Problems of Screening for Poverty Status

John F. Moeller and Nancy A. Mathiowetz[1]

Abstract: The 1987 National Medical Expenditure Survey used a screening interview conducted in the fall of 1986 to identify individuals of interest for oversampling, including the poor and near poor. To minimize response burden and nonresponse, a single question was asked to determine poverty status. In addition to the potential for measurement error, the dynamic nature of poverty confounds the problem of identifying "survey-year" poor in a screening interview. This paper examines a more efficient approach for screening for poverty status, given both the propensity for underreporting income and the dynamic nature of the measure.

Key words: Measurement error; oversampling; misclassification; statistical modeling; expected income.

1. Introduction

Survey sample designs often call for over-sampling population subgroups of particular interest. One method by which oversampling is operationalized is to conduct a screening interview to identify characteristics of these subgroups. Although this procedure is quite effective for static characteristics subject to low levels of measurement error and little or no change over time, the approach works less well for characteristics such as poverty status due to both the dynamic nature of the measure and the propensity of error in reports of income.

The 1987 National Medical Expenditure Survey (NMES) used a screening interview conducted in the fall of 1986 to identify individuals of interest for oversampling: the elderly, poor, functionally limited, blacks, and Hispanics. In the NMES screener, poverty status was based on the response to a single income question concerning total family income for the prior twelve months. The use of a single income question, while minimizing non-response, may result in measures of income that are significantly understated as compared to data obtained from several detailed questions (U.S. Bureau of the Census 1979). In addition, the movement into and out of poverty, estimated to be between 25% and 46% in any two adjacent years (Duncan 1984; U.S. Bureau of the Census 1989) further confounds the problem of screening for poverty status.

The NMES data provide a unique opportunity to address the issue of misclassification of poverty status and to develop and test a more efficient means for identifying income classes for the purpose of oversampling. The efficiency models (Section 6) provide a practical solution for those researchers in need of methods to oversample the poor.

2. Background

When developing a questionnaire to obtain information on the population necessary for sample selection, a conflict exists between the need for a brief instrument and the need for accurate measurement. There is a general belief that a brief screening instrument will both increase the response rate for the screening operation as well as serve as a "foot-in-the-door" for future contacts (Groves and Magilavy 1981).

With respect to income questions, there appears to be a relationship between the number of detailed questions asked the respondent and the accuracy of reported income. For example, comparisons of the reports of family income collected in the Current Population Survey (CPS) during the month of October with the responses to 11 questions identifying 23 different sources of income in the March income supplement have consistently shown a ratio of approximately .80 over a period from 1967 to 1976. Although the reference period for the two reports is somewhat different (previous 12 months for the October report; prior calendar year for the March supplement report), much of the 20% difference is believed to be a result of the different question forms (U.S. Bureau of the Census 1979).

Screening for poverty status is further complicated by the considerable movement into and out of poverty between any two

years. This means that even if a person's poverty status is accurately measured as "poor" in year "$t − 1$," there is at least a one in five chance that the person would be misclassified as in poverty in year "t." Duncan (1984) examined adjacent-year pairs of reports of income from 1969 to 1978 using data from the Panel Study of Income Dynamics (PSID) and found that 35% to 46% of those poor in one year were not poor in the following year. In a more recent publication, changes in poverty status were examined for individuals participating in the Survey of Income and Program Participation (SIPP) between 1984 and 1985 (U.S. Bureau of the Census 1989). The study examined both the movement into and out of poverty and found that 25% of those poor in 1984 were not poor in 1985, and 21% of those classified as poor in 1985 were not poor in the previous year.

3. NMES Design

The 1987 NMES Household Survey was a one-year panel study which collected information on health care status and utilization of health care services and associated expenditures for the civilian noninstitutionalized household population. As noted above, a screening interview was used to facilitate sample identification of population subgroups targeted for oversampling. The screening interview, conducted in October and November of 1986, was a relatively short questionnaire focusing only on the questions necessary for completion of the sample selection. Due to the complicated nature of simultaneously oversampling on several population characteristics, concurrent screening and sample selection were not feasible. Given the dynamic nature of poverty status in adjacent years, a concurrent screening and selection procedure would not have eliminated the

problem of misclassification that is the focus of this paper since selection for year t, the survey year, would still be based on income in year $t - 1$.

To minimize the number of questions asked during the screener, income information was obtained by asking a single question concerning the combined income for the reporting unit. In the NMES study, a reporting unit, hereafter referred to as an RU, was defined as a group of individuals related by blood, marriage, or adoption, living within a dwelling unit and included unmarried students, ages 17–21, who were living away at college. The income question read:

Which of these income groups represents your total (combined family) income for the past twelve months? Please include income from all sources such as wages, salaries, social security or retirement benefits, interest or dividends, rent from property and so forth. Just tell the letter.

Several showcards were prepared in advance which indicated (by letter) the poverty and 125% poverty lines for reporting unit sizes ranging from one to nine and over and for RUs with and without at least one person 65 years and older. The interviewer was instructed to hand the respondent the card corresponding to the RU size and, as noted above, the respondent reported the appropriate letter corresponding to the income for the reporting unit. Poverty thresholds were based on the official poverty thresholds for March 1986 released by the U.S. Bureau of the Census, and were adjusted according to the Consumer Price Index for the period between March 1986, and the month prior to the field period, September 1986. In addition, in an attempt to compensate for the known underreporting of income associated with a single income question, the defining points for the poverty and 125%

poverty lines were further reduced by 20%. The adjusted thresholds were rounded to the nearest $100.

In contrast, the income data collected for the 1987 calendar year paralleled the questions used in the March Supplement to the Current Population Survey. The income supplement was administered during the fourth round of data collection, conducted between January and March of 1988. Detailed information on 26 different income categories was collected separately for each RU member. Included among the income categories were wages, veteran's payments, Supplemental Security Income, Social Security, private pensions, and interest income.

Poverty status for 1987 was determined by summing all of the sources of income for all of the individuals within the RU and comparing that total with thresholds based on RU size released by the U.S. Census Bureau for 1987. Poverty status was assigned at the RU level and all individuals within the RU were given the same poverty status.

4. Poverty Status Gross Flows

According to NMES, nearly 62% of the 24 million persons who were classified as poor at the time of the screener were no longer poor in 1987, a turnover rate far in excess of the rates cited above from the PSID and the SIPP studies. By contrast, the gross flow of persons into poverty between the two years is similar to other longitudinal surveys. In NMES, 15.4% of the persons who were near-poor in 1986 (between 100% and 125% of the poverty line) became poor in 1987, while only 4.1% of the nonpoor in 1986 became poor in 1987. The comparable percentages from SIPP data for 1984/1985 were 21.8% and 1.6%, respectively.

The statistics cited above are based on person-level estimates; the NMES person-level estimates are provided as a means for comparison between NMES data and other data that have examined transitions in poverty status over time. However, the analyses presented in this paper will be limited to RU-level analyses. In part, using the reporting unit as the unit of analytic interest facilitates the use of consumer-unit based permanent income models. In addition, since all members of an RU were assigned the same poverty status, using the person as the analytic unit of interest would result in nonindependence of observations.

Table 1 presents the cross-classification of reporting units by 1986 and 1987 poverty status. Fifty-six percent of the reporting units classified as poor in 1986 left poverty in 1987, whereas the movement into poverty was quite small, with 16% of the near poor and 4% of the nonpoor entering poverty in 1987. As noted above, the exit rate from poverty far exceeds that reported in other studies, whereas the rate of entrance into poverty is comparable.

We hypothesize that the magnitude of the movement from screener poverty to near-poor or nonpoor status in 1987 suggests that in addition to the dynamic nature of poverty status, income was likely to have been underreported by a significant percentage of those classified as poor at the time of the screener. This hypothesis is based, in part, on the underreporting of income documented in the CPS. The research presented here attempts to develop a method for improving the efficiency of screening for poverty status.

Table 1. Reporting units cross-classified by screener and 1987 poverty status (unweighted estimates)[a]

Screener poverty status (1986)	1987 Poverty status			
	Poor	Near poor[b]	Nonpoor[c]	Total
Poor	1,056	325	1,037	2,418
Row %	44	13	43	100
Column %	65	45	11	20
Near poor[b]	275	242	1,230	1,747
Row %	16	14	70	100
Column %	17	34	13	15
Nonpoor[c]	302	155	7,397	7,854
Row %	4	2	94	100
Column %	18	21	77	65
Total	1,633	722	9,664	12,019
Row %	14	6	80	100
Column %	100	100	100	100

Source: National Medical Expenditure Survey.
[a]Restricted to reporting units with non-missing screener and 1987 income.
[b]Near poor refers to reporting units with annual incomes between the poverty threshold and 125% of the poverty threshold.
[c]Nonpoor refers to reporting units with annual incomes above 125% of the poverty threshold.
Note: Row and column percentages may not sum to totals because of rounding.

5. Permanent Income Models

Rather than using previous year's income to screen for poverty status, an alternative set of questions based on the permanent income theory may provide a more efficient means for identifying the persistent poor. Expected and permanent income models have a long-standing tradition in the economics literature on consumer behavior (Friedman 1975; Modigliani and Brumberg 1954; Mayer 1972). Like Paxson (1992), we define permanent income over a relatively short time horizon as expected income in year t based on the RU's characteristics and resources during year t, where t is defined as the year of the study, 1987. The NMES permanent income model was estimated as a function of various socioeconomic and demographic attributes of the RU and its members including age, family composition, family size, health status, home ownership, race/ethnicity, census division, SMSA status, education, employment status, gender and marital status. The sample size consisted of 12,019 RUs reporting income on the screener and in Round 4 of NMES. Several of the independent variables, such as functional status, health status, race, and education were measured prior to Round 4, while the remaining independent variables indicate characteristics of the reporting unit or reference person, i.e., the person or one of the persons who owned or rented the dwelling unit, as of Round 4.

A general version of this equation is as follows:

$$\ln Y = a_0 + a_1 z_1 + a_2 z_2 + \ln e$$

where

a_j for $j = 0, 1, 2$ are parameters;

z_1 is a vector of socioeconomic and demographic characteristics of the RU and its members;

z_2 is a vector of 2-way interactions between the age, education, and gender/marital status variables in z_1;

Y is income of the RU in 1987;

$\ln e$ is a normally distributed residual term with a zero mean and variance S^2; and

e is a lognormally distributed residual term with a unit mean and variance, $\exp(S^2) - 1$.

The results of the permanent income model (Table 2) show that RUs are likely to have relatively high expected income in 1987 if the residence is owner-occupied and in a large SMSA outside the southern or midwestern regions, and if the reference person is older, white, married, employed, highly educated and in excellent health. Also contributing to relatively high expected income are large-sized families, particularly if they contain at least one person 65 years or older and no children under age 16. Although family incomes tend to decline if anyone in the unit has difficulty with at least one activity of daily living, this effect on family income was not significant.

Over half of the variation in observed family income in NMES is explained by these systematic attributes of expected or normal income. The remainder is attributed to random or nonsystematic effects which produce what is usually labeled as "transitory income." In any given year there may be households who are above the official poverty threshold because of a typical "windfall" income, but who would "normally" be below poverty based on household characteristics correlated with their normal or expected income. Thus, families may escape poverty either because of a change in characteristics associated with their expected income, or because of an income windfall, that is, a

Table 2.   Permanent income model: Linear regression coefficients predicting the natural log of family income in 1987

Independent variables	Coefficient
Intercept	8.13**
LN age (reference person)[a]	0.17**
Own home	0.32**
RU composition	
LN RU size	0.54**
At least 1 child, age <6	−0.34**
At least 1 child, age 6–15	−0.28**
At least 1 person age 65+	0.09**
Activities of daily living (ADL)[a]	
No limitations	—
1 limitation	−0.06
2+ limitations	−0.02
Activities of daily living: other RU member	
No limitations	—
1 limitation	−0.09
2+ limitations	−0.01
Health status[a]	
Excellent	0.10**
Good	0.03
Fair	−0.07*
Poor	−0.18**
Not ascertained	—
Race/ethnicity[a]	
White, non-Hispanic	—
Black, non-Hispanic	−0.20**
Hispanic	−0.17**
Census division	
New England	0.06
Mid Atlantic	−0.03
East North Central	−0.05*
West North Central	−0.07*
South Atlantic	−0.04
East South Central	−0.21**
West South Central	−0.18**
Mountain	−0.09**
Pacific	—
SMSA size	
1 of 19 largest SMSAs	0.24**
Other SMSA	0.13**
Not an SMSA	—
Education[a]	
Less than 8 yrs/missing	—
Some high school	−0.40
High school graduate	0.09

Table 2. Continued

Independent variables	Coefficient
Education (continued)	
Some college	0.15
College graduate	0.77**
Post-graduate work	−0.11
Gender/marital status[a]	
Nonmarried male	−0.42*
Nonmarried female	−0.20
Married RUs	—
Employment status	
Reference person employed	0.46**
Other RU members employed	0.30**
Age × education	
Age ∗ <HSG	0.15*
Age ∗ HSG	0.06
Age ∗ some college	0.07
Age ∗ college grad	−0.05
Age ∗ grad school	0.20*
Age × marital status/gender	
Age ∗ nonmarried male	0.08
Age ∗ nonmarried female	−0.01
Education × marital status/gender	
<HSG ∗ nonmarried male	−0.04
<HSG ∗ nonmarried female	−0.07
HSG ∗ nonmarried male	0.08
HSG ∗ nonmarried female	−0.05
Some college ∗ nonmarried male	0.07
Some college ∗ nonmarried female	−0.04
College grad ∗ nonmarried male	0.24**
College grad ∗ nonmarried female	0.04
Graduate ∗ nonmarried male	0.20**
Graduate ∗ nonmarried female	0.21**
R^2	0.53**

Source: National Medical Expenditure Survey.
*$p < 0.05$, **$p < 0.01$.
[a]Reference person is defined as one of the persons who owns or rents the dwelling unit. Person level characteristics, unless otherwise noted, indicate characteristics of the reference person.

positive-valued change in their transitory income.

6. Comparing Alternative Screening Procedures

As demonstrated above, the use of a single question to screen for income, if the NMES experience is typical, is inefficient. Over half of the cases sampled at the time of the screener as "below poverty" were categorized as being *above* the poverty line for the survey year of interest. If reported prior-year poverty status in response to a single income question is a poor way to screen for poverty status, will the use of a

permanent income model approach provide a more efficient means for identifying households likely to be poor in the study year?

To address this issue, two separate logistic regression models were estimated for the probability of an RU being in poverty during 1987, the target group for oversampling purposes. The first model specification is nearly the same as that of the permanent income model (except for the omission of the interaction terms for age, education, gender, and marital status) and can be interpreted as a model of the probability of having expected income in 1987 below the poverty threshold. The second model contains only a single explanatory variable indicating income below poverty at the time of the screener.

Comparing the results from these two models shows that the likelihood ratio index value (Kmenta 1986) for the permanent income model is 10 percentage points (50%) greater than the value for the model containing a single income question (Table 3). This comparison suggests that screening in year $t-1$ on the characteristics associated with a low value of an RU's expected income could prove to be a more efficient way to identify the poor in year t than by screening on their poverty status in year $t-1$ based on a single income question.

One concern with the use of our permanent income model as a predictor of 1987 income is that several of the variables were measured coterminously with income (e.g., age, home ownership, marital status, and employment). However, when the model was reestimated using characteristics as measured at Round 1 (January through April 1987) other than home ownership, there was no effect on the direction or the significance of the coefficients. Although not an ideal means by which to improve the model, since the Round 1 variables

were still measured during the calendar year of interest with respect to income, and while some efficiency may be lost related to the measurement of these characteristics in year $t-1$ as opposed to year t, the findings suggest that the measures identified in the model are robust predictors.

What gains in efficiency are realized using the permanent income model variables as a screening procedure? Using an estimated probability of 0.6 that an RU will be poor (based on the coefficients in Table 3), 65% of those RUs predicted to be poor during the survey year were in fact poor during 1987. If in fact RUs were screened and oversampled on the characteristics associated with low expected income (and a relatively high probability of being in poverty), this result suggests that as many as 65% of these units would be below poverty in the survey year. This is an efficiency rate over 20 percentage points greater than the 44% rate achieved using the reported screener income measure, where efficiency rate is defined as those identified in the screener as poor who were also poor in the survey year. If both demographic characteristics *and* the screener reported income are used for selection, the percentage of RUs predicted to be poor in the survey year rises to almost 71%, reducing the inefficiency (that is, the percentage of screener poor who are not poor in the survey year) observed in NMES from 56% to 29%.

The cutpoint of 0.6 in the analysis above reflects the estimated probability based on the coefficients from Table 3 that an RU is below poverty. The choice of 0.6 is arbitrary, but note that the efficiency of the sampling changes as the cutpoint and the estimated probability are changed. At 0.5 estimated probability, the efficiency declines to 60%, whereas at an estimated probability of 0.7, it increases to 67%.

Table 3. Estimates of the logistic equations for the probability of an RU being poor in 1987 (The permanent income and single question models)

Independent variables	Coefficient
Permanent income model	
Intercept	4.62**
LN age (reference person)[a]	−1.13**
Own home	−0.93**
RU composition	
LN RU size	−0.19
At least 1 child, age <6	0.53**
At least 1 child, age 6–15	0.59**
At least 1 person age 65+	−1.18**
Activities of daily living (ADL)[a]	
No limitations	—
1 limitation	0.19
2+ limitations	0.16
Activities of daily living: other RU member	
No limitations	—
1 limitation	0.44
2+ limitations	−0.18
Health status[a]	
Excellent	−0.35*
Good	−0.25
Fair	0.04
Poor	0.40*
Not ascertained	—
Race/ethnicity[a]	
White, non-Hispanic	—
Black, non-Hispanic	0.93**
Hispanic	0.66**
Census division	
New England	−0.40
Mid Atlantic	0.10
East North Central	0.10
West North Central	0.20
South Atlantic	0.18
East South Central	0.73**
West South Central	0.63**
Mountain	0.42*
Pacific	—
SMSA size	
1 of 19 largest SMSAs	−0.58**
Other SMSA	−0.48**
Not an SMSA	—

Table 3. Continued

Independent variables	Coefficient
Education[a]	
Less than 8 yrs/missing	—
Some high school	-0.50^{**}
High school graduate	-1.05^{**}
Some college	-1.24^{**}
College graduate	-2.14^{**}
Post-graduate work	-2.23^{**}
Gender/marital status[a]	
Nonmarried male	0.07
Nonmarried female	0.46^{**}
Married RUs	—
Employment status	
Reference person employed	-1.19^{**}
Other RU members employed	-1.28^{**}
Likelihood ratio index	0.30^{**}
Single income question model	
Intercept	-2.75^{**}
RU classified as below poverty at screener	2.63^{**}
Likelihood ratio index	0.20^{**}

$^{*}p < 0.05$, $^{**}p < 0.01$.
[a]Reference person is defined as one of the persons who owns or rents the dwelling unit. Person level characteristics unless otherwise noted indicate characteristics of the reference person.

Another alternative survey strategy could have been to collect more detailed information in the screener about the 1986 income and poverty status of the RU. Apart from the additional time and dollar cost, as well as the threat of nonresponse, we note that even if more accurate 1986 income data were collected, such an approach does not prevent a low efficiency rate because a relatively high percentage of those RUs below poverty in 1986 would still be expected to escape poverty by the following survey year. At the risk of achieving a poverty sample weighted toward the permanently poor, the expected income approach to screening avoids this problem and should produce a higher efficiency rate at identifying the poverty population for the survey year of interest.

7. Conclusion

Data from the 1987 National Medical Expenditure Survey show that an unexpectedly large percentage of households and individuals moved from below poverty as reported in a screener interview in the fall of 1986 to above poverty thresholds as reported for the 1987 NMES survey year. Our findings suggest that this misclassification problem was due to both the dynamic nature of poverty status and to the method used for collecting income information on the screener interview, a single RU-level income question which tends to produce underreported income and overreported poverty status.

Because of the rather sizable flows out of poverty that are ordinarily expected

between successive years, we do not recommend survey methods designed to collect more accurate income and poverty status on a screener. Instead, a potentially more efficient way of assuring that the survey-year poor will be oversampled would be to oversample on the basis of characteristics associated with low levels of expected income that could be identified in a screener interview. Indeed, the 1986 NMES screener did in fact collect information on gender, marital status, presence of children, and geographic location of the RU and its members. Additional variables correlated with expected income such as education, employment status, health status, and housing tenure could also be collected in a screener interview without placing an undue burden on the respondents and thereby provide an efficient alternative for sampling the poor and near poor populations.

8. References

Duncan, G.J. (1984). Years of Poverty, Years of Plenty. Ann Arbor, MI: Institute for Social Research.

Friedman, M. (1975). A Theory of the Consumption Function. Princeton, NJ: Princeton University Press.

Groves, R.M. and Magilavy, L.J. (1981). Increasing Response Rates to Telephone Surveys: A Door in the Face for Foot in the Door. Public Opinion Quarterly, 1, 346–358.

Kmenta, J. (1986). Elements of Econometrics (2nd ed.). New York: Macmillan Publishing Company, p. 556.

Mayer, T. (1972). Permanent Income, Wealth, and Consumption. Berkeley, CA: University of California Press.

Modigliani, F. and Brumberg, R. (1954). Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data, in Kenneth Kurihara, Post Keynesian Economics. Brunswick, NJ: Rutgers University Press, pp. 388–436.

Paxson, C.H. (1992). Using Weather Variability to Estimate the Response of Savings to Transitory Income in Thailand. The American Economic Review, 82, 15–33.

U.S. Bureau of the Census (1979). Vocational School Experience: October, 1976. Current Population Reports Series P-70, No. 343, Washington, D.C.: U.S. Government Printing Office.

U.S. Bureau of the Census (1989). Transitions in Income and Poverty Status: 1984–85. Current Population Reports, Series P-70, No. 15-RD1, Washington, D.C.: U.S. Government Printing Office.