for estimates for New York State school districts only are 24 percent and 53 percent, respectively, for a method analogous to method (2); 22 percent and 49 percent, respectively, for a method based on free lunch participants; and 24 percent and 52 percent, respectively, for a method based on free and reduced-price lunch participants.

The absolute differences in all three methods of estimating poor school-age children in 1989 for New York State school districts are similar and large in

magnitude. Even though the school lunch data pertain to the same year as the 1990 census comparison estimates, neither set of school lunch-based synthetic estimates is much more accurate than the 1980 census-based synthetic estimates. However, looking at both absolute differences and category algebraic differences, the use of free lunch participants as the basis for estimates is marginally more accurate than the other two methods that were evaluated. This finding suggests that it could be worthwhile to conduct a similar analysis for other states to determine if there is enough consistency across jurisdictions in the relationship of school lunch program data to school-age poverty to warrant further consideration of the use of school lunch data for school district estimates. If these data were to be used, a major effort would be needed to improve the reporting of the data to NCES for use by the Census Bureau for estimation purposes.

POPULATION TOTALS

Estimation Procedures

The Census Bureau was charged to produce estimates at the school district level not only of poor school-age children in 1995, but also of the total population and total number of school-age children as of July 1996. Estimates of total school-age children are needed so that the Department of Education can compute poverty rates for school districts, which are a factor in the Title I allocation formulas and the hold-harmless provisions. Estimates of total population are needed so that a state knows which districts have fewer than 20,000 people if it wants to take advantage of the provision in the legislation that permits states to aggregate the Title I allocations for these districts and to redistribute the funds on some other approved basis.

The procedures used by the Census Bureau to estimate total population and total school-age children for school districts are similar to those used for estimating poor school-age children. The method for producing 1996 estimates of total population and total school-age children for districts was to retabulate the 1990 census data according to 1995-1996 school district boundaries, determine the 1990 census county share in each district or part of a district for total population and total school-age children, and apply those shares to the Census Bureau's 1996 county estimates of total population and total school-age children, respectively. The 1990 census school district shares are based on data from the complete count (short form) and are not subject to sampling error. The 1996 county estimates are derived from the Bureau's demographic estimates program, which uses administrative records on births, deaths, and migration to update the previous census (see National Research Council, 1998:App.B).

Evaluations

As it did for the estimates of poor school-age children, the Census Bureau evaluated its method for estimating total population and total school-age children at the district level by using the 1980-1990 evaluation file to compare three sets of 1990 estimates with 1990 census numbers. The three sets of estimates were derived by three methods: (1) applying 1980 census school district shares within counties to 1990 demographically derived county estimates; (2) applying 1980 census school district shares within counties to 1990 census county numbers; and (3) applying 1980 census school district shares within the nation as a whole to the national 1990 census number.

Tables 3-5 and 3-6 provide summary statistics for the three sets of school district estimates of 1990 total population and 1990 total school-age children, respectively, compared with the 1990 census numbers. The statistics provided are the average absolute difference between the estimates from a method and the census expressed as a percent of the average total population or total school-age children in the census, and the average proportional absolute difference between each set of estimates and the 1990 census numbers. For comparison purposes, the last row of each table provides the same statistics for county estimates of total population and total school-age children in 1990 from the Census Bureau's demographic estimates program. (As noted above, this program uses administrative records, such as births and deaths, to update population numbers from the previous census.)

The county demographic estimates of total population and total school-age children for 1990 differ little from the 1990 census numbers: the average absolute differences are 2 percent and 5 percent, respectively (Tables 3-5 and 3-6, first column); the average proportional absolute differences are 4 percent and 6 percent, respectively. The school district estimates show larger differences, although the differences are much smaller than those for school district estimates of poor school-age children (see Table 3-3). For school district estimates of total population under method (1), the average absolute difference is 10 percent of the average total population; for school district estimates of total school-age children under method (1), the average absolute difference is 12 percent of the average total school-age children. By comparison, for school district estimates of poor school-age children under method (1), the average absolute difference is 22 percent of the average number of poor school-age children. The corresponding average proportional absolute differences are 13 percent (total population), 17 percent (total school-age children), and 60 percent (poor school-age children).

Evaluations of Census Bureau population estimates for states and counties have shown that the proportional differences of the estimates in comparison with census numbers are larger on average for small areas than for large ones. The proportional differences of the estimates also tend to be larger for areas in which the population is changing rapidly than for areas that are more stable (see Na-

TABLE 3-5 Comparison of Synthetic Estimates and 1990 Census School District Numbers of Total Population in 1990

Model	Average Absolute Difference, Relative to Average Total Population (in percent)[a]	Average Proportional Absolute Difference (in percent)[b]
1990 School District Estimates		
(1) Synthetic method using 1980 census shares applied to 1990 county model estimates	9.6	13.3
(2) Synthetic method using 1980 census shares applied to 1990 census county numbers	9.2	12.6
(3) National stable shares method using 1980 census shares applied to 1990 census national number	13.9	18.9
1990 County Estimates from Census Bureau's Demographic Estimates Program	2.3	3.6

NOTES: School district estimates are based on 9,201 districts (9,243 districts in the 1980-1990 evaluation file minus 42 districts with estimated population 30 or less in 1980 or 1990). The 1990 census numbers used in the comparisons are from the complete count and are not subject to sampling error. The estimates from the three methods are controlled to the 1990 census national total population number before comparison to the 1990 census school district estimates.

[a]The formula, where there are n school districts or counties (i), and Y is the estimate (number) for the total population from a model (census), is

$$\Sigma[(\,|Y_{\text{model } i} - Y_{\text{census } i}\,|) / n] \,/\, [\,\Sigma(\,Y_{\text{census } i}\,) / n] \,.$$

[b]The formula is $\Sigma\,[(\,|Y_{\text{model } i} - Y_{\text{census } i}\,|) / Y_{\text{census } i}\,] / n$.

SOURCE: Data from Bureau of the Census; see also National Research Council (1998:75).

tional Research Council, 1998:75). The school district estimates of total population and total school-age children follow the same patterns.

Compared with the school district estimates of poor school-age children, the estimates of total population and total school-age children benefit from two factors. First, total population and total school-age children are larger quantities to estimate. Second, the census data that are used to form within-county school district shares of total population and total school-age children, while subject to measurement error, are not from a sample. Nonetheless, the estimates of total population and total school-age children for school districts are not nearly as accurate as the corresponding county estimates. The Census Bureau has begun, but has not had time to complete, an analysis of school enrollment data to deter-

TABLE 3-6 Comparison of Synthetic Estimates and 1990 Census School District Numbers of Total School-Age Children in 1990

Model	Average Absolute Difference, Relative to Average Total School-Age Children (in percent)[a]	Average Proportional Absolute Difference (in percent)[b]
1990 School District Estimates		
(1) Synthetic method using 1980 census shares applied to 1990 county model estimates	12.0	16.9
(2) Synthetic method using 1980 census shares applied to 1990 census county numbers	10.4	16.1
(3) National stable shares method using 1980 census shares applied to 1990 census national number	16.6	20.6
1990 County Estimates from Census Bureau's Demographic Estimates Program	4.9	6.3

NOTES: School district estimates are based on 9,201 districts (9,243 districts in the 1980-1990 evaluation file minus 42 districts with estimated population 30 or less in 1980 or 1990). The 1990 census numbers used in the comparisons are from the complete count and are not subject to sampling error. The estimates from the three methods are controlled to the 1990 census national number of total school-age children before comparison to the 1990 census school district estimates.

[a]The formula, where there are n school districts or counties (i), and Y is the estimate (number) of total school-age children from a model (census), is

$$\Sigma[(\,|Y_{\text{model } i} - Y_{\text{census } i}|\,) / n] \,/\, [\,\Sigma(\,Y_{\text{census } i}\,) / n] \,.$$

[b]The formula is $\Sigma\,[(\,|Y_{\text{model } i} - Y_{\text{census } i}|\,) / Y_{\text{census } i}\,] / n$.

SOURCE: Data from Bureau of the Census; see also National Research Council (1998:75).

mine if these data could be used to improve the school district estimates of total school-age children. Such work should be continued (see Chapter 5).

ASSESSMENT

It is difficult to draw firm conclusions from the evaluations of the Census Bureau's updated school district estimates of poor school-age children regarding their use for Title I allocations. On the positive side, the estimates are reasonably good for two groups of districts that contain many poor school-age children: districts that are coterminous with a county or more than one county, for which the county model provides estimates, and other districts with a total population of

40,000 or more, for which the Census Bureau's synthetic shares method produces estimates that are only somewhat less reliable than the county model estimates.[10] These two groups together (adjusting for the overlap between them) comprise only a small fraction of districts, 13 percent of the total as of 1990, but they contain a large fraction of poor school-age children, 62 percent of the total. On the negative side, the school district estimates are subject to high sampling variability for the remaining 87 percent of districts, which contain 38 percent of poor school-age children.

In terms of the mandate to the panel, the estimates might be judged to be "inappropriate or unreliable" for direct allocations of Title I funds to school districts. However, such a conclusion implies a definition of "inappropriate or unreliable" that does not take into account the allocation procedures that might otherwise be used. Given that some set of estimates must be used to make Title I allocations, the panel believes that "inappropriate or unreliable" should be defined in a relative sense. Applying a relative definition, one can argue that, in the context of currently available information, a direct allocation procedure that uses the Census Bureau's school district estimates is at least as good as and perhaps preferable to the alternative, which is for the states to continue to distribute the county allocations from the Department of Education to school districts by using a variety of data sources.

For suballocations of Title I funds, the states at present use several types of data:

- Seven states and the District of Columbia make no suballocations to districts because their school districts are coterminous with counties (three of these states make suballocations to a few districts in their states that are not coterminous with counties, such as a city that is a separate district from the remainder of the county).
- Eight states use 1990 census data alone.
- Ten states use 1990 census data and estimates of the other categories of formula-eligible children, such as foster children.
- Nine states use a combination of 1990 census data together with free lunch, or free and reduced-price lunch, or AFDC, or a composite of AFDC, food stamps, and Medicaid data.
- Eight states use free lunch data only.
- Three states use free and reduced-price lunch data.
- One state uses free lunch and state tax information.

[10]The 40,000 population size cutoff should be viewed as approximate. Examination of the evaluation results for a more detailed set of population size categories for school districts than discussed in the text indicated that the method (1) estimates for school districts approach the reliability of the county estimates somewhere in the range of about 30,000 to 50,000 population.

- Three states use AFDC data only or in combination with foster child data.
- One state uses food stamp data.

Most states are constrained to distribute the county allocations to school districts (or parts of school districts) within each county. However, the Department of Education permits nine states to make direct allocations of basic grants to school districts without regard to the county allocation amounts because so many of their school districts cross county boundaries.[11] Of these nine states, one uses 1990 census data to make direct allocations of basic grants; five use 1990 census data and estimates of the other categories of formula-eligible children; one uses a combination of 1990 census and free and reduced-price lunch data; one uses free lunch data; and one uses free and reduced-price lunch data.

The 18 states that rely on 1990 census data (either alone or together with estimates of the other categories of formula-eligible children) to distribute the county allocations to school districts could readily make use of the Census Bureau's school district estimates. In fact, the Bureau's census shares-based estimates are likely to be somewhat more accurate than the corresponding estimates that the states have been producing because the Bureau has access to 1990 census block data and so can more accurately retabulate the census data to reflect changes in school district boundaries; the states have had access only to public use census files for 1989-1990 school district boundaries.[12] In addition, the ratio-adjustment procedure employed by the Census Bureau to estimate census shares somewhat reduces their sampling variability. For the six states in this group that use 1990 census data to make direct allocations of basic grants to school districts without regard to the county allocation amounts, the use of the Bureau's census shares-based estimates would have the advantage of reflecting the updated county estimates from the Bureau's county model.

Twenty-five states currently use data sources other than the census, or in combination with the census, to suballocate county Title I funds to school districts. (Three of these states make direct allocations of basic grants to districts.) It was not possible to evaluate the accuracy of such sources as school lunch data across states. The analysis that was conducted for New York (see above) suggests that there are only marginal gains in accuracy from use of school lunch data. Moreover, it is not likely that the use of a shares approach based on school lunch data would produce results that are as consistent across states as the use of a shares approach based on census data: in some states, school lunch shares might

[11]No state is currently permitted to make direct allocations of concentration grants; see discussion below.

[12]The Census Bureau has provided the Department of Education with a file of 1990 census data for school districts defined according to 1995-1996 boundaries, to which the states can have access.

be better than census shares; in other states, they might be worse. This inconsistency could be a problem for direct allocation of concentration grants.

Overall, the panel finds four reasons to support use of the Census Bureau's school district estimates of poor school-age children for direct allocation of Title I allocation funds: the congressional mandate for direct allocations; the use of a uniform procedure to derive the Census Bureau's estimates; the somewhat greater accuracy of the Census Bureau's estimates of 1990 census shares compared with what the states can likely produce; and the absence of strong evidence that there are other, better data sources available for estimation. For the rest of our assessment, we consider more carefully the features of the basic grant and concentration grant allocation formulas and how they may interact with the provision in the 1994 legislation that states may redistribute the aggregate allocations for districts with fewer than 20,000 people by some other method that the Department of Education approves.

Basic Grants

Under the current two-stage allocation process, basic grants are allocated to school districts essentially as shares of the county total amounts. Whatever the data source used by a state to form the within-county shares (e.g., census data, school lunch data, combination of two or more data sources), the county totals remain as specified by the Department of Education. The exception, as noted above, is that the department currently allows nine states in which school district boundaries bear little correspondence to county boundaries to redistribute the total basic grant allocation for the state without regard to the county allocations. For other states, the county totals, which, in turn, reflect (approximately) the Census Bureau's updated estimates from its county model, are maintained.[13]

Direct allocation of basic grants to school districts by using the Census Bureau's synthetic shares estimates would have the same property of essentially respecting the county totals because the Census Bureau's estimation procedure controls the school district estimates to county estimates derived separately from its county model. The correspondence between the county totals from the two-stage allocation process and those from the sums of direct allocations to the districts in each county will not be exact for several reasons. One, the hold-harmless provisions applied at the county level will give a somewhat different result from applying the hold-harmless provisions to districts and aggregating the resulting amounts to counties. Also, in contrast to counties, a proportion of school districts (12% in 1995-1996, the most recent year for which the Depart-

[13]The county allocations under the current two-stage allocation process correspond only approximately to the county model estimates because of other factors in the allocation formula, such as hold-harmless provisions.

ment of Education has data) do not receive basic grants: although there is no eligibility threshold for counties to qualify for basic grants, school districts must have at least 10 formula-eligible children, and the number of eligible children must exceed 2 percent of the total number of school-age children in the district. Nonetheless, for basic grants, the county totals would likely be fairly similar whether direct allocations are made to school districts or the two-stage process is continued.

However, if states choose the option of redistributing the aggregate of the direct allocation amounts for school districts with fewer than 20,000 people by using some other data source (such as school lunch data), then the county totals for these districts may not be similar to the county amounts under the two-stage process.[14] The panel has a concern about this possible outcome: the county allocations that are made under the current two-stage process reflect (approximately) the Census Bureau's county estimates from its county model, and these estimates are the only small-area estimates of poor school-age children that have been thoroughly evaluated and determined to be reasonably reliable.[15] Direct allocations that use the Census Bureau's synthetic shares school district estimates would also reflect (approximately) the Bureau's county estimates, but state plans to redistribute the direct allocation amounts for school districts with fewer than 20,000 people by using some other data source may not have this desirable property.

Analysis with 1989 school lunch data for New York State districts with fewer than 20,000 people (476 districts, see the appendix, Table A-9) did not find evidence of this problem. The average absolute and average proportional absolute differences from 1990 census school district estimates of poor school-age children were about the same for estimates that were developed by using free lunch counts with and without county controls. However, this analysis pertains to only one alternate data source in only one state. In the absence of a complete analysis of alternate data sources, the panel believes it is desirable, to the extent possible, that the basic grant allocations reflect the county model estimates in all states, including those that choose the option of redistributing the aggregate of the direct allocations for school districts under 20,000 population by using another data source. The Department of Education can achieve this outcome by

[14]Presumably, the states that are more likely to choose this option are the 25 states that, at present, use another data source (e.g., free lunch data, free and reduced-price lunch data, or AFDC data) as the only factor or as one of the factors in allocating county allocations to districts. School districts with fewer than 20,000 people in these 25 states were 46 percent of total districts nationwide in 1990, containing 13 percent of total school-age children.

[15]For example, the county estimates of poor school-age children developed from the county model are much more reliable than county estimates developed by synthetic methods, such as applying within-state county shares of poor school-age children in the previous census to updated estimates from the Census Bureau's state model (see National Research Council, 1998:Table 4-2).

approving state reallocation plans that, in general, propose to aggregate the direct allocation amounts for districts under 20,000 population within counties and redistribute the county totals among the districts under 20,000 population in each county.

Concentration Grants

Concentration grants, in contrast to basic grants, are not allocated as shares of the county totals because only a fraction (less than half) of jurisdictions are eligible.[16] Under the current two-stage process, concentration grants are allocated to those counties that have at least 6,500 or more than 15 percent of formula-eligible school-age children. In turn, states allocate county concentration grants to those districts in eligible counties that exceed the threshold number or percentage of formula-eligible children: most districts that qualify for concentration grants will do so on the basis of exceeding the percentage threshold; few will do so on the basis of having more than 6,500 formula-eligible children.

Tabulations of 1990 census data in the evaluation file identified 30 percent of school districts, containing 60 percent of poor school-age children, as eligible for concentration grants under the current two-stage allocation process.[17] Eligible districts under the two-stage process were 65 percent of the total districts in eligible counties. (In states that use another data source to distribute county concentration amounts to districts, such as free lunch participants, a higher percentage of school districts in eligible counties may be classified as eligible for concentration grants; see below.)

The census tabulations showed that an additional 9 percent of school districts, containing 14 percent of poor school-age children, would be eligible for concentration grants if they were located in an eligible county. Currently, states may reserve up to 2 percent of their concentration grant funds to allocate to eligible districts that are not in eligible counties, but these amounts are probably not adequate for the children in those districts.

We note that the use of fixed thresholds for concentration grants places great demands on the quality of the estimates of those thresholds. An error of only one poor school-age child can make the difference between receiving a grant and not receiving a grant. For school districts that receive concentration grants to which they would not be entitled if true estimates of poor school-age children were available, these errors will be perpetuated through the hold-harmless provisions,

[16]In contrast, all counties and almost 90 percent of school districts are eligible for basic grants.

[17]The tabulations were limited to districts in the 1980-1990 evaluation file for which the boundaries did not cross county lines, totaling 6,434 districts, or 70 percent of the districts in the evaluation file. The classification of counties and school districts as eligible for concentration grants considered only the criterion of having a school-age poverty rate of more than 15 percent.

particularly if the hold-harmless rate is retained at 100 percent. (There are also fixed thresholds for school districts to receive basic grants, although they are low, as noted above.)[18]

Evaluation

One of the reasons for the legislation mandating direct allocations to school districts was to target concentration grants to all eligible school districts, including those in ineligible counties. To assess the appropriateness and reliability of the Census Bureau's updated school district estimates of poor school-age children for direct allocation of concentration grants, the panel first examined the rate of agreement between the Census Bureau's synthetic shares method (1) and the 1990 census in classifying school districts into one of two poverty rate categories for school-age children in 1989 that correspond to the concentration grant threshold: 0 to 15 percent and 15 percent or higher; see Table 3-7. The tabulations were prepared from the 1980-1990 evaluation file for districts that did not cross county lines.

The synthetic method (1) school district estimates and the 1990 census ratio-adjusted estimates for 1989 assigned the same poverty rate category (0 to 15% or 15% or higher) to 76 percent of school districts and 87 percent of poor school-age children. By comparison, the county model estimates and the 1990 census county estimates for 1989 assigned the same poverty rate category to 88 percent of counties and 92 percent of poor school-age children. The rate of agreement between the synthetic method (1) school district estimates and the 1990 census ratio-adjusted estimates was least for school districts with fewer than 5,000 people: 64 percent agreement for districts and 65 percent agreement for poor school-age children.[19] The rate of agreement was highest for school districts with 40,000 or more people: 92 percent for both districts and poor school-age children, slightly better than the rate of agreement for counties. For school districts for which the synthetic method and the 1990 census estimates were not in agreement (24% in terms of districts and 13% in terms of poor school-age children), the synthetic method classified a much higher percentage as having a school-age poverty rate of under 15 percent than did the census estimates.

To focus on the issue of concentration grant eligibility for school districts with direct allocations versus the current two-stage process, the panel examined the correspondence between the synthetic method (1) estimates and the 1990

[18]For a discussion of issues in the relationship of funding formulas and data sources; see Zaslavsky and Schirm, 1998.

[19]At least part of the explanation is that the census comparison estimates are subject to particularly high sampling variability for the smallest districts (see Table 3-2).

TABLE 3-7 Agreement Between Synthetic Method (1) Estimates and 1990 Census School District Estimates for Proportions of School-Age Children in Poverty in 1989

Method of Estimate	Percentage of School Districts	Percentage of Poor School-Age Children
Method (1) and Census Estimate, All Districts		
Both under 15%	50.0	25.6
Both 15% or more	25.7	60.9
(Total in agreement)	(75.7)	(86.5)
Census under 15%, method (1) 15% or more	8.8	2.5
Census 15% or more, method (1) under 15%	15.6	11.0
Method (1) and Census Estimate, Districts Under 5,000 Population		
Both under 15%	37.6	20.2
Both 15% or more	26.6	44.9
(Total in agreement)	(64.2)	(65.1)
Census under 15%, method (1) 15% or more	14.1	6.4
Census 15% or more, method (1) Under 15%	21.6	28.5
Method (1) and Census Estimate, Districts of 40,000 or More Population		
Both under 15%	59.8	22.0
Both 15% or more	31.8	70.0
(Total in agreement)	(91.6)	(92.0)
Census under 15%, method (1) 15% or more	2.4	1.3
Census 15% or more, method (1) under 15%	6.0	6.8
County Model and Census Estimate, All Counties		
Both under 15%	30.5	40.9
Both 15% or more	57.1	50.7
(Total in agreement)	(87.6)	(91.6)

NOTES: School district estimates are based on 9,243 districts in the 1980-1990 evaluation file. The 1990 census estimates for school districts are the ratio-adjusted estimates (see text). The method (1) school district estimates are produced by applying 1980 census within-county school district shares of poor school-age children to the county model estimates for 1989 and controlling to the 1990 census national estimate of poor school-age children in 1989.

SOURCE: Data from Bureau of the Census; see National Research Council (1988:Table 4-4 [model b]) for county model comparisons.

census estimates for cross-classifications of 1989 school district and county school-age poverty rate categories; see Tables 3-8 and 3-9. The synthetic method (1) estimated that 32 percent of districts, containing 59 percent of poor school-age children, would be eligible for a concentration grant under the two-stage process (cell f, Tables 3-8 and 3-9). Another 10 percent of districts, containing 12 percent of poor school-age children, would be eligible for a concentration

grant under direct allocations (cell o). These aggregate percentages are similar to those for the 1990 census, noted above (see cells h and q in Tables 3-8 and 3-9), but the synthetic method and the 1990 census classified a number of districts differently.

Of the districts and poor school-age children that the 1990 census estimated would be eligible for concentration grants under the two-stage process, the synthetic method (1) agreed for 86 percent of districts and 96 percent of poor school-age children (cell e divided by cell h). The other 14 percent of districts and 4 percent of poor school-age children would be ineligible for concentration grants under the two-stage process according to the synthetic method (1). There are also districts and poor school-age children that would be eligible under the two-stage process according to the synthetic method (1) but ineligible according to the 1990 census: they comprise 18 percent of the districts and 3 percent of the poor school-age children that are eligible according to the synthetic method (1) (cell d divided by cell f).

Of the additional districts and poor school-age children that the 1990 census estimated would be eligible for concentration grants under direct allocations (i.e., those in counties with school-age poverty rates under 15%), the synthetic method agreed for 53 percent of districts and 76 percent of poor school-age children (cell n divided by cell q). The other 47 percent of the additional districts and 24 percent of the additional poor school-age children would be ineligible according to the synthetic method (1). There are also additional districts and poor school-age children that would be eligible according to the synthetic method (1) but ineligible according to the 1990 census: they comprise 49 percent of the additional districts and 10 percent of the additional poor school-age children that are eligible according to the synthetic method (1) (cell m divided by cell o).

Overall, the classification differences between the 1990 census estimates and the synthetic method (1) estimates are relatively large for the additional districts that would be eligible under direct allocations (i.e., districts with 15% or more poor school-age children in counties with less than 15% poor school-age children). However, the classification differences are relatively small for the additional poor school-age children that would be eligible under direct allocations. In particular, the percentage of poor school-age children in the additional districts that would be eligible for concentration grants according to the synthetic estimates but would not be eligible according to the 1990 census estimates is relatively small (10%).

It should be kept in mind that these evaluations are limited in at least three ways. First, they apply only to a subset of school districts in the evaluation file, which are, themselves, a subset of total districts. Second, like all of the evaluations of the Census Bureau's school district estimates, they are based on a single time comparison. Third, the 1990 census estimates that are the standard of comparison are subject to high sampling variability for smaller school districts even with ratio adjustment.

TABLE 3-8 Comparison of Synthetic Method (1) and 1990 Census School District Estimates for Proportions of School-Age Children in Poverty in 1989, by 1990 Census County School-Age Poverty Rate: Distribution by Percentage of School Districts

CENSUS COUNTY SCHOOL-AGE POVERTY RATE 15% OR MORE

	Census School District Rate		
	Under 15%	15% or More	Total
Method (1) School District Rate			
Under 15%	10.8 (a)	4.3 (b)	15.1 (c)
15% or more	5.7 (d)	25.9 (e)	31.6 (f)
Subtotal	16.6 (g)	30.2 (h)	46.8 (i)

CENSUS COUNTY SCHOOL-AGE POVERTY RATE UNDER 15%

	Census School District Rate		
	Under 15%	15% or More	Total
Method (1) School District Rate			
Under 15%	38.9 (j)	4.4 (k)	43.3 (l)
15% or more	4.9 (m)	5.0 (n)	9.9 (o)
Subtotal	43.8 (p)	9.4 (q)	53.2 (r)
Total	60.4	39.6	100.0

NOTES: The two poverty rate categories used are those specified for concentration grants, 0-15 percent and 15 percent or more.

Cell entries are percentages of the 6,434 school districts in the 1980-1990 evaluation file for which the boundaries did not cross county lines. The 1990 census county and school district estimates are the ratio-adjusted estimates (see text). The method (1) school district estimates are produced by applying 1980 census within-county school district shares of poor school-age children to the county model estimates for 1989 and controlling to the 1990 census national estimate of poor school-age children in 1989. See text for discussion.

SOURCE: Data from Bureau of the Census.

TABLE 3-9 Comparison of Synthetic Method (1) and 1990 Census School District Estimates for Proportions of School-Age Children in Poverty in 1989, by 1990 Census County School-Age Poverty Rate: Distribution by Percentage of Poor School-Age Children

CENSUS COUNTY SCHOOL-AGE POVERTY RATE 15% OR MORE

	Census School District Rate		
	Under 15%	15% or More	Total
Method (1) School District Rate			
Under 15%	6.1 (a)	2.5 (b)	8.6 (c)
15% or more	1.5 (d)	57.5 (e)	59.0 (f)
Subtotal	7.5 (g)	60.0 (h)	67.5 (i)

CENSUS COUNTY SCHOOL-AGE POVERTY RATE UNDER 15%

	Census School District Rate		
	Under 15%	15% or More	Total
Method (1) School District Rate			
Under 15%	17.3 (j)	3.3 (k)	20.6 (l)
15% or more	1.2 (m)	10.7 (n)	11.9 (o)
Subtotal	18.5 (p)	14.0 (q)	32.5 (r)
Total	26.0	74.0	100.0

NOTES: The two poverty rate categories used are those specified for concentration grants, 0-15 percent and 15 percent or more.

Cell entries are percentages of poor school-age children in 1989 in the 6,434 school districts in the 1980-1990 evaluation file for which the boundaries did not cross county lines. The 1990 census county and school district estimates are the ratio-adjusted estimates (see text). The method (1) school district estimates are produced by applying 1980 census within-county school district shares of poor school-age children to the county model estimates for 1989 and controlling to the 1990 census national estimate of poor school-age children in 1989. See text for discussion.

SOURCE: Data from Bureau of the Census.

Understanding the limits of the evaluations and the alternatives available, the panel concludes, on balance, that the use of the Census Bureau's school district estimates for direct allocations of concentration grants would be an improvement over the current two-stage process. As intended by the 1994 legislation, many of the eligible districts that could not receive concentration grants with a two-stage allocation would receive such grants with direct allocations.

Reallocation of Concentration Grants

The option for states to redistribute concentration grant direct allocations for school districts with fewer than 20,000 people raises several issues. Presumably, states might propose to use another method to redistribute the allocations among the districts that the Department of Education determined to be eligible for concentration grants on the basis of the Census Bureau's estimates. Or, states might propose to use another method to redetermine both eligibility and allocation amounts. (The states that currently distribute county concentration grant allocations to districts on the basis of some other data source than the census use the alternate data source for both eligibility and amounts.)

The use of free lunch or free and reduced-price lunch data in place of estimates of poor school-age children to redetermine eligibility as well as to redistribute allocation amounts would likely have the effect that more districts receive concentration grants than they would with the use of the Census Bureau's school-age poverty estimates. The reason is that the income eligibility thresholds for free or reduced-price school lunches are higher than the poverty threshold. Consequently, more children fall below 130 percent of poverty (the threshold for free lunches) or below 185% of poverty (the threshold for reduced-price lunches) than fall below 100% of poverty.[20] (About 20% of school-age children nationally are in families with incomes below 100% of the poverty threshold, while about 26% are in families with incomes below 130% of the poverty threshold and about 38% are in families with incomes below 185% of the poverty threshold.)[21] For this same reason, it is likely that proportionately more districts are currently receiving concentration grants under the two-stage process in states that use school lunch data to determine eligibility than in states that use 1990 census data. In either case, the effect is to spread concentration grant dollars more thinly.

Analysis with 1989 school lunch data for New York State school districts with fewer than 20,000 people (476 districts; see the appendix, Tables A-5 through A-8) provides evidence of the effect of using estimates that reflect higher poverty thresholds. Under the two-stage process, 136 such districts in New York

[20]However, not all eligible children apply for reduced-price lunches.

[21]Data from panel tabulations of the March CPS for income years 1994-1996.

State would be eligible for concentration grants by using free and reduced-price lunch data and 112 would be eligible by using free lunch data, whereas only 76 districts would be eligible according to the synthetic method (1) estimates (or the 1990 census). Under direct allocations, the effect is much more pronounced: 294 districts with fewer than 20,000 people would be eligible for concentration grants by using free and reduced-price lunch data, and 214 districts would be eligible by using free lunch data, whereas only 109 districts would be eligible according to the synthetic method (1) estimates (115 districts according to the 1990 census).

The panel concluded that any redistribution of basic grant direct allocations for districts with fewer than 20,000 people should be performed for such districts within each county to the extent possible, thereby reflecting (approximately) the county estimates of poor school-age children. For concentration grants, the panel reaches the same conclusion, although it should be noted that there may be a problem with this approach when different data are used for reallocation. For example, if a county has two school districts and only one district is eligible for a concentration grant according to the Census Bureau's estimates of poor school-age children, but both districts are eligible by using school lunch data, then the first district will lose some of its dollars to the second district. Presumably, similar situations occur under the current two-stage allocation process, in which school district concentration grants are allotted from county totals.[22] However, such situations may be somewhat more likely to occur under direct allocations, which will provide concentration grants to eligible districts in counties that do not meet the concentration grant threshold.

One approach that could ameliorate this effect is to adjust school lunch data for school districts in a county to equal the Census Bureau's estimate of total poor school-age children for the county. The use of adjusted school lunch data to determine school-age poverty rates would be less likely to result in a much larger number of school districts qualifying for concentration grants than the use of the Census Bureau's estimates of school-age poverty rates. Analysis conducted for New York State confirmed this outcome (see Appendix, Tables A-7, A-8): 127 school districts with fewer than 20,000 people would be eligible for concentration grants under direct allocations by using adjusted free and reduced-price lunch data versus 294 districts that would be eligible by using unadjusted data. The corresponding figures are 124 districts and 214 districts by using adjusted and unadjusted free lunch data. By comparison, 109 districts would be eligible by using the synthetic method (1) estimates.

[22]The New York State analysis, in which more districts were eligible for concentration grants under the two-stage process by using school lunch data than by using the synthetic estimates, suggests that such situations currently occur.

Study of Allocation Process

Overall, by applying a relative standard for evaluation, the panel finds reasons to support the use of the Census Bureau's updated estimates of poor school-age children for direct allocation to school districts. Also, the panel concludes that, in general, it is desirable for both basic grant and concentration grant allocations to reflect the county model estimates in all states, including those that choose the option of redistributing the direct allocations for school districts under 20,000 population by using another data source. However, the panel recognizes that there are uncertainties about the operation of the formulas: for example, the extent to which the sum of direct school district allocations for counties will approximate the allocations that would result for counties under the current two-stage process and the extent to which there may be significant reallocations of concentration grant dollars from poorer to less poor districts with county controls. For this reason, the panel believes it is critically important for the Department of Education to undertake a thorough study of the direct allocation process, both the methods used by the states and the results. Simulations of the allocations that would likely have been made under the two-stage process would be very helpful to inform the study.

4

Recommendations for Title I Allocations for the 1999-2000 School Year

The Department of Education, following the recommendation of the Panel on Estimates of Poverty for Small Geographic Areas (National Research Council, 1998), allocated Title I funds to counties for the 1998-1999 school year by using the estimates from the Census Bureau's revised county model of the numbers and proportions of poor school-age children for 1993. ("County model" is used in the broad sense to include the entire estimation procedure, which comprises a county regression model, a separate state regression model, and county population estimates.) The Census Bureau has now used its model to produce county estimates of poor school-age children for 1995. In addition, it has produced school district estimates of poor school-age children for 1995 by applying within-county school district shares of poor school-age children from the 1990 census to the 1995 county estimates. The Census Bureau has also produced school district estimates of total population and total school-age children for 1996 by using a similar synthetic shares method, to accompany the updated poverty estimates.

The 1994 Title I legislation mandates that, beginning with the 1999-2000 school year, the Department of Education use the Census Bureau's updated school district estimates to make Title I allocations directly to school districts unless the Secretaries of Education and Commerce determine that "some or all of the data" are "inappropriate or unreliable" for this purpose on the basis of the panel's study. A direct allocation procedure would replace the current two-stage procedure, in which the department allocates Title I funds to counties and the states then distribute the county funds to school districts. Under direct allocation, the states would have the option of aggregating the school district Title I allotments for districts with fewer than 20,000 people and redistributing the aggregate

amounts to districts on the basis of another method that is approved for this purpose by the Department of Education (e.g., by using counts of participants in the National School Lunch Program).

ASSESSMENT OF 1995 COUNTY ESTIMATES

The model that was used by the Census Bureau to produce county estimates of school-age children in 1996 who were in poor families in 1995 is essentially the same model that was used to produce the revised county estimates of poor school-age children for 1993. On the basis of the internal and external evaluations that were conducted of alternative 1993 county models, the panel supported the use of the revised 1993 county estimates for Title I allocations for the 1998-1999 school year.

Additional internal and external evaluations were conducted of the 1995 county model that focused on the behavior of the county model (and the state model) when estimated for several time periods. Both sets of evaluations—for 1993 and 1995—identified areas for further research and development, including several areas of work that should be completed well before the next round of estimates is produced in fall 2000 (see Chapter 5). Overall, however, the 1995 evaluations confirmed the results of the 1993 evaluations, which showed that the county model is performing well. There remains, however, the question of whether the 1995 school district estimates, produced by applying 1990 census within-county district shares of poor school-age children to the 1995 county estimates, are appropriate for direct Title I allocations. The panel summarizes the pros and cons of using the 1995 school district estimates in the next section and then presents its recommendations for Title I allocations for the 1999-2000 school year.

ASSESSMENT OF 1995 SCHOOL DISTRICT ESTIMATES

It is not possible to develop very reliable estimates of poor school-age children for most school districts with the currently available data. The available data are inadequate not only because most school districts are small in population size, but also because the boundaries of many school districts do not coincide with the boundaries for counties or other governmental units, the boundaries can and often do change over time, and some school districts do not serve all elementary and secondary grades.

For 1990, the data on school-age poverty (from the ratio-adjusted long-form sample of the census) have a large degree of sampling variability for many districts because of their small population size (see Chapter 3). Moreover, for years following the census, no data are currently available on a consistent basis for all districts with which to estimate changes in their poverty population (or total population) over time. Food stamp and federal tax return data are not

currently available for districts. Data on participation in the National School Lunch Program for school districts that are provided annually by the states to the National Center for Education Statistics are not consistent or complete. Also, variations in program participation make it unclear how reliably school lunch data indicate differences in poverty among school districts across the nation.

The limitations of available data for school districts constrained the Census Bureau to use a basic synthetic shares approach. In this procedure, 1990 census within-county school district shares of poor school-age children, reflecting 1995-1996 district boundaries, were applied to the 1995 county estimates of poor school-age children developed from the county model. By definition, this procedure reflects only the changes over time in the numbers of poor school-age children for school districts that parallel the changes in the counties in which they are located.

Evaluations of the Census Bureau's procedure over the 1980-1990 period, by applying 1980 census within-county school district shares of poor school-age children to the county model estimates for 1989, revealed large differences between the synthetic estimates of poor school-age children for school districts and the ratio-adjusted estimates from the 1990 census. The reasons for the differences include: (1) the sampling variability in the 1980 census estimates of school district shares; (2) within-county changes in school district shares of poor school-age children from 1980 to 1990, which the synthetic shares method cannot capture; (3) errors in the county model (which are not a major factor); and (4) the sampling variability in the 1990 census comparison estimates. Even if the sampling variability in the 1990 census estimates were removed, the differences between the synthetic estimates and the census estimates would be large for many districts.

However, for districts that are coterminous with counties, the estimates are reasonably precise because they come from the county model. In addition, for districts with 40,000 or more people, the estimates are not markedly worse than the county model estimates. Together, these two groups of districts comprised only 13 percent of all districts in 1990, but they contained 62 percent of poor school-age children.

Use of Estimates for Allocations

Although the level of error in the Census Bureau's 1995 estimates of poor school-age children is potentially high for many school districts, the panel nonetheless concludes that the estimates are not inappropriate or unreliable to use for direct Title I allocations to districts as intended by the 1994 legislation. Central to the panel's conclusion is that it interprets "inappropriate or unreliable" in a relative sense in this context. Some set of estimates must be used to distribute Title I funds to school districts. The question is whether the Census Bureau's 1995 estimates are more appropriate and reliable than those produced by the

current procedures. Moreover, the clear congressional intent to move to direct allocations argues for the use of the Census Bureau's 1995 estimates even if they are not better than the estimates that the states are currently using to distribute the county allocations to school districts, so long as they are not appreciably worse than those estimates.

Three main findings support the panel's conclusion that the Census Bureau's 1995 school district estimates are not "inappropriate or unreliable" in a relative sense for Title I allocations. First, many states currently use a method for distributing county Title I funds to school districts that is similar to the Census Bureau's 1990 census-based synthetic shares method. The Census Bureau's estimates of within-county district shares for these states are likely to be better than the estimates that the states have been producing because the Census Bureau has additional geographic information. The Census Bureau has access to the 1990 census block data for retabulating numbers of poor school-age children according to updated school district boundaries. Also, the Census Bureau's estimates of within-county district shares of poor school-age children are ratio adjusted, which somewhat reduces their sampling variability.

Second, limited evaluation of school lunch data for one state suggests that the within-county school district shares of poor school-age children produced from such data are not appreciably better than the 1990 census shares, even though the school lunch data used in the evaluation pertained to the same year as the standard of comparison. (Further evaluation of school lunch data for other states would be desirable.)

Third, direct allocation using the Census Bureau's 1995 estimates addresses the inequities that result with the current two-stage allocation process for concentration grants whereby some eligible school districts do not receive these grants because they are in counties that are not eligible. According to 1990 census data, about 30 percent of school districts, containing about 60 percent of poor school-age children, are eligible for concentration grants under the current two-stage allocation process. An additional 9 percent of school districts, containing about 14 percent of poor school-age children, would be eligible for concentration grants under direct allocations. The Census Bureau's synthetic shares procedure, based on 1980 census school district shares applied to 1989 county model estimates, identified similar percentages of school districts and poor school-age children that would be eligible for concentration grants under direct allocations but would not be eligible under the two-stage process. Only about half of the school districts that were classified by one source—the 1990 census or the synthetic estimates—as being eligible for concentration grants under direct allocations but not under the two-stage process were so classified by the other source. However, the degree of agreement was much higher when expressed in terms of numbers of poor school-age children (see Tables 3-8 and 3-9). On balance, these results indicate that direct allocations with the Census Bureau's 1995 school district

estimates could be expected to help redress the inequities of the two-stage allocation process for concentration grants.

Reallocations for Small School Districts

If the Census Bureau's 1995 school district estimates of poor school-age children are used for direct allocations to districts, the 1994 Title I legislation allows states to aggregate the Department of Education's allocations for districts with fewer than 20,000 people and redistribute the aggregate amounts by using some other method that is approved by the department. The department has determined that states can use an alternative approved method for redistributing the department's allocations for districts under 20,000 population both for basic grants and concentration grants, including using the alternative method to redetermine which districts are eligible for concentration grants. The 1990 census estimated that over 80 percent of the nation's school districts had less than 20,000 population, although these districts included only 27 percent of total school-age children.

The panel believes that it is important to take account of the 1995 county estimates from the Census Bureau's county model in any reallocation that states choose to do because these estimates are the only updated small-area estimates of school-age children in poverty that have been shown to be reasonably accurate on the basis of a thorough evaluation. Thus, reallocations for school districts under 20,000 population will likely be more accurate to the extent that they reflect the updated county estimates. The Department of Education can ensure that any reallocations reflect (approximately) the county estimates by requiring that, to the extent possible, plans aggregate the grant amounts for districts under 20,000 population into county subtotals and reallocate the county subtotals, rather than reallocating the total for districts with under 20,000 population without regard to counties. The panel believes aggregation to county subtotals should be performed separately for basic and concentration grants.

RECOMMENDATIONS

The 1994 Title I legislation expressed the congressional intent to move to a system of direct allocations of Title I funds to school districts provided that the Census Bureau's estimates are adequate for this purpose. The panel concludes that the Bureau's 1995 school district estimates of poor school-age children are not inappropriate or unreliable for this purpose in a relative sense. While subject to large errors for many districts, they are at least as good as, if not better than, the estimates that are currently being used for many district allocations. Moreover, direct allocation with the Census Bureau's estimates partly addresses known inequities in the two-stage allocation process for concentration grants.

(1) The panel recommends to the Secretaries of Education and Commerce that the Census Bureau's 1995 school district estimates of poor school-age children be used to make direct Title I allocations to school districts for the 1999-2000 school year.

The panel believes that it is important to reflect the Census Bureau's updated estimates of poor school-age children for counties in the direct allocations to school districts. The county estimates are the only postcensal small-area estimates of poverty that have been thoroughly evaluated. For states that accept the department's direct allocations for all districts, the county estimates will be reflected (approximately) in the allocations because of the use of the county estimates in the synthetic shares estimation procedure for districts. For states that choose to reallocate the amounts for school districts with less than 20,000 population, the Department of Education should require, in general, that their plans reallocate the amounts for school districts on a county-by-county basis.

(2) The panel recommends that any state plan approved by the Department of Education for redistributing the sum of the department's allocations for school districts with under 20,000 population maintain the county total amounts for such districts to the extent possible.

The panel recognizes that developing county aggregate allocations from school district allocations is not always straightforward. In particular, states must determine the individual county components for school districts that cross county lines. However, states already must determine the individual county components for such districts under the current two-stage allocation process. Moreover, the Department of Education currently allows several states—those in which school district boundaries bear little relationship to county boundaries—to ignore the county allocations in distributing the total allocation for the state among school districts, and the department could grant similar exceptions for state plans to reallocate amounts for districts under 20,000 population.

The panel also recognizes that the development of school district estimates of poor school-age children is a complex process for the Census Bureau and that the use of these estimates for direct allocations imposes burdens on the Department of Education to obtain the additional data that are needed to implement the formulas (e.g., counts of the other categories of formula-eligible children and the dollar amounts of Title I allocations that school districts received in the previous year).

Because direct allocation of Title I funds to school districts is a new procedure, it will be important to assess how it is implemented and its effects in comparison with the current two-stage allocation procedure. The Department of Education should plan to conduct a thorough study in this regard, including such aspects as the operation of hold-harmless provisions and the differences in the

number of districts and poor school-age children that are eligible for concentration grants.

LOOKING TO THE FUTURE

Marked improvement in school district estimates of poor school-age children will require a substantial research and development effort. However, modest improvements in the estimates can likely be made in the near term with additional work to modify and evaluate the current 1990 census-based synthetic shares method. Specifically, a priority for the near term should be work to further reduce the sampling variability in the estimates of the 1989 shares of poor school-age children within counties that are derived from the census long-form sample. One approach is to use additional short-form data in the estimation of poor school-age children. Another approach is to smooth the 1990 census school district estimates with the 1990 census county estimates (see Chapter 5).

Looking to the longer term, it will be necessary to obtain relevant administrative records data at the school district level if there are to be significant improvements in the school district estimates of poor school-age children. School district-level administrative data (e.g., federal tax returns with addresses geocoded to districts) could provide the basis for model-based estimates of poor school-age children for school districts that more fully reflect changes in the extent of poverty among districts over time. In addition, data from the 2000 census and the planned American Community Survey have the potential to markedly improve the estimates for school districts. Research and development in these areas should begin as soon as possible.

SPECIAL CASE: PUERTO RICO

The Title I allocations include Puerto Rico, which has been treated as a county equivalent under the two-stage allocation process and will be treated as a single school district coterminous with a county for direct allocations. While the commonwealth's 1990 decennial census provides estimates for 1989, no estimates of Puerto Rican school-age children in poverty can be made for 1995 from the Census Bureau's county model because the appropriate federal tax and food stamp data are not available for Puerto Rico. The Census Bureau computed 1995 estimates for Puerto Rico from data collected in the 1996 Puerto Rican Family Income Survey that was conducted in the commonwealth in February-March 1997. Several adjustments had to be made to produce the estimates of school-age children in poverty in 1995.

The Census Bureau previously used a similar approach to compute 1993 estimates of poor school-age children with data from an earlier round of the Puerto Rico income survey. The panel concluded in its first interim report that the Bureau's approach for producing updated estimates of poor school-age chil-

dren in Puerto Rico seemed a reasonable one given the data available, although there was limited information about the quality of the data (see National Research Council, 1997:App. F). Additional information was obtained from Puerto Rico about the quality of the income survey that, in general, supported the use of the survey data to develop 1995 estimates of the number of poor school-age children for Puerto Rico. Consequently, the panel recommends that the 1995 estimates for Puerto Rico be used in the direct Title I allocations for the upcoming 1999-2000 school year.

The Puerto Rico Family Income Survey is expected to be conducted at regular intervals in the future. It will presumably be the basis of updated estimates of poor school-age children in Puerto Rico for 1997 and later years. Through cooperative work with Puerto Rico, the Census Bureau should continue its evaluations of the quality of the estimates and their comparability with the model-based estimates for U.S. counties to determine if there are ways in which the data and estimation procedures for Puerto Rico can be improved for use in Title I allocations.

5

Future Research and Development

There are several reasons that make it important for the Census Bureau to continue an active program of research and development for methods of estimating poverty for school-age children at the county and school district levels. For counties, although there is clear evidence that the county model is performing well, the county (and state) model evaluations have identified a number of issues that warrant investigation as a priority in the short term to determine how to further improve the estimation procedures. Also, with a model-based approach, it is important to examine carefully the continued applicability of a model each time it is used and to modify it appropriately when necessary. In addition, research is needed to take account of likely future developments in the availability and characteristics of data sources that have implications for the modeling effort and to work on longer term modeling issues. Continued work to improve the county model is important not only for county estimates, but also to improve school district estimates that are developed by using the basic synthetic shares estimation procedure.

For school districts, the important short-term priority is to investigate ways to improve the synthetic shares method for developing updated estimates of total and poor school-age children. Also, it is not too soon to begin research on ways to take advantage of likely future developments in available data that could make it possible to develop an estimation method that (unlike the shares method) captures changes in school-age poverty among districts within counties as well as changes between counties.

The chapter begins by reviewing the schedule for the Census Bureau to provide updated small-area estimates of poor school-age children. It then consid-

ers short-term and longer term research priorities for county and school district estimates. It concludes by noting the requirements for an ongoing program of small-area income and poverty estimates, particularly for thorough evaluation and full documentation of models and results.

SCHEDULE CONSIDERATIONS

Over the next 5 years, there are three legislatively mandated deadlines for the Census Bureau to deliver updated school district estimates of poor school-age children to the Department of Education for use in Title I allocations:

- October 2000: estimates for 1997 (or later) for use for allocations for the 2001-2002 and 2002-2003 school years
- October 2002: estimates for 1999 (or later) for use for allocations for the 2003-2004 and 2004-2005 school years
- October 2004: estimates for 2001 (or later) for use for allocations for the 2005-2006 and 2006-2007 school years

In each case, three estimates are needed: number of total and poor school-age children and the total population. Although the legislation does not require county estimates, they will be needed as long as the method for producing school district estimates includes an adjustment or control to county estimates. There is also interest in state and county estimates of poor children for other important public policy uses, such as evaluating the effects of changes in welfare programs.

Priorities for short-term and longer term research should consider the important changes that are likely to occur in the availability of data for modeling over the next 5 years and beyond, which include:

- current and future changes to welfare programs and tax systems that may affect the comparability or applicability of Food Stamp Program and Internal Revenue Service (IRS) data for use in small-area estimation models;
- the income and poverty estimates for small areas that will be available from the 2000 decennial census long-form sample of about 17 million households (likely to be available in 2002 for counties but not until later for school districts); and
- the planned introduction of the American Community Survey (ACS) as a large-scale, continuing sample survey of U.S. households, conducted primarily by mail, that will provide estimates similar to those provided by the decennial census long-form sample, including income and poverty estimates for small areas. The ACS is currently being tested in 4 sites; under current plans, it will be implemented in 31 sites in 1999-2001 for comparison with the 2000 census. For each year from 2000 to 2002, the ACS will sample about 70,000 households nationwide. Beginning in 2003, the ACS will sample 250,000 households each

month throughout the decade, for an annual sample size of about 3 million households. The current plan is that the ACS (as well as the 2000 census long-form sample) will oversample small jurisdictions. Unlike the 1990 census, the oversampling in the 2000 census and the ACS will include small school districts (see Alexander, 1998).

SHORT-TERM PRIORITIES

County Estimates

The panel identified seven types of research that should be pursued as a priority to determine if the current estimation procedure for counties can be improved: modeling of CPS county sampling variances; estimation of model error and sampling error variance in the state model; methods to incorporate state effects in the county model; discrete variable models that include counties in the CPS sample that have no sampled households with poor school-age children; ways to reduce the time lag of the estimates; evaluation of food stamp and other input data; and large category differences and residual patterns for the state and county models. This research, much of which the Census Bureau has planned, should be conducted and the results fully evaluated well before the next delivery of updated county estimates of poor school-age children, scheduled for October 2000.

Modeling of CPS County Sampling Variances The residual variance for the county model comprises two components: the model error variance and the sampling variance of the dependent variable. These two components need to be reasonably well estimated for the application of the model (e.g., to determine the relative weights of the regression estimate and the direct estimate in the shrinkage procedure). The current approach for estimating these components is to assume that the model error variance from the 1989 regression equation with the dependent variable formed from 1990 census data is the same as the model error variance when the dependent variable is formed from the 3 years of CPS data that are used for the county model equation for the target year. The total sampling variance is then obtained simultaneously with the regression parameter estimates through use of maximum likelihood estimation. As part of this procedure, the sampling variance for a particular county is assumed to be inversely proportional to the CPS sample size in that county.

There is ample evidence that the function that is now used to distribute the total sampling variance to counties is incorrect (see Chapter 2). Experimentation with other functions, which has already begun at the Census Bureau (specifically, investigating a function in which the sampling variance is inversely proportional to the square root of the CPS sample size in a county), should be pursued to eliminate or reduce the problem of variance heterogeneity with respect to both

the CPS sample size and poverty rate that is evident in the county model regression output. Research on this topic should include an assessment of the effects of alternative variance functions on the county estimates.

In addition, the Census Bureau should pursue an alternative approach, which is to estimate the CPS sampling variances for counties with adequate sample size on the basis of direct calculations of these variances that take account of the clustered sample design within these counties, and then use a generalized variance function for modeling the sampling variances for all counties with CPS sampled households. With this approach, the model error variance is calculated by subtracting the total sampling variance from the total squared error. This approach thus avoids the questionable assumption that the model error variances for the 1989 census equation and the CPS equation for the target year are equal. Census Bureau staff have begun work on fitting a generalized variance function to the CPS sampling variances. This work should continue and should include an early assessment of the effects on the county estimates to determine if the benefits justify continued refinement of the variance modeling.

Model Error and Sampling Error Variance in the State Model In the state model the model error variance is obtained from a maximum likelihood procedure that estimates the coefficients of the predictor variables and the model error variance, given estimates of the sampling error variances of the direct state estimates. For most years for which the state model has been estimated, this procedure estimates the model error variance as zero, which results in zero weight being given to the direct CPS estimates. In effect, the model is assumed to be without error, which is not credible. A likely explanation is that the Census Bureau's estimates of sampling error variance for the direct state estimates are overestimates, which results in a value of zero for the model error variance when the state sampling variances are used in a maximum likelihood procedure that estimates the coefficients of the predictor variables and the model error variance. The Census Bureau should investigate its procedures for estimating sampling error variance. And without waiting for the results of that work, it should also examine the effects of a simple correction, such as putting a small weight on the direct estimates in weighting the estimates from the CPS equation for a target year.

State Effects The magnitude of the state raking factors that are used to adjust the county estimates warrants further investigation. The Census Bureau should estimate the variance in the state raking factors for 1993 and 1995 to determine if their variability is consistent with sampling variation. If it is not, then research should be conducted to find an explanation for the variation. One part of this research could be to examine the effect of using 3 years rather than 1 year of CPS data in the state model, as is done in the county model.

More generally, work should be conducted to determine if there are idiosyncratic state effects that should be captured in the county model. The Census Bureau did some preliminary research on adding fixed state effects to alternative formulations of the county model (see National Research Council, 1998:App. A). While the addition of fixed state effects reduced some nonrandom residual patterns in the regression output, a fixed state effects model did not perform better than other models in comparison with the 1990 census estimates (see National Research Council, 1998:App. C and D). Some preliminary work with a random state effects model with two components of variance, one for state and one for county within state (see Fuller and Goyeneche, 1998), suggested that state effects may be present and that further research on a random state effects model should be conducted.

Discrete Variable Models that Use Counties with No Sampled Poor School-Age Children When using a logarithmic transformation of the number of poor school-age children as the dependent variable in the county regression model, all counties in the CPS sample for which none of the sampled households have poor school-age children (262 of 1,247 counties for the 1995 model) have to be removed from the regression analysis. The dropped counties are generally smaller counties with small CPS sample sizes.

While the dropped counties would have little influence in any regression equation due to their small size, the exclusion of 21 percent of the counties in the CPS sample is a cause for concern. Moreover, the internal and external evaluations of the county model suggest that although the current approach provides reasonably good estimates for small counties for 1989, 1993, and 1995, they could be improved. For example, there is a slight tendency in the county model equation to overpredict poverty in small counties (see Chapter 2). It is important to investigate the development of discrete variable regression models, such as Poisson regression or other forms of generalized linear models, that permit the inclusion of data for those counties that have no sampled families with children in poverty.

There are two factors that complicate the development of discrete variable models in this context: the lack of fully developed hierarchical models and related shrinkage procedures and the lack of methods for optimal incorporation of CPS sampling variances. However, Markov Chain Monte Carlo implementation of hierarchical models can be used to address the first issue, and, with additional research and development, can also probably address the second issue.

Ways to Reduce the Time Lag of the Estimates The Title I fund allocations for the 1999-2000 school year will be based on estimates of school-age children in 1996 who were in poor families in 1995, and these estimates will also be used for the 2000-2001 school year allocations. It is important to explore the

extent to which this time lag can be reduced for the county estimates, which will correspondingly reduce the time lag for the school district estimates.[1] The Census Bureau began some exploratory work on this topic in June 1997 but had to put it aside. Now that the county estimation procedure has been developed and put on a production basis, it is important to resume this work.

One of the causes of the lag is the availability of food stamp data, which must be obtained from individual states in some instances and which are not available until almost 2 years after the year to which they refer. It might be possible to overcome this problem, without seriously harming the performance of the county model, by using food stamp data for the year prior to the estimation year. Another possibility is to control the estimates from the county model to the state model estimates for the latest of the 3 years of CPS data used in the county model, instead of to the middle year. These ideas and others need to be evaluated to determine if the lag between the time period of the estimates and the year of allocation of funds can be reduced.

Evaluation of Food Stamp and Other Input Data Regular evaluation of the continued suitability of food stamp and other data for input to the state and county models is important for the Census Bureau's small-area estimation program. Changes in welfare programs and the accompanying data systems (especially those resulting from the 1996 Personal Responsibility and Work Opportunity Reconciliation Act) will almost certainly affect the comparability of food stamp data over geographic areas. For example, legal immigrants, many of whom are no longer eligible for benefits, are very unevenly distributed geographically. Comparability is an important assumption in both the county and state regression models, and, therefore, the way in which food stamp data are used as a predictor variable in the models may need to be modified. Changes in the tax system could also affect the usefulness of IRS data for small-area poverty estimation. More generally, it is important to continually evaluate the input data to the state and county models to assess errors or inconsistencies in them and to develop methods to account for those errors in the modeling process.

Large Category Differences and Residual Patterns for the State and County Models The internal and external evaluations (see Chapter 2) demonstrated that the state and county models are generally well behaved with respect to the estimates for various categories of states and counties. However, it is important to investigate further the residual patterns and category differences to determine if

[1]It would also be desirable to reduce the time lag in the school district boundary survey so that the allocations are made to current school districts. However, that survey is conducted every 2 years, and it may not be possible to carry it out more frequently or to complete it more quickly.

the regression models could be improved either through a modification of the model form or through the addition of predictor variables.[2]

As an example of a pattern that is worth further investigation, when compared with CPS aggregate estimates, the county model exhibited a tendency in 1989, 1993, and 1995 to underpredict the number of poor school-age children in counties with large percentages of Hispanics. Also, from examination of the standardized residuals, the state model exhibited a tendency to underpredict the proportion of poor school-age children in some states in the West Region.

More generally, as a model is estimated for additional years, it is important to look for consistent patterns of residuals and category differences to understand their causes and to take corrective action when necessary. While it may be necessary to tolerate overprediction or underprediction for a particular type of area in any one year, a consistent pattern of overprediction or underprediction needs to be addressed.

In the evaluation of residuals and category differences, particular attention should be paid to states and counties that have experienced large demographic or socioeconomic changes that may correlate with changes in numbers of poor school-age children. For example, the federal tax return data that are used to estimate internal migration for the demographic population estimates might be used to classify states and counties into categories by migration rates and the performance of the models compared for these categories. Also, the performance of the models might be compared for categories of counties classified by overall population change since the 1990 census. In turn, adding predictor variables to the models from the decennial census and the demographic estimates program, possibly including interaction terms, may prove a fruitful way to address persistent patterns of overprediction or underprediction for these and other categories of states and counties.

[2]The evaluations conducted to date of the county estimates include examination of the residual patterns from the regression model, comparisons of the model estimates for 1989 with 1990 census estimates, and comparisons of the model estimates for 1989, 1993, and 1995 with aggregate CPS estimates. Another evaluation that could help determine what portion of the errors in the county estimates is due to problems with the model–rather than measurement differences and sampling variability–is to fit the model to 1990 census data (prior to shrinkage and raking to the state model) and to compare the estimates to 1990 census values for aggregates of counties. This evaluation is similar to the county model-CPS aggregate comparisons, but it has the advantage that the sampling error in the census is much less than in the CPS. The county model estimates are not shrunk for this evaluation because the resulting estimates would have considerable weight on the census direct estimates and so be less informative about possible problems with the regression model.

School District Estimates

There cannot be marked improvements in the school district estimates without a substantial effort to improve the data sources for districts and to develop models to use them. Nonetheless, work should go forward to further evaluate the current estimation method and to seek to effect modest improvements in it. Three important areas for research are: investigation of methods to reduce the variance of the 1990 census estimates of poor school-age children; use of school enrollment data to improve estimates of the total number of school-age children; and investigation of the possible use of National School Lunch Program data to improve estimates of poor school-age children.

Reducing the Variance of the 1990 Census Estimates of Poor School-Age Children Because so many school districts are so small in size, the 1990 census estimates of poor school-age children, which derive from the long-form sample, are subject to high sampling variability. In addition to affecting the quality of the 1995 estimates that were developed by the Census Bureau's synthetic method, the sampling variability in the 1990 census estimates affects the1980-1990 evaluations. The evaluation measures reported in Chapter 3 overstate the degree of error in the synthetic estimates because of this sampling variability. The Bureau should conduct research to determine the extent of this overstatement for school districts of different sizes and compute adjusted evaluation measures in which the effect of this sampling variability is removed. A simple approach would be to use the mean square error as an evaluation measure. This measure may then be readily adjusted by subtracting out the sampling variance of the census estimates, thereby producing a more valid measure of the quality of the synthetic estimates.

The 1990 census school district estimates of poor school-age children that were used in the 1995 estimates and as the standard of comparison in the 1980-1990 evaluations were developed by ratio adjustment. This procedure, which applies the long-form-sample-based estimates of the school-age poverty rate to the complete-count estimates of total school-age children, reduces the variance of the 1990 census estimates to a modest extent. Other ways to further reduce the variance should be investigated.

One approach is to incorporate other characteristics from the census short form that are known to be related to poverty in estimating school district numbers of poor school-age children from the 1990 census. For example, such characteristics as race and ethnicity, home tenure (owner, renter), family type, and residence (e.g., central city) could be used for this purpose. A very simple form of this type of estimation procedure would be a stratified ratio adjustment with strata defined using short-form information.

Another approach is to smooth the 1990 census school district estimates with the 1990 census county estimates. By carefully constructing smoothed school-district estimates as combinations of school-district and county-level estimates, it

might be possible to produce school-district estimates with lower mean square errors than the direct 1990 census estimates. It would be desirable to make use of knowledge about model error and sampling variances at the school-district level—if available—to tailor the degree of smoothing for each school district. If successful, smoothing procedures might substantially improve the estimation of census school-age poverty rates in small school districts. They would add some bias because county poverty rates differ from poverty rates for school districts contained within them, but they could potentially substantially reduce variance, thereby improving mean square error.

The development of a smoothing approach should include a thorough evaluation. As part of that evaluation, it would be useful to compare 1990 census estimates of poor school-age children for school districts with three sets of estimates that differ in the calculation of 1980 census within-county shares that are applied to the 1989 county model estimates: unsmoothed 1980 census within-county shares (as in synthetic method (1), see Chapter 3); smoothed 1980 census within-county shares; and 1980 census within-county shares that use the 1980 census county school-age poverty rates for all school districts within each county. The third method represents a complete smoothing of the school district poverty rates within counties.

If one or both methods for reducing the variance of the 1990 census school district estimates of poor school-age children (smoothing and using other characteristics in the estimation) are successful, then the revised 1990 census estimates should be employed with the synthetic shares approach if it is used again in the future. The revised estimates should also be used as the standard of evaluation for assessing the synthetic shares estimates of poor school-age children in 1989.

Use of School Enrollment Data to Improve Estimates of the Total Number of School-Age Children The method for estimating total school-age children is similar to that for estimating poor school-age children, namely, to apply the 1990 census school district shares within each county to updated county estimates. The method is more robust for total school-age children (and total population) than for poor school-age children because the numbers being estimated are larger and because the 1990 census shares for total school-age children (and total population) are based on complete-count data that are not subject to sampling error. But the synthetic shares method still does not capture within-county changes in school district populations that have occurred since the 1990 census.

Public school enrollment data are collected annually by the National Center for Education Statistics (NCES) for school districts. Research should be conducted to determine if these data could be used to update the within-county school district shares of total school-age children. Research could begin by examining reported school enrollment in the 1980 and 1990 censuses for school districts to determine if the within-county enrollment shares in 1990, or, alternatively, the changes in enrollment from 1980 to 1990, produce estimates of total

school-age children that are more accurate for 1990 than the 1980 census-based shares. (Work is under way along these lines at the Census Bureau.) Research would also be needed to evaluate the quality of the NCES enrollment data and to determine if such factors as changes in public versus private school enrollment present a problem for estimation.

If it is determined that the use of enrollment data would improve school district estimates of total school-age children, it will be necessary to modify the estimation procedure for poor school-age children so that the estimates of both groups (total and poor) are consistent. One way to achieve consistency would be to apply 1990 census school-age poverty rates for districts to the updated estimates of within-county shares of total school-age children that are developed from enrollment data.

Possible Use of School Lunch Data to Improve Estimates of Poor School-Age Children There are many reasons that school lunch data are not necessarily a good proxy for school-age poverty (see Chapter 3). Moreover, at present, there is no complete, accurate source of school lunch data by school district that is readily available to the Census Bureau. Nonetheless, participation in the National School Lunch Program is an indicator of low income, and it seems worthwhile to pursue for other states the research that the panel undertook for New York.

The Census Bureau may be able to work through its state data centers for selected states to obtain school lunch data by district for 1989-1990 to evaluate whether within-county school lunch participation shares in 1989-1990 produce estimates of poor school-age children in 1989 that are more accurate than those produced from the 1980 census-based shares. Another approach to evaluate is whether a combination of school lunch data and census data would be preferable to using either data source alone. The research should also look at the effects of using school lunch data, solely or in combination with census data, to estimate school-age poverty rates because of the role that rates play in concentration grants. If the results of this research are promising, it would be necessary for the NCES to improve the reporting of participation in the National School Lunch Program that it collects in the Common Core of Data.

LONGER TERM PRIORITIES

State and County Models

In the longer term, research should proceed on multivariate approaches to state and county estimation that take advantage of the multiple data sources that are likely to become available in the next decade. These sources are the March CPS, the 2000 decennial census, and the monthly ACS.

Multivariate State and County Models

Use of multiple data sources (from separate surveys or multiple years of the same survey) in a system of equations can be advantageous for small-area modeling. For the state model, the Census Bureau has initiated work on a multivariate approach to incorporating the data from several years of the CPS, instead of just one year, into the regression equation (see Otto and Bell, 1997).

For the county model, the Census Bureau developed, as an alternative to the separate use of CPS and census county regression equations (with the census equation being used only to estimate the model error variance for the CPS model), a *bivariate* county regression model, in which the two dependent variables are the CPS and the previous census estimates of poor school-age children. This formulation has some very real advantages (see National Research Council, 1998:App. C). First, the internal evaluation of the regression output for the bivariate models for 1993 indicated that they are as good as or possibly better than their single-equation analogues. In addition, tests of the constancy of the parameter that distinguished between the single-equation and bivariate formulation clearly showed the benefit of the bivariate approach. Unfortunately, lack of administrative records data for 1979 prevented the Census Bureau from conducting an external evaluation of the bivariate models in comparison with the 1990 census. Therefore, given the novelty and relative lack of evaluation of these models, the panel did not recommend using them for the production of 1993 or 1995 county estimates of poor school-age children. However, research into this approach should continue, including an external evaluation as soon as that is feasible using the 2000 census data.

Similarly, integrating multiple years of the March Income Supplement of the CPS into the county estimation procedure by means of a multivariate model, as opposed to the current procedure of averaging the data for 3 years, may be advantageous. A multivariate model, with estimates from more than one CPS year and the census as dependent variables in a linear system of equations, might provide an effective way of using more of the available information. In the future this model could also incorporate data from the ACS, possibly by adding equations for the estimates from that survey.

More broadly, a wide variety of approaches that combine information over time and over geographic areas should be considered as such a combination might prove effective at modeling poverty for small areas. Because poverty very likely has commonalities over time and across areas that are similar in economic conditions, efforts to exploit this structure could prove advantageous and should be examined.

American Community Survey

The American Community Survey, when it is fully operational, will be an important component of any approach to providing updated estimates of poor

school-age children for small areas.[3] For states and counties, it is possible that several months (or years) of data from the ACS might be used to provide direct estimates of poor school-age children. Alternatively, ACS data could be used indirectly as a dependent variable in a model-based approach for state and counties, parallel to the manner in which CPS data are currently used.

However, given that each year of the CPS and the 2000 census will also provide information on poverty,[4] it will be important to find ways to use all three sources of information together, for multiple time periods (for the CPS and ACS), to produce the best state and county estimates. Furthermore, given that all three data sources will have their own measurement biases[5] and that they are available for different time periods—the decennial census year, multiple years of the CPS March Income Supplement, and many months of the ACS—it is unlikely that simply pooling estimates from the three data sources can be justified. Some adjustment or modeling procedure will be needed. Such a procedure will have to take account of available information about the variances and biases of the estimates from each data source.

Continued research and development on measurement error and time-series models will be needed to develop effective multivariate models for small-area poverty estimates that use multiple data sources for multiple time periods.[6] A specific research issue is to determine how best to use the 2000 census information, which will have lower sampling variance but possibly substantial measurement bias and which may be biased if the economic conditions during the census reference period differ markedly from the period for which estimates are needed.

In order to learn as much as possible about the measurement differences between the census, the March CPS, and the ACS, the Census Bureau should plan now for an exact match of the 2000 census with both the March 2000 CPS and the national ACS sample of about 70,000 households that will be in the field in that year. These two matches would provide a wealth of information about the three different income measurement systems. They would also provide key inputs to the development of a CPS-census measurement error model, which could help resolve some remaining issues about the state and county models. For example,

[3]The ACS, together with the 2000 census Master Address File, may also provide the means to improve small-area estimates of total population and population by age.

[4]If the ACS is implemented as planned, it is likely that the 2010 census and subsequent censuses will not include a long form and, hence, will not provide income and poverty information.

[5]The data collection methods for the census, CPS, and ACS differ in many respects, including the length of the questionnaire, the primary data collection technique (face-to-face interviews or mail questionnaires), the definitions of variables, the reference period for income measurement, and editing and imputation methods. Any of these differences can lead to different measurement biases.

[6]Measurement error models, by attempting to model effects over time and across states resulting from changes in program administration, could also be used to adjust administrative data that are used as predictor variables in estimation models for differences due to time or state effects.

some of the category differences observed in the 1990 census comparisons for the county model could be due to differences between the CPS and census measurements of poverty. A CPS-census measurement error model could also provide information from which to determine how to use data from the 2000 census with the current CPS-based estimation procedure to minimize discontinuities in the Title I fund allocations that may occur when the data from the 2000 census are incorporated into the models.

School Districts

The planned implementation of the ACS and the availability of 2000 census data hold out the prospect for markedly improved estimates of poor school-age children for school districts, as well as for states and counties. However, the availability of 2000 census and ACS data alone will not likely be sufficient to provide estimates of acceptable quality for school districts that reflect within-county as well as between-county changes in school-age poverty for districts.[7] It is likely that modeling will be necessary, and modeling, in turn, will require sources of data to serve as predictor variables. With the Master Address File that will be completed for the 2000 census, it should be possible to geocode most federal tax return data to the school district level. In fact, if a high proportion of tax return addresses can be geocoded in the near future, even before the census itself is completed, that information could be used to improve the current synthetic shares estimation method.[8] It may also be possible to undertake a federal-state cooperative effort to provide food stamp data that are geocoded to school districts.

A substantial research and development effort will be needed for improved school district estimates of poor school-age children for which work should begin now. The panel will comment further on the long-term prospects for improvement, in its final report, due at the end of 1999.

DOCUMENTATION AND EVALUATION

The development of small-area estimates of income and poverty is a major effort that includes data acquisition and review, database development, geographic mapping and geocoding of data, methodological research, model development and testing, and documentation and evaluation of procedures and out-

[7]For many school districts, data from the ACS will have to be pooled across several years to produce direct estimates of adequate precision. Because the ACS will not be fully phased in until 2003, the first 5-year pooled estimates, for example, will not be available until 2008. Moreover, such estimates will still be subject to high sampling variability for many districts, similar to the census.

[8]However, to obtain complete geocoding of these data would likely require that all tax returns be filed by the address of the residence of the tax filer rather than the address of the tax preparer.

puts. Since the production of small-area poverty estimates supports a range of important public policies for federal, state, and local governments—including the allocation of funds—it is essential that the Census Bureau have adequate staff and other resources for all components of the estimation program, including evaluation and documentation. It is the responsibility of any agency that produces model-based estimates to conduct a thorough assessment of them, including internal and external evaluations of alternative model formulations.

An integral part of the evaluation effort is the preparation of detailed documentation of the modeling procedures and evaluation results. No small-area estimates should be published without full documentation. Such documentation is needed for analysts both inside and outside the Census Bureau to judge the quality of the estimates and to identify areas for research and development to improve the estimates in future years.

Appendix
Use of School Lunch Data in New York State for the Estimation of School-Age Children in Poverty: An Analysis

James H. Wyckoff and Frank Papa

This analysis uses data from the National School Lunch Program in New York State as an alternative to census data in estimating the number of poor children (age 5-17) for use in the allocation of Title I funds to school districts. This analysis considers two uses of poverty estimates in the Title I allocations. First, for the purpose of estimating the number of school-age children who are in poor families in 1989, we compare estimates from using school lunch data for 1990 with estimates from the Census Bureau's synthetic or constant-share method that is based on 1980 census data. Second, we examine the sensitivity of various methods in estimating the 15 percent threshold for concentration grants. In conclusion, we examine some of the difficulties we encountered in attempting to use school lunch data for this purpose. Although this analysis may provide some interesting insights to some evaluation questions, it only reflects the experience in one state; other states may well differ in critical ways that would lead outcomes to change as well.

The data for this analysis cover public schools and come from the New York State Education Department Report 325 for February 1990, printed on July 10, 1992. The 325 Report is an accounting of the number of eligible applicants for free and reduced-price school lunches by school. Our data include all public school reports.[1]

[1]Some of the state's 3,279 public schools did not send reports to the New York State Education Department: most of those 389 schools did not operate a school lunch program. Those observations are treated as zeros in the analysis.

Reports from private schools are also available but they have not been included in this analysis: 804 private schools reported 42,828 free and reduced-price school lunch applicants in February 1990. This number represents about 12 percent of all school lunch applicants.

ESTIMATES OF POOR CHILDREN

The school lunch method for estimating the number of poor children in each school district is conceptually similar to the census constant-share method. County totals of poor children are allocated to specific school districts on the basis of an estimate of the ratio of poor children in the district to the county total. The school lunch ratio is computed by the ratio of free (or free and reduced-price) school lunch applicants in a school district to those in the county. This ratio is then multiplied by the total number of poor school-age children in the county (from the 1990 census) to arrive at the school district estimate. When districts cross county boundaries, the district is assigned to the county in which the school district administrative office is located.[2] In summary:

$$\tilde{Y}j' = \frac{SL_{ij}^{90}}{SL_{i.}^{90}} Cen_i^{90} \quad , \tag{1}$$

where:

$\tilde{Y}j'$ is the school lunch estimate of poor school-age children in school district j,

SL_{ij}^{90} is the number of school lunch applicants in county i, school district j in 1990,

$SL_{i.}^{90}$ is the number of school lunch applicants in county i in 1990, and

CEN_i^{90} is the 1990 census estimate of poor school-age children in county i.

The evaluation below compares these estimates of poor school-age children to those estimated using the census constant-share method, which applies the 1980 census shares of poor school-age children for school districts (or parts of school districts) within counties to the 1990 census county estimates of poor school-age children (synthetic method (2) in Chapter 3). Mean algebraic and absolute percentage errors are estimated for each method by using the 1990 census totals for school districts as "truth." Tables A-1 to A-3 summarize these results.

Table A-1 illustrates the distribution of the algebraic percentage errors, unweighted and when each district is weighted by the number of school-age

[2]We also computed estimates by employing school-level data to form county pieces when schools of a district are located in more than one county. Roughly 35 percent of the districts cross county boundaries. This estimation method produces estimates that are very close to the method that does not account for the county pieces. As a result, we present only the results that assign a whole district to the county of the district's administrative office.

TABLE A-1 Distribution of Algebraic Percentage Errors for Children Age 5-17 in Families in Poverty, Various Models, Unweighted and Weighted, New York State School Districts in Evaluation Universe, 1990 (N = 623), in percent

Distribution of Algebraic Percentage Errors	Census Constant 1980 Share	Free Lunch	Free and Reduced-Price Lunch
Unweighted			
Mean	31.2	7.1	14.0
Less than –40.0%	10.3	20.4	17.7
–40.0 to –20.0%	11.4	12.4	12.8
–19.9 to –10.0%	10.6	9.5	9.1
–9.9 to –0.1%	9.0	9.5	10.1
0.0 to 9.9%	10.6	12.0	9.8
10.0 to 19.9%	7.7	6.6	7.5
20.0 to 39.9%	11.9	11.9	10.3
40.0% and more	28.6	17.8	22.6
Weighted by Related Children Age 5-17 in Poverty, 1990 Census			
Mean	0.8	1.6	1.3
Less than –40.0%	5.0	8.1	6.8
–40.0 to –20.0%	16.0	11.0	13.7
–19.9 to –10.0%	17.0	10.0	12.1
–9.9 to –0.1%	28.6	10.8	29.8
0.0 to 9.9%	7.2	33.8	11.8
10.0 to 19.9%	8.1	8.1	5.0
20.0 to 39.9%	8.1	7.9	8.9
40% and more	10.1	10.3	11.9

NOTES: The census constant 1980 share estimates are calculated as described in Chapter 3 (synthetic method (2)). The school lunch estimates are formed by multiplying the 1990 census estimates of related children age 5-17 in families in poverty for the county by the school district's share of the county's free (free and reduced-price) lunch participants. The mean unweighted algebraic percentage error is the sum over all school districts of the algebraic difference between the estimate of poor school-age children from a model and the 1990 census estimate as a proportion of the census estimate for each district, divided by the number of districts. The weighted mean weights each difference by the census number of poor school-age children in the district.

TABLE A-2 Mean Absolute and Algebraic Percentage Errors for Children Age 5-17 in Families in Poverty, Various Methods, New York State School Districts in Evaluation Universe, 1990, Unweighted, in percent

Category	Percent of Districts (N = 623)	Census Constant 1980 Share	
		Mean Absolute % Error	Mean Algebraic % Error
Total	100.0	53.4	31.2
1990 School District Population			
Under 2,500	11.9	66.3	34.0
2,500-4,999	14.3	41.2	15.8
5,000-7,499	17.5	57.7	32.6
7,500-9,999	10.8	58.7	28.3
10,000-14,999	12.5	61.3	45.1
15,000-19,999	9.5	43.5	29.8
20,000-29,999	10.8	67.2	55.8
30,000-39,999	5.3	36.5	11.6
40,000-49,999	2.9	42.6	25.1
50,000-99,999	3.9	24.9	10.7
100,000 or more	0.8	12.0	–12.0
1980-1990 Population Growth			
Decrease of 10.0% or more	3.9	45.5	30.6
Decrease of 5.0-9.9%	12.0	54.7	34.9
Decrease of 0.1-4.9%	24.4	48.1	27.1
Increase of 0.0-4.9%	21.8	50.6	28.1
Increase of 5.0-9.9%	15.9	58.2	35.8
Increase of 10.0% or more	22.0	59.4	33.4
Percentage Poor School-Age Children, 1990			
0.0%	2.3	0.0	0.0
0.1%-5.9%	34.2	97.7	83.6
6.0-8.9%	16.1	42.1	20.1
9.0-12.4%	17.0	33.1	8.7
12.5-16.4%	15.1	26.4	3.0
16.5-23.9%	11.9	23.6	–15.6
24.0% or more	3.5	24.6	–20.6

Free Lunch		Free and Reduced-Price Lunch	
Mean Absolute % Error	Mean Algebraic % Error	Mean Absolute % Error	Mean Algebraic % Error
48.7	7.1	52.1	14.0
57.4	4.6	59.8	10.1
40.1	–0.4	41.9	4.2
65.8	30.4	71.2	38.8
47.2	–12.1	45.7	–10.1
53.2	22.1	60.4	31.3
39.3	–6.3	43.1	1.3
47.3	14.1	54.3	27.8
37.2	–9.0	38.3	–3.3
36.7	–14.4	36.9	–8.0
24.2	–7.0	26.5	–2.9
5.1	5.1	4.7	–0.9
23.4	–0.9	31.8	10.0
63.0	7.0	67.5	11.0
39.3	–10.3	40.4	–4.9
47.9	20.4	51.2	27.7
43.0	12.0	47.2	20.1
60.6	11.3	64.6	19.4
0.0	0.0	0.0	0.0
81.4	22.2	90.2	37.6
41.9	2.7	44.8	8.9
41.0	6.8	41.7	10.6
22.6	–2.0	22.0	0.8
24.3	–10.4	23.4	–12.3
24.2	–16.3	24.5	–21.4

TABLE A-2 Continued

Category	Percent of Districts (N = 623)	Census Constant 1980 Share: Mean Absolute % Error	Census Constant 1980 Share: Mean Algebraic % Error
Change in Poverty Rates for Children, 1980-1990			
Decrease of 10.0% or more	4.5	132.1	129.3
Decrease of 5.0-9.9%	11.9	95.9	93.4
Decrease of 0.1-4.9%	46.1	55.8	50.2
Increase of 0.0-4.9%	29.2	23.9	–18.7
Increase of 5.0-9.9%	7.1	37.6	–37.1
Increase of 10.0% or more	1.3	59.1	–59.1
Percent of Population Black, 1990			
0.0-0.9%	15.1	29.9	9.9
1.0-4.9%	36.9	48.6	23.9
5.0-9.9%	34.7	64.5	42.0
10.0-24.9%	13.3	64.5	47.1
Percent of Population Hispanic, 1990			
0.0-0.9%	22.6	38.4	16.0
1.0-4.9%	49.3	48.8	28.4
5.0-9.9%	28.1	73.7	48.3

NOTES: The census constant 1980 share estimates are calculated as described in Chapter 3 (synthetic method (2)). The school lunch estimates are formed by multiplying the 1990 census estimates of related children age 5-17 in families in poverty for the county by the school district's share of the county's free (free and reduced-price) lunch participants. The mean unweighted absolute (algebraic)

children from families in poverty. Each of the methods results in estimates with some very large errors. For example, consider the weighted results. All three methods have at least 15 percent of the districts with errors of at least 40 percent. This pattern is also illustrated in Table A-2, which shows unweighted estimates broken down by various school district characteristics. Regardless of method, the errors are very large on average and in most categories.

Weighting by the number of poor school-age children in 1990 substantially reduces the percentage errors across all methods, as shown in Table A-3. This approach yields results that are quite similar across all three models. Mean

Free Lunch		Free and Reduced-Price Lunch	
Mean Absolute % Error	Mean Algebraic % Error	Mean Absolute % Error	Mean Algebraic % Error
101.8	48.1	103.3	51.8
64.5	47.5	73.4	57.2
53.4	11.5	57.0	20.5
33.8	–13.5	34.5	–8.6
26.4	–23.1	26.1	–21.5
44.2	–36.2	43.9	–42.0
33.3	5.3	34.8	9.8
46.7	17.2	50.0	22.0
54.0	0.2	57.5	8.2
57.8	0.9	64.2	11.7
41.3	13.8	45.3	19.2
40.8	9.9	44.1	18.4
68.5	–3.2	71.5	2.0

percentage error is the sum over all school districts of the absolute (algebraic or signed) difference between the estimate of poor school-age children from a model and the 1990 census estimate as a proportion of the census estimate for each district, divided by the number of districts.

algebraic percentage errors are relatively small; however, as one would expect, mean absolute percentage errors are much larger. Most of the patterns of errors with respect to school district attributes are as would be expected. For example, school districts with small total population have larger errors than districts with larger populations.

An important result of this analysis is that even after some effort in data preparation, the school lunch method is still not meaningfully better than the census constant-share method. At least in New York State it does not appear that

TABLE A-3 Mean Absolute and Algebraic Percentage Errors for Children Age 5-17 in Families in Poverty, Various Methods, New York State School Districts in Evaluation Universe, 1990, Weighted by Children Age 5-17 in Poverty, 1990 Census, in percent

Category	Percent of Districts (N = 623)	Census Constant 1980 Share	
		Mean Absolute % Error	Mean Algebraic % Error
Total	100.0	23.9	0.8
1990 School District Population			
Under 2,500	11.9	43.4	13.8
2,500-4,999	14.3	30.4	4.2
5,000-7,499	17.5	31.6	–0.1
7,500-9,999	10.8	32.5	–4.4
10,000-14,999	12.5	34.8	13.2
15,000-19,999	9.5	21.6	4.0
20,000-29,999	10.8	37.8	21.4
30,000-39,999	5.3	31.5	–2.0
40,000-49,999	2.9	33.6	9.4
50,000-99,999	3.9	18.3	–0.4
100,000 or more	0.8	10.4	–10.4
1980-1990 Population Growth			
Decrease of 10.0% or more	3.9	31.2	26.2
Decrease of 5.0-9.9%	12.0	13.7	–4.2
Decrease of 0.1-4.9%	24.4	20.8	–3.7
Increase of 0.0-4.9%	21.8	31.1	9.9
Increase of 5.0-9.9%	15.9	30.1	2.2
Increase of 10.0% or more	22.0	32.4	2.9
Percentage of Poor School-Age Children, 1990			
0.0%	2.3	0.0	0.0
0.1-5.9%	34.2	53.4	40.3
6.0-8.9%	16.1	34.0	9.6
9.0-12.4%	17.0	22.2	4.5
12.5-16.4%	15.1	22.7	–2.1
16.5-23.9%	11.9	19.0	–14.8
24.0% or more	3.5	11.3	–10.1

Free Lunch		Free and Reduced-Price Lunch	
Mean Absolute % Error	Mean Algebraic % Error	Mean Absolute % Error	Mean Algebraic % Error
22.3	1.6	24.2	1.3
38.5	–7.4	39.2	–4.5
31.6	–11.2	31.8	–7.3
34.2	1.3	34.9	3.9
32.6	–4.6	30.0	–4.5
32.0	6.8	35.6	10.9
24.9	0.9	28.7	5.3
36.3	13.8	39.7	21.1
27.9	3.1	28.1	3.4
34.0	–6.8	35.6	–6.3
21.0	–1.5	24.0	–3.2
3.4	3.4	5.3	–3.0
9.8	2.2	24.2	7.7
10.2	–3.8	17.0	9.8
17.2	1.9	17.1	1.1
32.5	8.6	33.1	11.5
29.9	0.3	30.6	3.3
39.0	2.0	38.0	4.6
0.0	0.0	0.0	0.0
47.1	–5.3	52.1	8.7
34.3	2.8	36.7	8.6
36.3	19.5	38.5	23.5
19.6	–1.6	20.0	–2.3
18.2	1.4	16.2	–3.3
5.4	–0.6	9.1	–7.8

TABLE A-3 Continued

Category	Percent of Districts (N = 623)	Census Constant 1980 Share: Mean Absolute % Error	Census Constant 1980 Share: Mean Algebraic % Error
Change in Poverty Rates for Children, 1980-1990			
Decrease of 10.0% or more	4.5	79.3	75.3
Decrease of 5.0-9.9%	11.9	47.0	38.1
Decrease of 0.1-4.9%	46.1	32.1	23.1
Increase of 0.0-4.9%	29.2	19.0	–15.0
Increase of 5.0-9.9%	7.1	16.9	–15.7
Increase of 10.0% or more	1.3	10.6	–10.6
Percent of Population Black, 1990			
0.0-0.9%	15.1	21.4	–0.5
1.0-4.9%	36.9	24.4	0.5
5.0-9.9%	34.7	30.1	2.4
10.0-24.9%	13.3	16.9	–0.6
Percent of Population Hispanic, 1990			
0.0-0.9%	22.6	22.7	0.7
1.0-4.9%	49.3	20.1	–0.5
5.0-9.9%	28.1	34.8	4.1

NOTE: See notes to Table A-2.

using school lunch data results in significant gains in estimating school-age children from poor families.

ESTIMATES OF THE CONCENTRATION GRANT THRESHOLD

Eligibility for Title I concentration grants is based on having a school-age poverty rate of at least 15 percent or at least 6,500 poor children.[3] Current Title

[3]Children eligible for Title I are not limited to school-age children from poor families (see Chapter 1). However, for the purpose of this analysis, which is to examine the census constant-share estimates of school-age children from poor families, eligibility is so characterized.

Free Lunch		Free and Reduced-Price Lunch	
Mean Absolute % Error	Mean Algebraic % Error	Mean Absolute % Error	Mean Algebraic % Error
64.9	19.5	68.2	23.5
41.4	31.7	43.1	32.4
36.0	7.1	37.9	12.0
20.9	–4.8	21.1	–3.6
8.5	–5.2	12.0	–10.5
5.5	–1.2	7.8	–7.7
22.4	–2.6	23.8	–0.9
23.6	0.5	24.6	0.0
32.0	4.6	32.1	4.1
9.8	0.0	14.5	–0.3
23.5	0.0	25.5	0.7
15.0	0.5	18.1	0.0
40.8	5.4	39.3	5.2

I allocations employ a two-stage eligibility criterion. A district must be in a county that meets the 15 percent (or 6,500) rule, and the district itself must meet that criterion. Under the proposed direct allocation system, grants will be made directly to districts and, as such, eligibility will be determined solely with regard to district poverty rates, without regard to county poverty rates. The proposed direct allocation method also permits states to aggregate the allocations to districts that have total population of less than 20,000 and reallocate this total based on alternative data, such as those from the National School Lunch Program. It is of interest to examine eligibility for concentration grants in those districts with less than 20,000 population under three different scenarios: the current two-stage process, the direct allocation process to districts without controls, and direct

allocations when school district poverty estimates must sum to the census county totals. We examine how concentration grants eligibility differs under these circumstances when school lunch data are used rather than census constant-share estimates, using 1990 census ratio-adjusted counts as the measure of truth.

Of the 623 districts in New York State that are in the census evaluation universe, 476 are in districts that had less than 20,000 total population in 1990. As shown in Table A-4, these 476 districts represent 76 percent of all districts in the evaluation universe for New York, but they contain only 35 percent of the poor children age 5-17 in the census evaluation universe.

Tables A-5 to A-8 examine estimates of the number of districts and percentage of school-age children who are in poor families under alternative estimation methods in 1990. The census counts are the ratio-adjusted estimates of school-age children who are in poor families from the 1990 census. Census-based estimates (synthetic method (2) estimates) use the 1990 census counts of county school-age children who are in poor families and allocate these totals to school districts by the school district's share of county totals from the 1980 census. The model-based estimates (synthetic method (1) estimates) use a similar approach, but with the county estimates of school-age children who are in poor families in 1989 produced from the Census Bureau's county model. The school lunch estimates are produced, as outlined above, by using 1990 county ratio-adjusted estimates of school-age children who are in poor families from the 1990 census and allocating them to constituent school districts by the share of that school district's free (or free and reduced-price) school lunch eligibles relative to the county total.

Tables A-5 and A-6 provide estimates for the two-tier concentration grant eligibility for districts with total population (from the 1990 census) of less than 20,000. That is, districts must be in counties where at least 15 percent (or 6,500) of the school-age children are poor and in a district that also meets this criterion.[4] If we take the census counts as our measure of "truth," then employing school lunch data will likely overstate eligibility. As shown in Table A-5, roughly 50 percent more districts and school-age children are estimated to be eligible with free school lunch data than with the census counts. This problem is further magnified when the free and reduced-price lunch counts are employed. Table A-6 illustrates where each method errs relative to the eligibility categorization of the census counts: as might be expected, the school lunch estimates produce a substantial number of false positives.

Tables A-7 and A-8 provide a similar analysis for direct allocations. Now districts must only meet the single criterion that the district has at least 15 percent

[4]In Tables A-5 and A-6, county eligibility is determined by the county counts from the 1990 census. Within each of these eligible counties, the alternative methods listed are used to determine school district eligibility.

TABLE A-4 New York State Districts in Evaluation Universe Above and Below the 20,000 Population Threshold for Pooling Allocations (N = 623)

	Districts		Poor Children Age 5-17	
Category	Number	Percent	Number	Percent
School District Total Population				
Less than 20,000	476	76.4	61,236	35.0
At least 20,000	147	23.6	113,556	65.0

TABLE A-5 Concentration Grant Eligibility at County and School District Level, Various Methods for New York State Districts in Evaluation Universe with Less than 20,000 Population, 1990 (N = 476)

	Districts		Poor Children Age 5-17	
Method	Number	Percent	Number	Percent
Census Counts	76	16.0	16,689	27.3
Census-based Estimates	78	16.4	14,162	23.1
Model-based Estimates	76	16.0	14,134	23.1
Free Lunch[a]	112	23.5	21,662	35.4
Free and Reduced-price Lunch[a]	136	28.6	24,515	40.0

NOTES: Cell entries are for school districts and poor school-age children that would be eligible for concentration grants according to various methods (see text) under the current two-stage allocation process (i.e., both county and school district have more than 6,500 or more than 15% poor school-age children). The total number of poor school-age children in districts with less than 20,000 population is 61,236.

[a]Some school districts (54 or 11.3%) did not report school lunch data.

TABLE A-6 Concentration Grant Eligibility at County and School District Level, Census Counts Compared to Various Other Methods for New York State Districts in Evaluation Universe with less than 20,000 Population, 1990 (N = 476)

	Census Not Eligible				Census Eligible			
	Estimate Not Eligible		Estimate Eligible		Estimate Not Eligible		Estimate Eligible	
Method	% Districts	% Poor Children Age 5-17	% Districts	% Poor Children Age 5-17	% Districts	% Poor Children Age 5-17	% Districts	% Poor Children Age 5-17
Census-based Estimates	78.6	68.0	5.5	4.8	5.0	8.9	10.9	18.4
Model-based Estimates	78.8	68.1	5.3	4.6	5.3	8.8	10.7	18.4
Free Lunch[a]	75.4	63.6	8.6	9.2	1.1	1.0	14.9	26.2
Free and Reduced-price Lunch[a]	71.2	59.8	12.8	13.0	0.2	0.2	15.8	27.1

NOTE: See notes to Table A-5.

[a]Some school districts (54 or 11.3%) did not report school lunch data.

TABLE A-7 Concentration Grant Eligibility at School District Level, Various Methods for New York State Districts in Evaluation Universe with Less than 20,000 Population, 1990 (N = 476)

	Districts		Poor Children Age 5-17	
Method	Number	Percent	Number	Percent
Census Counts	115	24.2	25,343	41.4
Census-based Estimates	114	24.0	19,596	32.0
Model-based Estimates	109	22.9	18,285	29.9
Free Lunch[a]	214	45.0	39,222	64.1
Free and Reduced-price Lunch[a]	294	61.8	48,835	79.8
Free Lunch with Controls[a,b]	124	26.1	25,024	40.9
Free and Reduced-price Lunch with Controls[a,b]	127	26.7	24,045	39.3

NOTES: Cell entries are for school districts and poor school-age children that would be eligible for concentration grants according to various methods (see text) under a direct allocation process (i.e., the school district has more than 6,500 or more than 15% poor school-age children). The total number of poor school-age children in districts with less than 20,000 population is 61,236.

[a]Some school districts (54 or 11.3%) did not report school lunch data.

[b]Controls are imposed at the county level so that number of poor children and number of children in the school district must sum to county census counts for 1990.

of its school-age children who are poor (or at least a total of 6,500). These estimates also show the effect of imposing county controls on the use of school lunch estimates. (The county controls are equivalent to the estimates produced by equation 1, above.) The school lunch estimates without controls greatly overstate concentration grant eligibility. Imposing county controls substantially improves the accuracy of these estimates.

Table A-9 shows mean algebraic and absolute percentage errors for the various estimation methods. Here the school lunch estimates have either been controlled to the statewide total of school-age children living in poor families for the 476 districts with populations of less than 20,000 or to a similar county total. With these controls in place, each of the methods has roughly the same algebraic and absolute percentage errors. This result is interesting as the school lunch estimates with county controls had the potential to be either better or worse than the estimates with state controls. We would in general expect them to be better as there is a tighter level of control imposed. It is possible that they are worse as a result of lack of precision that occurs when school districts cross county boundaries and school lunch data are coded to the county where the district office is located.

TABLE A-8 Concentration Grant Eligibility at School District Level, Census Counts Compared to Various Other Methods for New York State Districts in Evaluation Universe with less than 20,000 Population, 1990 (N = 476)

	Census Not Eligible				Census Eligible			
	Estimate Not Eligible		Estimate Eligible		Estimate Not Eligible		Estimate Eligible	
Method	% Districts	% Poor Children Age 5-17	% Districts	% Poor Children Age 5-17	% Districts	% Poor Children Age 5-17	% Districts	% Poor Children Age 5-17
Census-based Estimates	65.3	49.5	10.5	9.1	10.7	18.5	13.5	22.9
Model-based Estimates	66.6	52.1	9.2	6.5	10.5	18.1	13.7	23.3
Free Lunch[a]	52.5	34.0	23.3	24.7	2.5	2.0	21.6	39.4
Free and Reduced-price Lunch[a]	37.0	19.7	38.9	38.9	1.3	0.6	22.9	40.8
Free Lunch with Controls[a,b]	66.8	50.0	9.0	8.7	7.1	9.2	17.0	32.2
Free and Reduced-price Lunch with Controls[a,b]	65.6	49.8	10.3	8.9	7.8	11.0	16.4	30.4

NOTE: See notes to Table A-7.

[a]Some school districts (54 or 11.3%) did not report school lunch data.

[b]Controls are imposed at the county level so that number of poor children and number of children in the school district must sum to county census counts for 1990.

TABLE A-9 Mean Absolute and Algebraic Percentage Errors for Children Age 5-17 in Families in Poverty, Various Methods, Unweighted and Weighted, New York State School Districts in Evaluation Universe with Less than 20,000 Population, 1990 (N = 476)

	Unweighted		Weighted by Poor School-Age Children	
Method	Mean Absolute % Error	Mean Algebraic % Error	Mean Absolute % Error	Mean Algebraic % Error
Census-based Estimates	54.9	30.8	30.6	4.3
Model-based Estimates	54.3	27.2	31.5	0.1
Free Lunch[a]	52.9	9.8	32.3	–0.1
Free and Reduced-price Lunch[a]	55.1	13.4	32.6	–0.1
Free Lunch with Controls[b]	54.6	12.6	31.7	–0.1
Free and Reduced-price Lunch with Controls[b]	55.7	13.3	31.7	–0.1

NOTES: See text for the calculation of poor school-age children by each method. The mean unweighted algebraic percentage error is the sum over all school districts of the algebraic difference between the estimate of poor school-age children from a model and the 1990 census estimate as a proportion of the census estimate for each district, divided by the number of districts. The weighted mean weights each difference by the census number of poor school-age children in the district.

[a]Controls are imposed at the state level so that the number of poor children and number of children in the school district must sum to the state census count for 1990 for districts with less than 20,000 population.

[b]Controls are imposed at the county level so that number of poor children and number of children in the school district must sum to county census counts for 1990 for districts with less than 20,000 population.

PROBLEMS WITH THIS APPROACH

Using school lunch data to estimate the number of poor children for each school district has several potential problems, based on the experience in New York State. As has been widely acknowledged:

- Participants in the National School Lunch Program are not the target population of Title I:
    - There are differences in eligibility between Title I and school lunch.
    - There are differences in reporting geography: Title I counts residents, school lunch counts by location of the school the child attends.
- Not all eligible students apply for the school lunch program, and application rates appear to be uneven across schools.
- Some schools choose not to participate in the school lunch program.

Other difficulties include:

- New York State has a number of regional (groups of counties) educational authorities with students, and they participate in the school lunch program; how to allocate these students is an issue.
- In New York State, the school lunch program is administered separately from most other programs, which can make use of the administrative data difficult (e.g., schools sometimes have separate identification numbers, which makes matching to other data very time consuming).

References and Bibliography

NOTE: Many of the cited reports and papers by Census Bureau staff are available on the Census Bureau's web site: http://www.census.gov. Other reports related to small-area income and poverty estimates may be obtained by contacting: David Waddington, Division of Housing and Household Economic Statistics, U.S. Bureau of the Census, Washington, D.C. 20233.

Papers that were prepared for the panel—marked by an asterisk (*)—are available from the Committee on National Statistics, 2101 Constitution Ave. NW, Washington, D.C. 20418.

Abelson, R.P., and J.W. Tukey
1963 Efficient utilization of nonnumerical information in quantitative analysis: General theory and the case of simple order. *Annals of Mathematical Statistics* 34:1347-1369.

Alexander, C.H.
1998 Recent Developments in the American Community Survey. Paper prepared for presentation at the Joint Statistical Meetings, Dallas. Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (August).

Alexander, C.H., S. Dahl, and L. Weidman
1997 Making Estimates from the American Community Survey. Paper prepared for presentation at the Joint Statistical Meetings, Anaheim, Calif. Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (August).

Bell, W.R.
1997a A County CPS Model with "Census Residuals." Statistical Research Division, Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (July).
1997b Regression Diagnostics for Models for County Poverty Estimates–Complete. Statistical Research Division, Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (September).

Belsley, D.A., E. Kuh, and R.E. Welsch
1980 *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley.

Bureau of the Census
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.
1993 *Poverty in The United States: 1992*. Current Population Report, Series P60-185. Washington, D.C.: U.S. Department of Commerce.
1995 *Subnational Estimates of Total Population by the Tax Return Method*. Population Estimates Branch, Population Division. Washington, D.C.: U.S. Department of Commerce.
1998 1995 SAIPE State and County Models. Draft documentation. By P. Cardiff, R. Fisher, P. Siegel, A. Strand, D. Waddington, W.R. Bell, and M. Kramer, with assistance from R.E. Fay and G. Train. Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (October).
1999 Estimates of Poor and Population for School Districts: School Year 1995-1996. Draft documentation. By E. Miller, M. Kramer, A. Strand, R. Fisher, W.R. Bell, and P. Siegel. Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (January).

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, Washington, D.C. (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.

Fay, R.E.
1996 Methodology for Poverty Estimation for 1993. Section 2. Demographic Statistical Methods Division, Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (November).
1997a Comparisons of CPS and Census Poverty Levels. Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (October).
1997b Notes on Direct Variance Estimates for Counties. Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (September).

Fay, R.E., and R.A. Herriot
1979 Estimates of income for small places: An empirical Bayes application of James-Stein procedures to census data. *Journal of the American Statistical Association* 78:269-277.

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.

Fay, R.E., and G.F. Train
1995 Aspects of survey and model-based postcensal estimation of income and poverty characteristics for states and counties. *American Statistical Association Proceedings of the Section on Government Statistics*. Alexandria, Va: American Statistical Association.
1997 Small domain methodology for estimating income and poverty characteristics for states in 1993. *American Statistical Association Proceedings of the Social Statistics Section*. Alexandria, Va: American Statistical Association.

Fisher, R., and P. Siegel
1997 Methods used for small area poverty and income estimation. *American Statistical Association Proceedings of the Social Statistics Section*. Alexandria, Va: American Statistical Association.

Fuller, W.A., and J.J. Goyeneche
1998 *Estimation of the State Variance Component. Working Paper. Statistical Laboratory, Iowa State University, Ames. (June).

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.

National Research Council
1995 *Measuring Poverty: A New Approach*. Constance F. Citro and Robert T. Michael, eds. Panel on Poverty and Family Assistance: Concepts, Information Needs, and Measurement Methods, Committee on National Statistics. Washington, D.C.: National Academy Press.
1997 *Small-Area Estimates of Children in Poverty, Interim Report 1, Evaluation of 1993 County Estimates for Title I Allocations*. Constance F. Citro, Michael L. Cohen, Graham Kalton, and Kirsten K. West, eds. Panel on Estimates of Poverty for Small Geographic Areas, Committee on National Statistics. Washington, D.C.: National Academy of Press.
1998 *Small-Area Estimates of Children in Poverty, Interim Report 2, Evaluation of Revised 1993 County Estimates for Title I Allocations*. Constance F. Citro, Michael L. Cohen, and Graham Kalton, eds. Panel on Estimates of Poverty for Small Geographic Areas, Committee on National Statistics. Washington, D.C.: National Academy of Press.

Otto, M.C., and W.R. Bell
1995 Sampling error modeling of poverty and income statistics for states. *American Statistical Association Proceedings of the Section on Government Statistics*. Alexandria, Va.: American Statistical Association.
1997 Bayesian Inference About Poverty and Income for States. Paper prepared for presentation at the Joint Statistical Meetings, Anaheim, Calif. Bureau of the Census, U.S. Department of Commerce, Washington, D.C.

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

Prasad, N.G.N., and J.N.K. Rao
1990 The estimation of mean squared errors of small area estimators. *Journal of the American Statistical Association* 78:47-59.

Reed, J.M.
1996 Puerto Rico Estimates and Projections Based on Demographic Component Method: 1990-1995. International Programs Center, Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (November).

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.

Santos, R., and D. Waddington
1999 *Obtaining Estimates of Child Poverty in Puerto Rico: An Overview of Current Practices and Recommendations for Improvement. Report prepared for the Panel on Estimates of

Poverty for Small Geographic Areas, Committee on National Statistics, National Research Council. National Opinion Research Center, Chicago, Ill., and Bureau of the Census, U.S. Department of Commerce, Washington, D.C.

Siegel, P.
1997 Background on Poverty Estimates for School Districts [Tables]. Housing and Household Economic Statistics Division, Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (May).

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

Spencer, B.D., and Che-Fu 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.

Thiel, H.
1971 *Principles of Econometrics*. New York: John Wiley and Sons.

Voss, P.R., R. Gibson, and K. Morgen
1997 *Assessment of the Initial 1993 Estimates of School Age Children in Poverty. Report prepared for the Panel on Estimates of Poverty for Small Geographic Areas, Committee on National Statistics, National Research Council. Department of Rural Sociology, University of Wisconsin–Madison.

Zaslavsky, A., and A. Schirm
1998 *Interactions between the American Community Survey and Federal Funding Formulae: A Look Ahead. Paper prepared for the American Community Survey Workshop. Committee on National Statistics, National Research Council, Washington, D.C. (September).

Biographical Sketches of Panel Members and Staff

GRAHAM KALTON (*Chair*) is a senior statistician and senior vice president of Westat. 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 president 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 Dame. His previous positions have been as a visiting scholar at the Joint Center for Poverty Research of the University of Chicago and Northwestern University, 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 senior program officer for 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 assisting the Panel on Alternate Census Methodologies and the Panel on Estimates of Poverty for Small Geographic Areas. He previously directed 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, Princeton University. His general area of research is the use of statistics in public policy, with particular interest in census undercount, model validation, and robust estimation. He received a B.S. in mathematics from the University of Michigan and an M.S. and a Ph.D. 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 the School of 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 chairman of the Design and Analysis Committee of the National Assessment of Educational Progress (NAEP). She has served as a member of the Technical Advisory Committee for the National Adult Literacy Survey and the Graduate Record Examination and 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. She is a member of the American Educational Research Association and the National Council on Measurement in Education. In addition, she has served on the editorial boards of the *Educational Researcher* and the *Journal of Applied Measurment in Education*. 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 a lecturer in the School of Public Health at the University of California, Los Angeles. Previously she taught at the RAND Graduate School of Public Policy Studies and 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 editor of *Statistical Science* and as an associate editor for the *Journal of the American Statistical Association*, is past 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. 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 the 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 wel-

fare, 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.

MICHELE VER PLOEG, a research associate for the panel, is also serving as study director for the Panel on Data and Methods for Measuring the Effects of Changes in Social Welfare Programs of the Committee on National Statistics. Her research interests include the effects of social policies on families and children, the outcomes of children who experience poverty and changes in family composition, and individuals' education attainment choices. She received a B.A. in economics from Central College and M.S. and Ph.D. degrees in consumer economics and housing from Cornell University.

PAUL R. VOSS is professor 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.

JAMES H. 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.