AN EVALUATION OF THE SAMPLING ALGORITHMS IMPLEMENTED IN CAPI FOR THE NATIONAL MEDICAL EXPENDITURE SURVEY - 3 FIELD PRETEST

Mamatha S. Pancholi, Steven B. Cohen, Agency for Health Care Policy and Research
Mamatha S. Pancholi, AHCPR, 2101 E. Jefferson St. Suite 500, Rockville, MD 20852

Key Words: Poverty status, high expenditures, sampling

Any opinions, findings, and conclusions or recommendations expressed here are those of the author and do not necessarily reflect the views of the Agency for Health Care Policy and Research.

1. INTRODUCTION

The use of computer software in survey data collection is a growing trend. Within the past decade a number of changes have occurred within the survey field. One important one to focus upon is the introduction and further development of the computer assisted information collection (CASIC) methods. CASIC provides data collection organizations with opportunities in obtaining better quality data at a reduced cost. Casic is broadly defined to include the Computer Assisted Coding and Editing System (CACE), the Computer Assisted Telephone Interviewing System (CATI), the Computer Assisted Personal Interviewing System (CAPI) and other data collection methods. It is more generally defined to be the application of computer technology to the data collection process.

The National Medical Expenditure Survey (NMES) targets the civilian non-institutionalized population living in the United States. It was last fielded in 1987 as NMES-2 and is now being designed and is planned for fielding in 1996 using CAPI software. The purpose of NMES-3 is to allow for the estimation of medical utilization and expenditures for the U.S. civilian non-institutionalization population at a national level.

In any national survey, sampling algorithms are vital to the screening and selection of nationally representative sample of persons. Once the population of interest has been targeted for study, the next task is to determine how one goes about selecting a representative sample of that population. Before computers became such an integral part of the survey interview, the concurrent selection of the sample was performed by the interviewer. Today, the automation of the survey data collection process alleviates this task for the interviewer and can be performed within a CAPI system.

The household component, referred to as FAMES, of NMES-3 is designed to oversample those persons in the target population who are in poverty or who have high medical expenditures as well as those among other policy relevant population subgroups. Due to the dynamic nature of both of these characteristics across time, there have been measurement error associated with them. At the time of screening, a household may be in poverty or have individuals who incur high medical expenditures, but by the time actual data collection begins, this may not be true. Therefore, in order to reduce this problem, two predictive models were developed to help identify those households which have a high probability of being in poverty and those with persons who have a high probability of incurring high medical expenditures.

Prior to fielding a national survey as large as NMES-3, it is typical practice to first implement a Pretest for the national survey. The Pretest is a smaller sample of the same target population and is designed to be as similar to the national survey as possible. In January 1995, the NMES-3 Pretest went forward and data was obtained for the Screener/Baseline of the Pretest. The Screener/Baseline round in the FAMES component of the NMES-3 Pretest was specified to test sampling procedures to be used in the national study (sample selection is concurrent). The performance of the sample algorithms were evaluated, including prediction models for poverty status and high expenditures. Both measures are important criteria for selection into the FAMES sample. This paper presents a summary of the sample design of the NMES-3 FAMES Field Pretest and the sampling approach implemented in CAPI for the FAMES Pretest. Descriptions of the models, results from quality control reviews performed on the models and yields obtained through this method from the Pretest will be presented.

II. NMES-3 Pretest

Sample Design

The FAMES HS field pretest sample design includes selection from area segments and special lists. The segment selections consist of a random sample of households (including group quarters and other non-institutional dwelling units) chosen to represent 11 sampling domains of the civilian non-institutionalized population.

The Pretest is designed to yield a complete series of NMES interviews (through Round 3) with 200

reporting units. A reporting unit is defined to be a person or group of related persons, including unmarried couples living as partners, who are to be interviewed at the same time.

Within selected segments, the selection of households for the pretest was modelled after the approach planned for the national study. After listing dwelling units in each sampled segment, a random sample of households was selected for screening. The sample of dwelling units per sampled segment averages 10 dwelling units per sampled segment. A subsample of households which successfully completed the Screener questionnaire was selected for participation in the FAMES HS Pretest. The decision to select this subsample of households was based upon the demographic characteristics of the residents of the household and was made immediately after completing the screening interview.

The FAMES HS Screener/Baseline Pretest field tested the sampling procedures planned in sampled segments for the national survey. The sampling algorithm classified individuals in each pretest RU according to their allocation with respect to 11 demographic groups:

1. Black/ non-Hispanic, 65 years and over
2. Black/ non-Hispanic, < 65 years
3. Whites and other races, 18-64 years of age, predicted to have high medical expenditures
4. Hispanic, 65 years and over
5. Hispanic, < 65 years
6. Whites, 80 years and over
7. Whites 65-79 years
8. Whites, < 65 years and at below poverty and near poor
9. Other races, at or below poverty and near poor
10. Other races, other income
11. White < 65 years, other income

These domains are ranked hierarchically with respect to their sampling rates for the pretest. The domains identify household with at least one member having the specified demographic characteristic. They are ordered based upon the demographic characteristics of the "highest priority" individual. Each screened household was classified into only one out of the eleven demographic domains. Pretest sample selection rates varied from a low of 0.75 for the white, < 65, other income group to a certainty selection for the black,non-Hispanic , 65 years and older group.

Sampling rates were established and sampling algorithms implemented to test sampling procedures. Once this demographic information was entered into CAPI for each DU member, the computer software informed the interviewer whether or not the DU was to be retained for the Baseline questionnaire.

The CAPI program also included an imputation procedure which was to be applied to DUs with incomplete screener data. This would thereby eliminate missing data and facilitate sample selection. However, for methodological purposes, imputation flags were created to identify those cases for which an imputation was performed.

Using Predictive Models for Domain Assignments

Since a reporting units poverty status classification in 1996 will be unknown at the time of the administration of the HS Screener interview (fall 1995), a prediction model will be used to determine whether a household is to be oversampled (Mathiowetz and Moeller). More specifically, a logistic regression model has been developed that estimates the probability that a reporting unit will have a family income less than 1.25 times the poverty level in a subsequent year based on the poverty status classification and other predictive measures obtained during the screening interview. Households with predicted probabilities above a certain threshold value will be oversampled. In addition to facilitating an oversample of individuals with family incomes less than 125 percent of the poverty level, use of this prediction model will facilitate an oversample of individuals with family incomes less than 200 percent of the poverty level.

The results listed below were observed based on an evaluation of the models performance at the Reporting Unit level, using data from NMES2, and using a predicted probability of .3 or greater (derived from the logistic regression prediction model) as the criterion to target reporting units most likely to have members with family income less than 200 percent of the poverty level in 1996:

1. Based on the NMES2 experience, the expected prediction rate for true positives (family income less than 200 percent of the poverty level) is 83.1 percent among the 19.5 percent of reporting units predicted to have members with family income less than 200 percent of the poverty level.

2. The expected prediction rate for false negatives is 17.1 percent among the 80.5 percent of reporting units predicted to be other income and with members under the age of 65.

Furthermore, among the 30 percent of reporting units with family income less than 200 percent of the poverty level, 54 percent were predicted to have

members with family income less than 200 percent of the poverty level. Alternatively, among the 70 percent of reporting units with family income above 200 percent of the poverty level, 95.3 percent were predicted to have members with family income above 200 percent of the poverty level.

The logistic regression model under consideration is specified at the reporting unit level and requires data on the following measures obtained in the screening interview:

1. Age of reference person;
2. Home ownership;
3. Reporting Unit size;
4. Whether children of specific ages (<6, 6-15) are present in the RU;
5. Whether someone in the RU other than the reference person is at least 65 years of age;
6. Health status of reference person;
7. Race/ethnicity of reference person;
8. Census Division;
9. MSA status of PSU;
10. Education of reference person;
11. Marital status and gender of reference person;
12. Whether reference person or spouse was employed in the previous 3 months;
13. Whether the family income of the reporting unit as less than 1.25 times the poverty level;
14. Whether anyone in the RU was covered by Medicaid.

A sample selection monitoring system was developed which was directly tired to the transmission of data collected in the Screener interview through the CAPI mode of data collection. A system will also be developed to allow for a modification to the sample selection rates during the screening field period of the main study. The system that will allow for a modification in the selection rates during the field period for the screening interview, providing new sampling rates (or different threshold probabilities for the model) that are to be electronically transmitted from a central location to the interviewers' personal computers. This will result in the new sampling rate replacing the original rate (or new threshold probabilities replacing the original threshold probability of .3) in the CAPI selection program for remaining screening interviews.

Also among the groups to be oversampled in the main survey are individuals between the ages 18-64 who are predicted as likely to incur high medical expenditures in the subsequent year. An individuals medical care expenditures in a future year will be unknown at the time of the administration of the HS Screener interview (fall 1995); therefore, a prediction model based on NMES2 data will be used to determine whether a household is to be oversampled as part of the high medical expenditures group because one or more of the family members are expected to incur high medical expenditures in the subsequent year. More specifically, a logistic regression model has been developed that estimates the expected probability an individual who is between the ages of 18-64 will incur high medical expenditures (top 15 percent of the health expenditure distribution) in a subsequent year based on predictive measures obtained during the screening interview. Households with at least one such person with a predicted probability above a certain threshold value will be oversampled. The group is restricted to individuals who are between the ages 18-64, since the persons 65 or older are separately targeted to be oversampled.

The logistic regression model is specified at the person level and requires data on the following measures obtained in the screening interview:

1. Gender;
2. Health status;
3. Marital status;
4. Poverty status;
5. Whether the person lives alone;
6. Age;
7. Whether the person's health keeps him/her from working at a job, doing work around the house or going to school;
8. Whether the person is unable to do certain kinds or amounts of work, housework, or schoolwork because of his/her health;
9. The number of visits to a medical doctor or other medical care provider the person has had during the last 6 months;
10. The number of times prescribed medicines were purchased or obtained for the person's use in the last 6 months;
11. Census Division; and
12. MSA status of PSU.

The results listed below were observed based on an evaluation of the models performance at the individual level, using data from NMES2, and using a predicted probability of .4 or greater (derived from the logistic regression prediction model) as the criterion to target individuals who are between the ages 18-64 and considered likely to incur high medical expenditures in the subsequent year:

1. Based on the NMES2 experience, the expected prediction rate for true positives is 37.7 percent among

the 14.1 percent of individuals in reporting units with members between the ages 18-64 who are predicted to incur high medical expenditures in the subsequent year. It should be noted that when restricting the evaluation to the subset of individuals (8.1 percent) that are predicted to incur high medical expenditures, the expected prediction rate for true positives is 65.3 percent. 2.The expected prediction rate for false negatives is 11.3 percent among the 85.9 percent of individuals in reporting units with members between the ages 18-64 who are predicted to not incur high medical expenditures in the subsequent year.

III. Testing the Sampling Algorithm

A. General procedures to test CAPI outlined in literature

The use of CAPI and other CASIC systems in the survey environment complicates an already difficult data collection operation. This computerization approach towards data collection inherently requires a great amount of system testing. The testing process also seems to follow two paths. One path deals with the mechanics of the software. Are there any programming flaws that freeze or terminate the application? This type of review is usually performed by those who develop the software. The other focuses upon the operational aspect of collecting the data accurately. It is this aspect that will be focused upon within the context of this paper.

The alternative to using a CAPI sampling algorithm would be to have interviewers manually select the sample. This method was used in the 1987 NMES-2. Manually selecting the sample could potentially become a difficult task depending upon the complexity of the sample design. Since the sampling task is now subject to human error as well as time constraints, it is usually preferable to keep the sampling task as simple as possible.

B. Procedures used to test the CAPI sampling algorithm

The data from the NMES-3 Screener Baseline Pretest was used to evaluate the sampling algorithm developed for the national survey. This evaluation occurred in several ways. Frequencies were first generated on all variables to determine if any missing values were present. Specific attention was paid to the variables used in the key sampling variables and imputation flags.

The second stage of this evaluation process focused upon the prediction models developed for predicting people with high medical expenditures and for predicting reporting units in poverty. This process entailed computing the log odds of having high medical expenditures for select persons and the log odds of being poverty for select reporting units. These calculations were compared to the corresponding values generated by the CAPI algorithm. The beta coefficients from both models developed by Mathiowetz and Moeller were used in this computation. The following are equations used for the high medical expenditure model and the poverty status model respectively

For selected persons in the Screener/Baseline, Y_h was computed where

$$
\begin{aligned}
Y_h = & - 3.0638 + 0.3099*\text{female} + 0.1522*\text{goodhlth} \\
& + 0.4209*\text{fairhlth} + 0.9791*\text{poorhlth} \\
& + 0.2949*\text{neng} - 0.1754*\text{mdatl} - 0.2004*\text{enctrl} \\
& - 0.2227*\text{wnctrl} - 0.1413*\text{satl} \\
& - 0.4253*\text{esctrl} - 0.3468*\text{wsctrl} - 0.1175*\text{mtn} \\
& + 0.2794*\text{lgstmsa} - 0.005464*\text{otmsa} \\
& + 0.2065*\text{married} - 0.1320*\text{widow} \\
& + 0.2809*\text{divorce} - 0.4241*\text{poor} + 0.1553*\text{alone1} \\
& - 0.07502*\text{hage1} - 0.2946*\text{hage2} - 0.1458*\text{hage3} \\
& + 0.2680*\text{wrklim1} + 0.2611*\text{wrklim2} \\
& + 0.5545*\text{amb1a} + 1.3778*\text{amb1b} + 1.9357*\text{amb1c} \\
& + 0.3427*\text{pmed1a} + 0.8677*\text{pmed1b} \\
& + 1.6338*\text{pmed1c} + 1.5884*\text{pmed1d}.
\end{aligned}
$$

For selected reporting units, Y_p was computed where

$$
\begin{aligned}
Y_p = & - 0.8513 + 0.2761*\text{age1} + 0.3584*\text{age2} \\
& + 0.3264*\text{age3} - 0.6533*\text{owns} - 0.9212*\text{lnrusize} \\
& + 0.9651*\text{kids05} + 0.8397*\text{kids615} - 0.2264*\text{ad65} \\
& + 0.1396*\text{ghlth} + 0.3628*\text{fhlth} + 0.7630*\text{phlth} \\
& + 0.2478*\text{black2} + 0.3627*\text{hisp2} - 0.6951*\text{neng} \\
& + 0.0518*\text{mdatl} - 0.1083*\text{enctrl} - 0.2474*\text{wnctrl} \\
& + 0.0536*\text{satl} + 0.3920*\text{esctrl} + 0.4356*\text{wsctrl} \\
& + 0.3788*\text{mtn} - 0.3929*\text{msa1} - 0.4446*\text{msa2} \\
& - 0.2260*\text{shschl} - 0.5898*\text{hschlg} - 0.7915*\text{scllg} \\
& - 1.4499*\text{cllgg} - 1.2503*\text{grad} - 0.01161*\text{mnmarr} \\
& + 0.2547*\text{fnmarr} - 1.1763*\text{hsemply} \\
& + 1.8170*\text{scrnpov} + 0.3676*\text{mcaid};
\end{aligned}
$$

These values were then used in the transformation: $p = e^y/1+e^y$ (i.e. the log odds) and were then compared to the log odds outputted by the CAPI sampling algorithm. Note that Y_h is a person level measure while Y_p is a family level measure.

The third stage of this evaluation focused upon the sampling domains which were previously outlined. The CAPI algorithm was reprogrammed in SAS by an

individual other than the original CAPI programmer. Frequency of dwelling units by domain obtained through CAPI were compared to the SAS domain classification of these DUs.

IV. Results of testing procedures

Upon reviewing the log odds for both models it was immediately evident that the values obtained through CAPI and SAS were identical (Table 1). There were no discrepant cases found. The results from reclassifying the domains by another program are presented in Table 2.

Table 1. Results from Quality Control on High Medical Expenditure and Poverty Status Models

	High Medical Expenditure	Poverty Status
% Match	100%	100%
% No Match	0%	0%

Table 2. Results from Quality Control on Domain Classification

Domain	#DUs Classified w/out CAPI	#DUs Classified by CAPI
Black/ non-Hispanic, 65 years and over	9	9
Black/ non-Hispanic, < 65 years	42	45
Whites and other races 18-64 years of age, predicted to have high medical expenditures	59	59
Hispanic, 65 years and over	2	2
Hispanic, < 65 years	15	15
Whites, 80 years and over	9	9
Whites 65-79 years	36	36
Whites, < 65 years and at below poverty and near poor	20	20
Other races, at or below poverty and near poor	3	3
Other races, other income	3	3
White <65 years, other income	113	113
Total	311	314

V. Summary/Conclusion

The results from the evaluation of the CAPI sampling algorithm provide confidence as to the reliability of the CAPI algorithm in the NMES-3 Pretest experience. The values obtained through the CAPI algorithm and SAS algorithm for the log odds in the poverty status and high medical expenditure models were identical. The domain classification results indicate no difference between the two methods used. The only discrepancy is due to 3 missing odus in category 2. These were empty odus and hence did not link properly during analysis. Note that the totals differ by these 3 cases.

The increase in use of CASIC systems to perform data collection operations will require the survey managers and methodologists to further develop standard quality control procedures for assessment of these computerized instruments. This is particularly important for the sample selection component of the instrument.

It must also be noted that given the dynamic nature of government sponsored survey organizations, the NMES-3 that was originally planned for 1996 has been redesigned in hopes of integrating government surveys. With that in mind, this algorithm will not be used in 1996 but will be modified for the 1997 redesigned NMES-3. In 1996, the NMES-3 sample will be selected from the 1995 HIS sample. In 1997, the sample will be drawn from the 1996 HIS sample. The 1996 HIS will be fielded with a paper/pen instrument, not a CAPI one. However, for the 1996 NMES-3, a CAPI instrument will still be fielded.

BIBLIOGRAPHY

National Medical Expenditure Survey Screener/Baseline OMB Clearance Package, No. 0935-0093.
United States Public Health Service, 1995.

Moeller, John F. and Nancy A. Mathiowetz, Problems of Screening for Poverty Status. Presented at the American Statistical Association Annual Meeting, August 1990.