

SMALL-AREA INCOME AND POVERTY ESTIMATES: PRIORITIES FOR 2000 AND BEYOND

Constance F. Citro, National Research Council; Graham Kalton, Westat

Constance F. Citro, National Research Council, 2101 Constitution Ave., N.W. Washington, D.C. 20418

Key Words: Income, Poverty, Small-Area Estimation

CONCLUSION

Recent trends in federal policy for social and economic programs have increased the demand for regularly updated small-area estimates of income and poverty. More than \$130 billion of federal funds are allocated each year to states and localities by means of formulas that include such estimates, and the estimates are used for program evaluation and other purposes as well. States also use small-area income and poverty estimates to allocate their own and federal funds to substate areas. The funds support a wide range of activities and services, including child care, community development, education, job training, nutrition, public health, and others.

The newest source of small-area income and poverty estimates is the Census Bureau's Small Area Income and Poverty Estimates (SAIPE) Program, which was begun in the early 1990s to provide estimates that would be more timely than those from the decennial census. The 1994 "Improving America's Schools Act" called for the use of the SAIPE estimates of poor school-age children for counties and school districts to allocate more than \$7 billion annually for programs for educationally disadvantaged children under Title I of the Elementary and Secondary Education Act. The 1994 act also required a panel of the Committee on National Statistics at the National Research Council to determine if the estimates were sufficiently reliable for Title I allocations and to make recommendations for their use and future development. (The panel members and staff are listed at the end of the article.)

The first state and county SAIPE estimates were issued in early 1997 (for income year 1993); they included estimates of median household income, numbers of poor people, poor children under age 5 (states only), poor children aged 5-17 in families, and poor people under age 18. The estimates in early 1999 (for income year 1995) also included the numbers of poor school-age children in families for more than 14,000 school districts. The U.S. Department of Education has used the SAIPE estimates for Title I allocations since 1997, and some other programs use them as well.

Because there is no one data source that can provide the SAIPE estimates, the Census Bureau develops them by using statistical modeling techniques that combine data from household surveys, the decennial census, and administrative records. The SAIPE estimates, consequently, are "indirect," and, as such, their quality depends on the choice of a suitable statistical model.

In the coming decade, it should be possible to develop more accurate and timely income and poverty estimates for small areas by using new and improved sources of data from household surveys and administrative records. However, none of the existing or planned surveys or administrative records sources can, by itself, provide direct estimates of sufficient reliability, timeliness, and quality of responses for all of the SAIPE income and poverty estimates. Therefore, the panel concludes that the SAIPE Program must continue to rely primarily on models that combine data from more than one source to produce indirect estimates.

USING ESTIMATES IN PROGRAMS

The use of small-area income and poverty estimates for allocating funds or related program purposes imposes significant requirements if the estimates are to satisfy the intentions of program legislation. Such requirements include the desired concept or definition of poverty or income measured, the level of geographic detail, the level of population or demographic detail, the timeliness of production and updating, and the accuracy of measurement. The selection of a set of estimates to use in a given program will generally involve tradeoffs among competing goals. For fund allocation, it is important to consider features of the specific allocation formula, some of which may be sensitive to the level of accuracy in the estimates. For example, if a formula has a threshold for eligibility for funding, an area that is erroneously estimated to be below the threshold will not receive any funds, even if the degree of underestimation is small.

For program use, policy makers should consider the advantages and disadvantages of alternative sources of income and poverty estimates and choose estimates that are most in accord with program goals. Data from the decennial census have the advantage of providing small-area poverty estimates on the basis of a very large household survey (the sample of households that receives the census long form), but the census estimates are only available every 10 years. For very small areas, they also have considerable error due to sampling variability. In contrast, administrative records sources of poverty estimates, such as school district counts of children approved for free or reduced-price lunches under the National School Lunch Program, are timely and not subject to sampling error. However, they may not relate

in a consistent manner to poverty across areas because income eligibility guidelines for programs often differ from the poverty thresholds, and program participation may vary substantially across areas.

Evaluations of the SAIPE poverty estimates found them to be a marked improvement over outdated census estimates for states and counties and at least as good as, if not better than, other estimates that were being used for school district allocations. However, the evaluations also found that the level of inaccuracy in the estimates could be sizable, particularly for small school districts.

RECOMMENDATIONS FOR PRODUCERS AND USERS

The SAIPE Program indirect estimates of income and poverty, which use official concepts and are updated on a regular basis, are likely to become more widely used for fund allocation and other program purposes in the future. We recommend practices that we believe are critically important for the SAIPE Program in the production of estimates and that are important for users to follow in applying estimates for program use.

Producers

The producing agency for a program of model-dependent estimates, such as SAIPE, should, first of all, have adequate staff and other resources for all the component operations. The producing agency should also:

- o maintain regular contact with key users, so that the estimation program produces those estimates that are most needed and appropriate for important program uses within the constraints of available resources;
- o as a matter of routine practice every time a new round of estimates is prepared, check the input data for errors and assess each data source for its continued suitability for use in estimation models;
- o search for possible new data sources whose use might lead to improved estimates and consider pilot efforts as appropriate to establish their value for use in models;
- o pursue efforts, such as reducing the lag in availability of key data sources, to reduce the time between release of estimates and the year to which they refer;
- o carry out research and development on methods that may improve the estimates in terms of their variability, bias, and timeliness;
- o thoroughly evaluate the estimates every time they are produced, by conducting internal evaluations of

the model outputs and, to the extent possible, external evaluations with other data sources; and

o document the evaluations and results in detail, make the documentation available to users, and provide research access to the input data and models to permit independent replication and evaluation, taking care to address confidentiality concerns.

Users

Agencies that use estimates for fund allocation or other program purposes should:

- o carefully review the documentation provided by the producing agency to understand the properties of the estimates;
- o periodically obtain independent reviews of the estimates and alternatives to them; and
- o regularly study the effects of using the estimates for the allocations made by the agency and, where appropriate, for suballocations made by others.

Policy makers need to have information about the effects of alternative estimates and formula provisions to consider in making decisions about program uses of estimates. For this purpose, we urge that policy makers:

- o commission assessments of formulas and the estimates used in them to identify key issues and develop detailed alternatives for consideration in the early stages of crafting new or modified program legislation.

RECOMMENDATIONS FOR SAIPE

Internal and external evaluations of the 1993 and 1995 estimates of poor school-age children for small areas from the SAIPE Program found that the state and county models are working reasonably well but identified areas for further research and development for both the models and the data sources used in them. For school districts, the Census Bureau was constrained to use relatively crude estimation procedures because of the lack of suitable data at the school district level with which to develop a more effective statistical model. Marked improvements in the estimates for school districts and other subcounty areas will require new sources of data for use in models.

Research and Development for Current Models

The Census Bureau's SAIPE Program estimates poverty and income for states and counties by combining the estimates from statistical regression models that are

based on the March Current Population Survey (CPS) with the direct CPS estimates (where available). The procedure for combining the regression predictions and the direct estimates weights them by their relative accuracy.

The use of regression models is necessary because of the high sampling variability of CPS estimates for all but the largest states and counties and the lack of any sample households for two-thirds of the counties. In the state regression models, the state's direct CPS estimate of poverty (or income) for the reference year is the dependent variable, and the predictor variables are obtained from such sources as Internal Revenue Service (IRS) tax returns, Food Stamp Program records, population estimates from the Census Bureau's demographic estimates program, and the previous census. The county regression models use the same general approach. One difference is that the dependent variable in the county models is a 3-year average CPS estimate, centered on the reference year, rather than a single-year estimate. Another difference is that the county regression models are estimated from data for only the counties that have some households in the CPS, whereas the state regression models are estimated from data for all states. The formulation of the county poverty models also requires that a county have at least one poor household in the CPS sample with a member in the relevant age group in order to be included in the model. For poverty models, the county models estimate numbers of poor, while the state models estimate the proportions poor. As a last step in developing county poverty estimates, each of the county estimates in a state is multiplied by a state raking factor so that the sum of the adjusted county estimates equals the state estimate from the state model.

From its review of the state and county models for poor school-age children, the panel identified the following areas for research and development by the Census Bureau in the near term. The Bureau has already begun work in these areas, which would likely benefit the models for other age groups as well.

- o The state and county models, while similar in broad outline, differ in many important details that raise questions about possible inconsistencies in their estimates. A goal for the future should be closer integration of the state and county models. In the interim, work should be conducted to determine the usefulness of including state effects in the county models, for example, by developing a random state-effects model.

- o The current formulation of the county model has the disadvantage of excluding counties from the estimation that have households in the CPS sample but no sampled households with poor school-age children. Work should proceed on estimation techniques, such as generalized linear mixed models, that would include all counties with households in the CPS sample.

- o Both the state and county models have problems in estimating the relative weights that are used to combine the regression predictions and the direct CPS estimates. Procedures that the Census Bureau is developing to address these problems in the short term should be evaluated and implemented, as appropriate, while awaiting the results of longer term research and development.

- o Looking to the future, as more data become available from such sources as the American Community Survey and the 2000 census, the use of time-series and multivariate modeling techniques that make use of multiple years of data from the same survey, separate surveys, or both, could be advantageous. Work on such models should proceed, building on the Census Bureau's previous efforts along these lines.

- o SAIPE model estimates are currently produced with a 3- to 4-year lag between the release date and the income reference year. Work should proceed to find ways to reduce the time lag. For example, for the county model, the estimates might be raked to the state model estimates for the latest of the 3 years of CPS data used in the county model instead of to the state model estimates for the middle year.

The current school district estimation procedure uses 1990 census data to estimate each school district's share or proportion of its county's total number of poor school-age children. These estimates of shares, which are then applied to updated estimates from the county model, have considerable error due to sampling variability for many small school districts. Work should proceed on ways to reduce the sampling variability in the census estimates beyond what has already been achieved by using a simple ratio-estimation technique.

Role of Survey Data

New sources of household survey data may support significant improvements in SAIPE Program estimates in the next decade and beyond. These sources include the 2000 census long-form survey, which will provide income and poverty estimates for 1999 from a sample of about 18 million housing units, and the planned American Community Survey (ACS), which when fully implemented in 2003 will provide income and poverty estimates on a continuous basis from a large sample of more than 2 million responding housing units each year. In addition, two smaller ongoing surveys, the March CPS and the Survey of Income and Program Participation (SIPP), will continue to provide income and poverty estimates.

The panel reviewed these surveys and the possible ways in which estimates from them might be used in the SAIPE Program in light of such

considerations as error due to sampling variability (see Table 1), timely availability of updated estimates, and likely quality of responses and comparability with the current CPS-based estimates and reached several conclusions and recommendations.

General

To inform decisions about the use of the 2000 census long form, ACS, March CPS, and SIPP for SAIPE, the Census Bureau should conduct research to understand and document the differences in their measurement of income and poverty. For this purpose, the Census Bureau should conduct a series of exact matches and analyses of each survey with the 2000 census data and also with data from IRS tax returns for income year 1999.

American Community Survey

Research and development by the Census Bureau should begin now to explore two possible uses for the ACS in SAIPE models for counties. One use is for ACS estimates to form one of the predictor variables in regression models for which the official source of income and poverty estimates, the March CPS, continues to provide the dependent variable. Another use is for ACS estimates to serve as the dependent variable in county models, which could thereby include all, or nearly all, counties in the estimation. The Census Bureau should also conduct research on using ACS estimates for school districts and other subcounty areas to form within-county shares or proportions to apply to updated county model poverty estimates.

If the American Community Survey is to fulfill its potential to play a major role in the SAIPE Program, it is important that the survey have sufficient funding for planned sample sizes over the next decade. Reductions in funding could jeopardize the usefulness of the ACS for SAIPE and, more generally, make it difficult to properly assess the potential uses of ACS data in small-area estimation.

2000 Census

The Census Bureau should plan to use 2000 census long-form estimates to form one of the predictor variables in the SAIPE state and county models. For SAIPE estimates for income year 1999, it could be possible to use the direct estimates from the 2000 census long form, but whether this will be feasible (the data may not be available in time) or desirable is not clear. The Census Bureau should consider the available options and

discuss them fully with users.

Role of Administrative Data

SAIPE estimates for school districts and other subcounty areas cannot currently be produced by using regression models similar to the state and county models, although such models would likely improve upon the current shares procedure. No administrative records data sources currently exist that can provide consistently measured predictor variables for a subcounty model, in the way that tax return and food stamp data are used in the state and county models.

The panel reviewed the advantages and problems of developing IRS tax return and food stamp data for use in subcounty models. Such use would require improvement of the Census Bureau's capabilities for geocoding addresses from administrative records to small geographic areas. The Bureau should give high priority to the continued development of its Topologically Integrated Geographic Encoding and Referencing/Master Address File (TIGER/MAF) system, and, as soon as possible after the 2000 census is completed, it should study the extent to which TIGER can be used to geocode addresses on tax returns to school districts.

Use of administrative records data requires regular reviews of their quality and consistency in terms of how they relate to income or poverty across geographic areas and over time. The review should include identifying possible changes to administrative records systems that would benefit estimation without undue cost to the data collection agency or burden on respondents. For the SAIPE poverty models, it is particularly important to review the interarea comparability of food stamp data in light of the changes in eligibility provisions and participation rates for food stamps that have occurred as a consequence of the changes in welfare programs beginning in 1996.

The panel also considered the advantages and problems of using data from the National School Lunch Program for improved poverty estimates, specifically for school districts. School lunch data might be used, alone or combined in some manner with census and ACS data, to form within-county shares to apply to updated county model estimates. Alternatively, school lunch data might be used as a predictor variable in a regression model for school district poverty estimates. Although issues of comparability across areas and the current lack of a centralized source for the data present problems in using school lunch counts to estimate poverty, the panel concludes that further evaluation may be warranted to determine the usefulness of those data for the SAIPE school district estimates.

Estimates of total population and population by age are required for many uses of small-area income and

poverty estimates from SAIPE; postcensal population estimates are developed by using administrative records such as tax returns. The panel recommends several areas for research and development to improve the estimates, including: ways to improve population coverage in tax return files, use tax returns for estimating population by age, and geocode tax returns to subcounty areas; reassessment of the usefulness of school enrollment data for county and school district estimates of school-age children; and ways to use the MAF and, perhaps, the ACS to improve population estimates.

NOTE

This article is based on the executive summary of the panel's final report:

National Research Council (2000a), *Small-Area Income and Poverty Estimates: Priorities for 2000 and Beyond*. Panel on Estimates of Poverty for Small Geographic Areas, Constance F. Citro and Graham Kalton, editors. Committee on National Statistics. Washington, DC. National Academy Press.

The panel previously published three interim reports, each of which evaluated a specific set of estimates of poor school-age children for use in allocating Title I funds to counties and school districts. These three reports have been combined in a volume that documents and evaluates the Census Bureau's current methodology for producing estimates of poor school-age children:

National Research Council (2000b), *Small-Area Estimates of School-Age Children in Poverty: Evaluation of Current Methodology*, Panel on Estimates of Poverty for Small Geographic Areas, Constance F. Citro and Graham Kalton, editors. Committee on National Statistics. Washington, DC. National Academy Press.

PANEL MEMBERS AND STAFF

Graham Kalton, (*chair*), Westat, Rockville, MD
David M. Betson, Department of Economics,
University of Notre Dame
Nancy E. Dunton, Midwest Research Institute,
Kansas City, MO
Wayne A. Fuller, Department of Statistics,
Iowa State University
Thomas B. Jabine, Consultant, Washington, DC
Sylvia T. Johnson, School of Education, Howard
University
Thomas A. Louis, School of Public Health,
University of Minnesota
Sally C. Morton, RAND, Santa Monica, CA
Jeffrey S. Passel, Urban Institute, Washington, DC
J.N.K. Rao, Department of Mathematics and Statistics,
Carleton University, Ottawa, Canada
Allen L. Schirm, Mathematica Policy Research, Inc.,
Washington, DC
Paul R. Voss, Department of Rural Sociology,
University of Wisconsin
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
Jamie Casey, Project Assistant

TABLE 1 Illustrative Coefficients of Variation (CV) for Estimates of an 18 Percent Poverty Rate for School-Age Children

Population Size	<u>Cumulative Percentage Distribution</u>				<u>CV from Simulated American Community Survey (percent)</u>			CV from Simulated 1990 Census Long Form (percent)
	<u>Counties</u>		<u>School Districts</u>		1-year	3-year	5-year	
	Areas	People	Areas	People	Average	Average	Average	
100,000	15	73	2	27	15	9	7	6
50,000	27	83	6	44	21	12	9	8
25,000	47	92	16	63	30	17	13	11
15,000	63	96	24	76	39	22	17	14
5,000	91	~100	53	94	67	39	30	25
2,500	96	~100	69	97	94	54	42	35

SOURCE: Adapted from Table 4-3 in National Research Council (2000a; see note above).