Estimation Details for Puerto Rico

**Obtaining Estimates of Child Poverty
in Puerto Rico:
An Overview of Current Practice
and
Recommendations for Improvement
Submitted to:
Panel on Estimates of Poverty for Small Geographic Areas
National Academy of Sciences**
Submitted by:
Robert Santos, National Opinion Research Center at the University of Chicago
David Waddington, Bureau of the Census, U.S. Department of Commerce
January, 1999

1. Introduction

The National Academy of Sciences (NAS) Panel on Estimates of Poverty for Small Geographic Areas (Panel) was convened by the Committee on National Statistics and has been charged with the task of reviewing the methodology used to develop state- and county-specific estimates of children in poverty. These estimates are used by the Department of Education for the disbursement of over $8 billion in Title I funds for education programs for disadvantaged children. Consequently, the scientific rigor of the estimation process and the quality of the input data are of principal concern to the government.

The Bureau of the Census staff (David Waddington, Alexander Strand, Joe Dalaker, and others) and a NAS consultant (Robert Santos of the National Opinion Research Center) have collaborated to conduct an investigation and produce this report, which has four objectives:

1) document the process used in assessing child poverty estimates for Puerto Rico;
2) provide a brief overview of the methodology used in the 1996 Puerto Rican Family Income Survey (PRFIS);
3) provide our assessment of the methodology used in the PRFIS, noting both strengths and limitations; and
4) document our recommendations for improving the PRFIS.

2. Process Used in the Assessment

The Bureau of the Census (Bureau) develops estimates of children in poverty to be used in the allocation and disbursement of Title I funds. Using data from the Current Population Survey, along with administrative records on food stamp participants and IRS data, the Bureau develops state and county specific estimates. Data from the PRFIS have been used for estimating the number of children in poverty for Puerto Rico.

The NAS Panel has evaluated the methodology for developing state and county estimates of children in poverty and published their findings under separate cover. (*Small-Area Estimates of School-Age Children in Poverty, Interim Report 1: "Evaluation of the 1993 County Estimates for Title I Allocations"*; *Small-Area Estimates of School-Age Children in Poverty, Interim Report 2: "Evaluation of Revised 1993 County Estimates for Title I Allocations"*; and *Small-Area Estimates of School-Age Children in Poverty, Interim Report 3: "Evaluation of 1995 County and School District Estimates for Title I Allocations".*)

Assessments of the data quality for the corresponding Puerto Rican estimates have been hindered by a paucity of documentation, as well as a language barrier. As such, it was necessary to devote special attention by the Bureau and the NAS Panel to this assessment effort.

The process used for the assessment involved the commissioning of an exploratory mission to Puerto Rico for the purpose of gathering sufficient information to conduct the evaluation of methodology and data quality (to the extent reasonable).

The Bureau of the Census contacted the Puerto Rican Bureau of Labor Statistics to obtain the 1996 PRFIS survey data and concomitant documentation. The following materials were received from Edna Ayuso Rosa, Chief of the Special Economics Studies Division, Bureau of Labor Statistics, Puerto Rico:
* documentation of the sample design used in the PRFIS;
* tabulations pertaining to the development of weights;
* tabulations of field results (response rates, cases completed per interviewer);
* the 2 page PRFIS questionnaire;
* interviewer manual for the 1996 PRFIS;
* two data files (stored as dBASE files) with data from the PRFIS.
Additionally, we gathered information from a conference call meeting with the Puerto Rican BLS staff, their consultants, the BOC staff and the NAS consultant. Specifically, on October 29, 1998, the following people met

via telephone: Edna Ayuso Rosa (PR-BLS), Dr. Carlos Toro (PR-BLS consultant), Dr. Jaime Bofill (PR-BLS consultant), Robert Santos (NORC), and David Waddington (BOC). Drs. Bofill and Toro are currently conducting an evaluation of the Family Income Survey at the request of the PR-BLS.

The meeting proved to be quite productive. The participants provided answers for many (but not all) of our questions. Santos and Waddington were invited to provide additional suggestions to the staff at the Puerto Rican Department of Labor regarding questionnaire design and methodology. However, the first order of business was the production of this report for the NAS Panel.

3. Overview of the 1996 Puerto Rican Family Income Survey

The following text summarizes the PRFIS survey design and methodology. The PRFIS survey is an area probability household sample of Puerto Rico that is conducted biannually as an offshoot of the PR Labor Force Survey (PRLFS). Face-to-face interviews are conducted during a field period that is concurrent with the monthly PRLFS and which utilizes the PRLFS interviewers. Interviews for this survey were conducted in February and March of 1997. The Bureau of the Census obtained the 1996 PRFIS survey data in mid-June, 1998.

Sampling. The PRFIS sample design utilizes multi-stage area probability sampling; households are selected with equal probability. The PRFIS is drawn from the Puerto Rican Labor Force Survey (PRLFS) (analogous in design to the CPS); 4 rotation groups (out of 8 total in any given Labor Force Survey month) were used from previous surveys: Nov. & Dec. 1995, and Nov. & Dec. 1996. These rotation groups are *inactive* for PRLFS at the time the PRFIS is conducted. (See Appendix A, *Revision of the Household Sample for the Labor Force Survey, 1990.*) The sample design for the 1996 PRFIS is virtually identical to that of the previous survey -- the 1994 PRFIS.

All addresses sampled into the PRLFS (regardless of response disposition to the PRLFS) were taken into the 1996 PRFIS sampling frame. That is, both responding and nonresponding households to the PRLFS were eligible to be drawn into the PRFIS. A total of 3152 housing units were selected into the 1996 PRFIS. Of these, 2560 were found to be eligible for interview, 2314 were successfully interviewed. This produces an interview response rate of 90 percent (i.e., 2314/2560).

Questionnaire. A two page questionnaire was administered to gather 1996 income from all sources. A questionnaire was to be completed for each household resident at ages 16 and over. For the designated household head, an additional question captured annual income from all sources received by the collection of all children under 16 years old residing in that household. Information was to be collected directly from each person 16 or older or from any knowledgeable family member (i.e., proxy reporting was allowed). When respondents reported sub-annual amounts (e.g., weekly pay), interviewers were directed to convert the response into an annual amount, and record the annual amount onto the paper questionnaire.

The 1996 PRFIS questionnaire is similar to that of the previous PRFIS (whose data collection occurred in 1995). One difference is that the 1994 survey collects age and sex of all household members in Part II. In contrast, the 1996 survey captures the same information in Part I from the household head/household informant, then requests it again from each individual 16 and over in Part II (household members under age 16 are excluded from being asked questions in Part II). It is a bit odd that such demographic data would be collected twice for members 16 and over-- once by the household respondent in Part I, then a second time by the individual in Part II.

Another difference between the 1994 and 1996 instruments is that the 1996 PRFIS instrument gathers total aggregate income among *children under 16 years of age*. This occurs in Part I of the questionnaire, which is administered only to the adult designated as household head. Part I gathers household composition and total aggregate income from all sources that were received from all children in the household under age 16.

In the *1994 PRFIS*, the survey collects income data for persons less than 16 years of age with two different questions:

A) a multiple response question asking for indications of all income sources received from any child; and
B) a single question using income categories as responses, asking for the aggregate annual income from all sources across all children under 16 years of age.

In contrast, the (current) *1996 PRFIS* gathers this information in a somewhat different fashion by asking for two amounts:

1) aggregate annual income in 1996 from *wages and salaries* (i.e., the combined total across all children under 16); and
2) aggregate annual income in 1996 from *all other sources* (i.e., the combined total across all children under 16).

Turning to individual income questions in Part II of the questionnaire, the structure of the questionnaire is the same between the 1996 and the 1994 PRFIS. Response categories for each of the 20 income items are structured so that you first ask a yes/no question regarding receipt of a specific income type. If the respondent

indicated receipt of a specific income type, he/she was subsequently asked for the annual income amount. There was no mechanism (i.e., a numeric response code on the questionnaire) for either (the 1994 or 1996 PRFIS) survey to indicate nonresponse to a specific income amount or category (e.g., don't know, refused). However, there was a response box for those who refused to answer any of the income questions.

Field Operations. The same interviewer force which implements the PRLFS is used to obtain the PRFIS interviews. One day of interviewer training was provided to the field interviewers (see Appendix D, *Instructional Manual*).

Data collection schedules for the PRLFS and the PRFIS are concurrent. Field staff may conduct interviews for both surveys on the same day. We note that the PRLFS work load is roughly twice as large (i.e., 6,000 households) as that of the PRFIS. With regard to quality control operations, neither validation re-interviews nor direct supervisor observations are conducted for the PRFIS. However, validation interviews are conducted for the PRLFS.

Based on a review of the documentation, field protocols for the 1996 PRFIS (with data collection occurring in 1997) appear to be very similar if not identical to those of the previous PRFIS (whose data collection occurred in 1995).

Interviewers have a *weekly* average of 55 PRFIS interviews per interviewer. For the PRLFS, the weekly average is 109 cases per interviewer (for a combined weekly average of 164 cases per interviewer).

Data Processing. Item nonresponse was reported to be very low (described as nonexistent). No imputation has been performed. However, data are taken from the PRLFS and appended to the PRFIS (the BLS staff alluded to household composition, but the exact nature is vague). Logical, deductive imputation is performed (but was reported to be "minor" given the high rates of item response).

Weighting. Weights are developed with the underlying assumption that households are selected with equal probability (i.e., the design is self-weighting in the absence of nonresponse and noncoverage).

Consistent with PRLFS weighting protocols, each *rotation group* is weighted separately to the population total for that month/year: Nov. 1995; Dec. 1995; Nov. 1996; Dec. 1996. Additionally, post-stratification weights were used for gender, age, and geography (within rotation group) to adjust to independent totals.

Single adjustments for unit nonresponse and post-stratification are *not* performed; rather, a single (combined) adjustment is developed via post-stratification. The post-stratification adjustment is based on a moving average (over time) of the population distribution by gender and age. The "control" distributions are developed independently by the Puerto Rican Planning Board.

Household Income. Household income is derived using a weighted sum of household resident incomes, normalized using the weight of the household head. Let W(h) denote the analytic weight of the head of household in household h.

Then, household income is obtained as follows:

where summation occurs over all persons *I* in household *h* with nonzero annual income, and where *INC* represents the reported annual income of person *I* within household *h*. It is unclear how this income measure is intended to be used.

The Bureau did not use this weighted income measure for calculating the child poverty estimate for Puerto Rico. The estimate that was provided to the Department of Education was based on the appropriate unadjusted dollar amounts. That is, the annual income for household *h* is:

where summation occurs over all persons *I* in household *h* with nonzero annual income, and where *INC* represents the reported annual income of person *I* within household *h*.

Other. Preliminary analyses conducted by Bofill and Toro suggest significant under-reporting of social program participation and benefits (e.g., food stamps). Although this indicates that measurement errors exist in the reporting of these income sources, it does not identify the nature of the error (e.g., question comprehension, social desirability, avoidance of disclosure, etc.).

The PRFIS is expected to continue as an ongoing biannual survey, probably with a similar design approach. The PR BLS has commissioned the consultants to make recommendations for the development of a research agenda to improve the quality of the PRFIS data.

4. Assessment of the Methodology

The methodology used in the 1996 PRFIS has strengths and weaknesses.

The *strengths* include:

* an equal probability sample design of households in Puerto Rico;
* a survey design which is integrated with the monthly PRLFS;
* a relatively high unit response rate;
* relatively low item nonresponse (deduced from the assertion that imputation was not necessary);
* a "piggyback" implementation strategy (from the PRLFS) which gathers survey data at marginal cost.

The *weaknesses* include:

* use of proxies for person level annual income reporting;
* mathematical manipulation of responses by interviewers for converting sub-annual amounts to annual amounts;
* evidence of measurement error (under-reporting of social program participation and benefits), suggesting problems of comprehension, fear of disclosure, etc.;
* potential for large interviewer effects due to very large average workloads per interviewer;
* no apparent quality control or validation protocols to substantiate the quality of the field work;
* little documentation is maintained for the data processing phase of the project -- little or no information about imputation exists;
* weighting utilizes a single post-stratification adjustment to compensate for (potentially) differential unit nonresponse and noncoverage;
* weighting to totals is inconsistent across rotation group sub-samples; the 1996 replicates are weighted to 1996 month-specific population totals, while the 1995 replicates are calibrated to 1995 month-specific population totals;
* calculations of total household income for a household use an unconventional algorithm whereby within-household income is averaged using person level analytic weights.

These weaknesses point to a number of recommendations that could easily and in many cases inexpensively improve the quality of the PRFIS data.

5. Recommendations

The following recommendations are based on the information that is available at this time. Despite our persistence, it is possible that additional, relevant documentation could be made available which would warrant changing some of these recommendations (through additions, deletions, or other types of revision).

Recommendation 1: Develop and implement quality control protocols for field work (e.g., validation, re-interviews, supervisor observations) to ensure the quality of the collected data.

Recommendation 2: Allow for the recording of sub-annual responses onto the questionnaire; perform conversion to annual amounts via computer at the data processing stage of the project; also, allow for specific coding of don't know, refusals and zero amounts to individual income items in the instrument.

Recommendation 3: Conduct cognitive lab experiments to investigate:

* respondent *comprehension* levels to survey questions;
* validation of *concepts and/or constructs* (e.g., do respondents know what a trusteeship is? Are there other missing or more appropriate constructs of income that could be used given the economy of Puerto Rico?);
* barriers to *disclosure* and how they can be overcome.

Recommendation 4: Investigate the extent of interviewer effects; diversify/expand the field work to a larger number of interviewers to reduce the average number of cases per interviewer.

Recommendation 5: To the extent feasible, establish a methodological report series which documents the design, field work, outcomes, and data processing that goes into the PRFIS.

Recommendation 6: Develop analytic weights with four components:

* a household sampling weight;
* a unit nonresponse weight adjustment at the *household* level which incorporates (at a minimum) sample design stratification (including geographic stratum, urbanicity, rotation group, etc.);
* a unit nonresponse weight adjustment at the *person* level, using the sampling and household weights, and taking into account demographics known for all sample persons (including the persons who failed to respond), including gender, age, and sampling variables;
* a post-stratification weight adjustment, standardized to a single population defined by the actual period of PRFIS data collection (rather than separate adjustments that point back to rotation group month/year combinations);
* the overall analytic weight would be the product of the appropriate weight adjustments and two analytic weights could be constructed:
    - those for HOUSEHOLD level analyses,
    - those for person level analyses.

Recommendation 7: Re-define household income so that it is simply the sum of incomes from the persons residing in the household (i.e., jettison the current weighted average approach to calculating HOUSEHOLD income) and use the household analytic weight for HOUSEHOLD level analyses.

Recommendation 8: Investigate patterns of item missing data, and the utility of more sophisticated imputation protocols to adjust for item nonresponse to critical items.

Recommendations 1 to 8 relate to design, methodological and estimation issues; additional recommendations can be made with regard to some *process* issues:

Process Recommendation 1: The NAS should consider adding to its membership an individual from the ranks of the Puerto Rican BLS; alternatively, such a person could play a consultancy role or act in an *ex-officio* capacity as an observer/participant to the NAS Panel. The goal is to secure the attention (and active engagement) of the PR BLS in the work of this NAS panel. Such a strategy would provide the Panel with more direct access to information about the PRFIS program. It would also serve to communicate to the PR BLS the importance of the data generated by the PRFIS program.

Process Recommendation 2: The NAS consultant (Santos) and the BOC staff (Waddington, Dalaker, Siegel, Nelson) worked well together on this investigation using a collaborative working relationship. Perhaps a NAS panel member could be assigned to work with the BOC to facilitate an ongoing program of investigation of quality issues as well as assistance to the PR BLS with regard to the PRFIS. This person could readily act as a liaison between the NAS panel, the Census Bureau, and the Puerto Rican BLS.

Process Recommendation 3: A concerted effort should be made to maintain an ongoing, supportive relationship between staff of the PR BLS, staff of the Census Bureau, and the NAS Panel members. The Puerto Rican BLS is currently amenable to suggestions and assistance to improve its design and operations in the PRFIS. Such an opportunity should not be overlooked, and it may not exist for long.

In closing, we note that the NAS-funded exploratory mission called for the NAS consultant to visit to Puerto Rico (if deemed necessary) to meet with the PR BLS and to more directly review methodological/design issues and documentation. It was not necessary to visit PR for this investigation, in large part due to the propitious use of English speaking Puerto Rican consultants contracted by the PR BLS to evaluate the PRFIS and recommend improvements. Future assessments may well necessitate travel to Puerto Rico.