RELIABILITY OF INCOME AND POVERTY DATA FROM THE CURRENT POPULATION SURVEY ANNUAL DEMOGRAPHIC SUPPLEMENT[1]

Jennifer W. Reichert, John C. Kindelberger, U.S. Census Bureau
Jennifer W. Reichert, U.S. Census Bureau, SFC #2 Room 2500-B, Washington, DC 20233

Key Words: response error, reliability, poverty, self versus proxy

1.0 Background

The Annual Demographic Supplement to the Current Population Survey (CPS) collects data on work experience, several sources of income, migration, household composition, health insurance coverage, and receipt of non-cash benefits [2]. The Supplement is also the source of the annual estimate of the national poverty rate. In 1998, for the first time ever, the Census Bureau used reinterview to evaluate response error in the Supplement. Response error is the result of respondent error in reporting or interviewer error in recording information in an interview.

In addition to the usual income and poverty data, the 1998 Supplement was the first to collect data on many of the new welfare programs implemented nationally and at the state level (e.g., transportation assistance, child-care assistance, job training, etc.). There was great interest in assessing the reliability of these data.

Reinterview programs allow us to detect problems in the questions, but usually they cannot identify causes of response error, nor correct the problems. High response variance indicates a problematic question, and moderate response variance suggests some problems with reliability.

2.0 Summary of Results

We found the income and poverty data from the March 1998 Annual Demographic Supplement to be fairly reliable. The most unreliable data were those from the new questions regarding the new welfare reform programs. We also observed some relationships among demographic characteristics, household poverty status, and the amount of unreliability in the data. The Findings section below provides more details on our analysis and results.

3.0 Methodology

3.1 Survey Design and Implementation

We conducted the reinterview for the March 1998 Annual Demographic Supplement. We used a random sample of 1,346 households. To collect enough data to get meaningful measures of response error for the poverty-targeting questions (i.e., public assistance and welfare reform items) in the Supplement, we oversampled poverty households. Poverty households represented approximately 19 percent of the households in the original Supplement interviews. For the reinterview, we designed the sample so that half of the reinterview sample consisted of households at or below the poverty level.

All reinterviews were conducted within two weeks of the original interview. Senior interviewers conducted the reinterviews mostly by telephone, but they conducted personal visits when telephone contact wasn't possible (the original interviews were conducted by personal visit only). Upon contacting the household, the reinterviewers proceeded to re-ask all of the same questions from the original interview. The reinterviewers attempted to contact the same respondent from the original interview, but proxy respondents were used when the original respondents couldn't be contacted.

3.2 Reinterview Model Assumptions

The response error reinterview model assumes the reinterview is an independent replication of the original interview.

Independence means that the response errors are not correlated between the original interview and the reinterview. If the respondents remember their original answers and consciously repeat them in the reinterview, the independence assumption is violated. Lack of independence generally results in underestimates of response variance.

---

[1] This paper reports the results of research and analysis undertaken by Census Bureau Staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

Replication means that the reinterview was conducted under the same conditions (same interviewer, same respondent, same interview mode, etc.) as the original interview. If the reinterview replicates the original interview, the distribution of the original and reinterview responses will be the same. With quantitative data, the means and variances of the original and reinterview responses will be equal. With categorical data, the difference between original proportion-in-category and the reinterview proportion-in-category, the net difference rate (NDR), will have an expected value of zero.

3.3 Measures Used to Estimate Response Variance

Random measurement errors in the survey process (nonsampling error) increase the mean-squared error of the data collected. When these measurement errors are not correlated with the responses or with each other, we call this variability "simple response variance." In categorical data, simple response variance can imply bias.

The **index of inconsistency (index)** and the **gross difference rate (GDR)** are the principle measures of response variance in categorical data. Overall estimates of the index and the GDR for categorical questions, the **aggregate index** and the **aggregate GDR**, apply to questions with three or more answer categories.

3.4 Index of Inconsistency

The index of inconsistency estimates the ratio of response variance to total variance for a question answer. It is a relative measure of response variance.

The aggregate index is similar to the index of inconsistency, but applies to the entire question rather than a specific answer category. It is a weighted average of the index of inconsistency across all categories for the question. For questions with two categories (e.g, yes/no questions), the index of inconsistency is equal to the aggregate index.

An aggregate index of zero means responses were in perfect agreement, but an index of 100 does not mean that all of the respondents changed answers. Rather, it means that we saw what we would expect if there were no relationship between original and reinterview answers beyond chance.

We use this rule of thumb to interpret the index of inconsistency and the aggregate index:

Index Value	Response Variance Level	Interpretation
Less than 20	Low	Usually not a major problem
Between 20 and 50	Moderate	Somewhat problematic
Greater than 50	High	Very problematic

Either of these factors may cause high response variance:

- The methods used to collect the data may need improvement or the questions may be unclearly written.
- The concept itself may not be measureable.

3.5 Gross Difference Rate

The gross difference rate (GDR) is the percentage of responses that fall in a category in the original interview but not in the reinterview, or vice versa. For a single-category question, one-half the GDR equals the simple response variance. The GDR estimates the consistency of reporting.

The aggregate GDR applies to an entire question rather than to a specific answer category. For questions with more than two categories, the aggregate GDR is the percentage of responses that change between the original interview and the reinterview.

The GDR is more difficult to interpret than the index of inconsistency. Large GDRs indicate serious response variance in the data. Unfortunately, a small GDR is no guarantee of good consistency. In a low-frequency category, even a small GDR can represent high response variance relative to total variance. If this is the case, the index of inconsistency will tell us.

3.6 Cross-Tabulations

For a "yes/no" question, the cross-tabulation looks like this:

Reinter-view Response	Original Response				
	Total	N/A	Sub-total	Yes	No
Total					
N/A					
Subtotal			n	a+c	b+d
Yes			a+b	a	b
No			c+d	c	d

where,

n = the number of respondents who answered the question in both the original interview and the reinterview
a = the number of respondents who answered "yes" both times
b = the number of respondents whose answer changed from "no" in the original interview to "yes" in the reinterview
c = the number of respondents whose answer changed from "yes" in the original interview to "no" in the reinterview
d = the number of respondents who answered "no" both times.

To compute the response variance measures, we only used cases where respondents answered the question in both the original interview and the reinterview.

In multi-category questions, these cross-tabulations show the movement among answer categories between the original interview and the reinterview. Patterns in this movement can provide clues to the reasons for inconsistent reporting. In some cases, such movement may suggest question revisions to reduce response variance.

3.7 Limitations of Analysis

The reinterview may not have been independent of the original interview due to the possibility that respondents remembered and repeated their answers from the original interview or were less cooperative because of the burden of the extra interview.

Operational constraints make it difficult to conduct the reinterview as an exact replication of the original. When a reinterview does not replicate the original interview perfectly, the differences in methodology may cause an overestimation or underestimation of the response variance.

3.8 Response Variance Formulas

These formulae use **a, b, c, d** and **n** from the preceding cross-tabulation table.

For multi-category questions, we treat "in category" as *yes* and "not in category" as *no*.

- Original Percentage – the percentage of original responses in a specific answer category. The formula is:

  $P_o = [(a+c)/n] \times 100$

- Reinterview Percentage – the percentage of reinterview responses in a specific answer category. The formula is:

  $P_r = [(a+b)/n] \times 100$

- Gross Difference Rate (GDR) – the percentage of the responses that change into or out of a specific answer category. The formula is:

  $GDR = [(b+c)/n] \times 100$

- Simple Response Variance (SRV) – the average variance of responses from the same units to the same question over repeated interviews. The simple response variance equals half of the GDR (expressed as a proportion). The formula is:

  $SRV = (b+c)/2n$

- Index of Inconsistency (index) – the ratio (scaled as a percentage) of simple response variance to the total population variance for a characteristic. The index represents the proportion of the total population variance for a characteristic caused by simple response variance.

  For categorical data, when $P = P_o = P_r$, that is, when the percentage reported in-category is the same on both original interview and reinterview, the formula is:

  $$\text{Index} = [SRV/P(1-P)] \times 100$$
  $$= [[(b+c)/2n]/P(1-P)] \times 100$$

  where the total population variance for the characteristic is P(1-P).

If P_o and P_r are not equal, P(1-P) is estimated by:

$$\frac{1}{2}[P_o(1-P_r) + P_r(1-P_o)].$$

- Overall GDR (aggregate GDR) – the percentage of people who change their answers to a question.

- Aggregate Index of Inconsistency (aggregate index) – a weighted average of indices of inconsistency across all categories of the question (weighted by percent-in-category).

4.0 Findings

Although we conducted our analysis on five broad groups of questions (i.e., work history, household income, public assistance programs, migration, and health insurance)[1], this paper concentrates on the results from the analysis of the household income and public assistance data items. Because of the high interest in the new welfare programs, we present the results for both poverty and non-poverty households. We discuss the affect of self versus proxy responses on the level of response variance. Finally, we attempted to fit the data to a model to determine what characteristics may affect the reliability of the respondents' answers.

4.1 Overall Findings

For the questions regarding household income, we found that the data collected were fairly reliable. Only 22 percent of the questions were found to have high response variance. The highest percentage of the questions (41%) were in the moderate range. Response variance was low for 37 percent of the questions.

The data on public assistance programs were less reliable than the household income data. We found moderate to high response variance for 83 percent of the questions, suggesting poor reliability. The remaining 17 percent had low response variance. Cognitive studies showed that for many of the public assistance questions, the high response variance may be due to the respondents' unfamiliarity with the terminology and program names [1]. If the lack of reliability was indeed due to a lack of understanding, we would hope to see an improvement in the reliability over time as the respondents become more knowledgeable of the programs and terminology.

Table 1. Number of Questions by Question Group and Response Variance Level

Question Group	Total Evaluated (%)	High (%)	Moderate (%)	Low (%)
All Questions	53 (100)	13 (24.5)	23 (43.4)	17 (32.1)
Income	41 (100)	9 (21.9)	17 (41.5)	15 (36.6)
Public Assistance	12 (100)	4 (33.3)	6 (50.0)	2 (16.7)

4.2 Poverty Versus Non-Poverty

One of the goals of this reinterview program was to evaluate the reliability of the data collected for the various new welfare programs implemented in 1997 and 1998. Because most of these programs are designed to benefit households below the poverty line, we expected there to be a significant difference in the reliability of the data for poverty versus non-poverty households.

For those questions that apply only to poverty households, such as public housing, food stamps, and other public assistance programs, there was a positive association between poverty status and response variance. As we expected, non-poverty households were very consistent in their "no" answers to these questions. In general, the data provided by the poverty population contained higher response variance for those questionnaire items that target the poverty population. However, we cannot conclude that poverty households are less reliable in general. In fact, the data provided by the non-poverty population were more unreliable for the questionnaire items that applied primarily to the non-poverty population.

In this part of the analysis, the GDR emerged as a more relevant measure for comparing discrepancies than the Index of Inconsistency. For nearly all of the questions targeting the poverty population, the index took a higher value for the non-poverty households than the poverty households. However, the GDR is clearly lower for non-poverty households (see Table 2). In addition, the differences between the poverty and non-poverty indexes were not statistically significant, but the GDRs between the two populations did show significant differences. The "balancing" seen in the indexes is due to the total variance in the survey responses, which is lower for non-poverty households with respect to the poverty-targeting variables. Since the index measures simple response variance relative to total variance, a lower total variance

drives up the index, countering the effect of a lower percent of discrepancies. Thus, the overall variance differences between poverty and non-poverty households caused the significant differences in the GDRs to be "washed out" in the Index.

Table 2. Poverty vs. Non-Poverty

	Poverty		Non-Poverty	
	Index	GDR	Index	GDR
WIC Program	28.8	10.9	41.3	5.7
Public Housing	42.5	13.8	29.2	4.0
Food Stamps	15.1	7.0	24.1	3.1
Bonds/ T-Notes	35.1	7.5	43.2	21.3
Dividends	42.0	4.7	40.6	18.4
Estates/ Trusts	100.7	2.1	93.3	4.5

4.3 Same Versus Different Respondents

In the reinterview, the interviewer tries to contact the same household respondent as was contacted in the original interview. However, proxies are accepted if necessary. We speculated that the moderate and high levels of response variance we observed may have been due to proxy reinterview respondents. Our findings did not strongly support this expectation.

The same respondent answered questions in both the original interview and reinterview in 91 percent of the cases. For those cases where the respondents were different household members, there were only a small number of questions where the amount of response variance was dependent on whether or not a different respondent answered the questions in the two interviews. And, the overall response variance for those questions improved only slightly when the households with different respondents were removed from the analysis (see Table 3.). In one case, for the question on estates/trusts, the overall index actually increased when we removed the proxies. This is because by removing the proxy households, we reduced the total variance thereby increasing the index.

Table 3. Same vs. Different Respondent

	Index of Inconsistency	
	With Proxies	Without Proxies
Checking Account	50.3	48.6
Supplemental Security Income	32.5	29.1
Dividends	36.6	33.3
Alimony	45.4	39.6
Estates/Trusts	95.0	101.1

4.4 Demographic Associations

One goal of the analysis was to determine if demographic characteristics were associated with the presence of response error. We used two tests in an attempt to answer that question. We tested for independence between demographic characteristics and response discrepancies, and we performed a logistic regression using percent of response discrepancies as the dependent variable and various demographic characteristics as the independent variables.

We performed a stepwise logistic regression to eliminate the problem of correlation between the independent variables. We attempted to fit the model using the following independent variables, which showed significant association in our tests for independence:

- poverty status:
  - below poverty line
  - at or above poverty line
- homeowner status:
  - owns home
  - does not own home
- respondent age
  - 21 or under
  - over 21
- respondent education level
  - high school or less
- race
  - black
  - not black
- labor force status
  - unemployed
  - employed or not in the labor force
- Hispanic status
  - Hispanic
  - not Hispanic

Our results were not surprising. Similar to our results from the comparison between poverty and non-poverty (section 4.2 above), we found that there are high discrepancies in the data provided by the poverty population for the questionnaire items that are most applicable to the poverty population. For those same questionnaire items, there are low discrepancies for the non-poverty population.

Table 4 illustrates the results of this analysis. A "+" indicates that for the poverty-related classification within the explanatory variable, there is a higher probability of discrepancies. A "–" indicates a lower probability of discrepancies for the poverty-related classification.

For example, the poverty-related classification within the Poverty Status variable is "poor." For the population that is classified as poor, there is a higher probability of discrepancies for the poverty variables and a lower probability of discrepancies for the non-poverty variables. Poverty variables are those variables most applicable to the poverty population (e.g., WIC, Public Housing, Food Stamps). Those variables that apply mostly to the non-poverty population are non-poverty variables (e.g., bonds/t-notes, dividends, estates/trusts).

Table 4. Demographic Associations

	Poverty Variables	Non-Poverty Variables
Poverty Status	+	-
Homeowner Status	NA	+
Respondent Age	+	NA
Education Level	+	-
Race	+	-
Labor Force Status	+	-
Hispanic Status	+	-

Notes: + → Higher probability of discrepancies
- → Lower probability of discrepancies
NA → No significant findings

The results from our demographic association tests are somewhat confounded. We know there are significant correlations between poverty status and the other demographic variables used as explanatory variables. Also, the demographics that appear to be associated with response variance are also associated with the question categories. In general, these results indicate that response error in the Supplement tended to increase with the variability within the question category, and is not attributable to any particular demographic groups.

5.0 References

1. Kindelberger, J.C. (January 1999), *Response Variance in the March 1998 Current Population Survey Income Supplement*, Census Bureau report.

2. Current Population Survey Technical Paper 63, Design and Methodology, U.S. Bureau of Labor Statistics and U.S. Census Bureau, March 2000.