# EVENT HISTORY ANALYSIS OF PROXY RESPONSE PATTERNS AND HEALTH CARE COVERAGE AMONG THE AGED: DATA FROM THE 1984 SURVEY OF INCOME AND PROGRAM PARTICIPATION PANEL

Larry S. Corder and Kenneth G. Manton, Duke University
2117 Campus Drive, Durham, NC 27706

KEY WORDS: Survey research, Grade of Membership, proxies

## INTRODUCTION

Proxy reporting, the act of interviewing a person about the characteristics and behavior of (an) other person(s) in a household, has long been an important part of both list samples of individuals and area probability samples of dwelling or households in the United States during a rapid period of growth in such data collection activities. Both cross-sectional and longitudinal study designs currently employ proxy reporting as a practical necessity associated with budget, field period, and maintenance of interviewing staff (e.g., Health Interview Survey, Current Population Survey, National Long Term Care Survey). However, experience with research on the use of proxy reports suggests that: 1) surveys have not employed standard proxy reporting rules and 2) no case control study of proxy reporting has been conducted (Moore, 1988). Rather, the most recent review of the limited literature on proxy reports clearly indicates that the preponderance of the research which have been carried out suffers from observational confounding or selection bias where the accuracy of reports is the research issue. We would expect that any research contemplated on proxy reports on a post hoc basis would encounter substantial accuracy and validity problems. Further, the population of interest, either from the study design or the selection of a particular subpopulation offers the opportunity for different or more focused results to emerge, based on analysis designed to address specific research questions, e.g., accuracy of Medicaid reporting among the aged. For example, a recent methodologically sophisticated publication dealt with the bias associated with proxy reporting among males 25-44 years of age in the labor force during three waves of the 1894 Survey of Income and Program Participation (SIPP). Our aim is to examine proxy reporting patterns among aged SIPP survey respondents.

The Survey of Income and Program Participation is a large, complex, sample survey of dwellings in the United States which has added substantially to the availability of information on income from multiple sources, and complete coverage of nonmeans and means tested program participation as its core mission. Conducted in 9 waves, each approximately four months apart, the study is organized into waves and panels according to a modified rotating panel design (Kasprzyk, 1988). Respondent rules in the SIPP require the interviewer to conduct a proxy interview with the spouse. If the spouse is not available, then one of the group of other adults in the household will be interviewed. In the second and subsequent waves, the proxy respondent from the previous wave is given higher priority for the conduct of the proxy interview than the spouse of the respondent.

No definitive evidence of a proxy reporting effect is available. Thus, concern over the possibility of such an effect on important policy related estimates should continue, e.g., out-of-pocket costs for prescription drugs among the aged. New research methods and results presented below illustrate the potential of a new class of analytical tools which may be applied to survey research methods and design problems as well as to the analysis of longitudinal data sets. These methods, Grade of Membership event history analysis, provide an opportunity to examine the pattern of proxy reporting over time (within a SIPP panel), determine the level of clustering of proxy events, and present information on the correlates of the proxy reporting event by identifying "pure" types. Further, these general methods may be employed to examine 1) the predictors of proxy interview status and 2) the time dependent changes in those predictors in an aged population.

### Study Background

The components of sampling and nonsampling error in sample surveys are well known in terms of both items and structure (Kish, 1987). Estimates of specific survey error components and development of the total survey error component have only recently been presented (Horvitz et al., 1988). The continuing discussion in the literature over data quality in general and the potential influence of proxy reporting rules in particular remains conceptual in part. The discussion is derivative or consequent to a model of sampling and nonsampling error, e.g., accuracy of reports in surveys have not been consistently estimated (Marquis et al., Hill). While these observations are true of the general civilian noninstitutional population studies that have been completed, little research has concentrated on the problems associated with contacting, locating, or interviewing older persons, particularly the fastest growing segment of the United States population; the oldest-old (85 years of age and older). The literature on proxy reporting as a source of nonsampling error clearly illustrates two issues in terms of the trade off between different types or categories of error (Hertzog and Rodgers, 1988). In the case of older persons, use of a proxy, according to different rules clearly and demonstrably reduces nonresponse. As age rises, both nonresponse and proxy reporting rise in a manner inconsistent with proxy reporting patterns of other population groups, e.g., employed persons or young males living at home. It seems sensible to suggest that proxy reports, particularly by a caregiver and/or spouse concerning an aged person, who may be in frail health or cognitively impaired, may reduce nonresponse in a qualitatively different way than interviewing a proxy because the respondent is absent from the household at a job. Further, interviewing a proxy about an institutionalized aged person who may not usefully be subjected to the interview process because of cognitive impairment or frail health enables the production of civilian population estimates. The assumption that proxy reports will be less accurate than self-reports has not been tested or proven in the aged population for any policy relevant variables of interest. Further, it has not been possible to separate out response quality from self-selection issues. In this study, our aim is to characterize the process of proxy reporting in a longitudinal survey with approximately three years of follow-up which included a large number of aged persons at the first wave (~6,000). Our fundamental view is that proxy reporting rules for this survey of the civilian institutionalized population where we concentrate on the aged group, achieved high levels of response across aged population groups thus minimizing error due to nonresponse associated with age related patterns of cognitive impairment and frail health. In the longitudinal context, proxy reports contribute to continuity across survey waves reducing missing data and the complex of item recall issues associated with various components of nonsampling error. Indeed, among a class of oldest-old, frail interviews, the proxy interview or the assisted interview could be the expected event.

## THE STRUCTURE OF THE SURVEY AND ANALYSIS

SIPP contains a wide variety of information concerning the financial status of sampled individuals and households as well as detailed information about the personal and family characteristics of survey respondents including information about the interview itself. Our study exploits the availability of that information in each of nine waves to create a relatively large data base for the study of older persons in general and survey methodological issues associated with them in particular. While a number of supplements and repeating/nonrepeating questionnaire batteries were administered to the panel during one or more of the waves, only a select proportion of all the available variables from any of the waves are available for all of the waves. For example, a short questionnaire battery on health and disability was included in wave three as part of a routine questionnaire supplement wave rotation. While health, disability, and its influence on proxy reporting among the aged is of interest, our purpose requires measurement of most variables, particularly those which could influence the decision to employ a proxy interview at each wave of the survey or interview point.

While we restrict our domain of interest to the aged survey person, the SIPP offers three distinct advantages for survey methods research on proxy reports. First, a large number of aged persons are available (> 6,000) and they are followed for an extensive period of time (9 waves, $\simeq$ 3 years). Thus, a unit of observation, wave/person, can be developed, dramatically increasing the available n for analysis. A principal concern of the analysis of data from the SIPP for the aged population has been the relatively small number of persons age 85 years and older available (Taeuber, 1986). These survey methods results do not suffer from that problem.

Second, a standard set of data collection methods and questionnaire batteries were administered over the life of the panel. Thus, our restricted sample (persons 65 years of age or older at the first wave) has had a uniform set of proxy reporting rules administered by a group of professional interviewers who tracked changes in family circumstances and composition as well as various reasons for wave nonresponse over an extended time. In this study, it is important to note that once a proxy is chosen, usually in the first interview, that person is the preferred subject for all remaining survey rounds. Further uniform questionnaire batteries were employed as the core element of each wave of interviewing. These questionnaire items and batteries enumerate the household, its characteristics as well as those of its member and the circumstances surrounding the interview. The core questionnaire is intended to repeatedly measure the kind, type and volume of cash and noncash income received from all sources. Further, enrollment, participation, or recipiency within means tested and nonmeans tested programs was assessed at each wave. Indeed, monthly interwave receipt of income and program participation are assessed in this ambitious survey effort. We employ only variables that are available from the core survey instrument consistent with the utilization of wave/person as the unit of analysis.

Third, the SIPP offer an opportunity, through its basic design to study the accuracy of reports for certain policy relevant variables, e.g., Medicare and Medicaid (Short et al.), the characteristics and impact of missing data on estimates before and after imputation which may serve as a separate area of investigation itself (E-M algorithm) and the characterization of the variables in a grade of membership structure which best describe the data (Woodbury) in which events occur over time, one of those events being the proxy report. We concentrate on this task in this paper. Further, we may produce a model, based on the above characterization which predicts proxy response (Manton et al.).

A small number of variables (36) were chosen which characterize the social and demographic composition of households containing an aged person at wave one, e.g., age, race, sex and others which were assessed at each wave, e.g., proxy response. Among these variables were indicators of the circumstances surrounding the interview, the identity of the proxy, if any, and the reason for subsequent nonresponse to the survey. Indeed, survey response status is available for each wave in some detail. Further, a wave nonresponse indicator was available. Also, certain program participation variables concerning Medicare, Medicaid, and private health insurance were included for the purpose at hand. This small number of participation variables are included as indicators of health and economic status as well as indirect accuracy checks on proxy reports (this facet of these variables' behavior will be examined directly in a subsequent paper). Specifically, these small number of variables are the items available to characterize the pattern of proxy response in a direct way. At this point, we do not employ a 'behavioral' model in our characterization of wave/person level proxy reports. Rather, we are defining pure types based on the structure of the data itself with internal variables which define the GoM structure and external variables which do not directly contribute to defining the pure types. In the case of SIPP, a very large number of variables concerning receipt of cash and benefits are available on a monthly or wave basis. Several hundred such variables are available, many with very low affirmative levels, and were used indirectly as external variables to characterize the ideal type structure of the data. The simple behavior models which include proxy reporting in the recent literature, problems of selectivity aside for the moment, must be concerned with finding a person at home and conducting an interview after getting a person to cooperate in that process. Age related levels of labor force participation, cognitive impairment, and physical frailty in terms of health require the construction of highly specific behavioral models of events of interest which include proxy reporting when the aged population is the group of interest. We have reviewed the respondent selection rules which restrict the interview in the selection of a respondent both for themselves and for other who cannot respond for themselves. What literature that exists suggests that the level of item level inaccuracy increases with proxy response but offsetting reduction in nonresponse error not corrected by imputation (a regression based approach) may be useful and practical (Norris). A behavioral model of reporting accuracy of coverage under various public and private health care may be constituted based on a three-stage process. First, wave/person based grade of membership structures or pure types will be identified in the 1984 SIPP panel which will generate indirect evidence on accuracy of certain reports. Second, the predictors of proxy reports will be identified in a grade of membership based regression. Third, a subsequent paper will show how proxy report status and its predictors will be employed in an analysis of the accuracy of Medicare and Medicaid reports in SIPP. Wave/month remains the unit of observation in each of the event histories. Further, we may then produce proxy and nonproxy life tables conditional on survey participation (Manton et al.). We complete the first two research tasks in this paper. While we can estimate the effects of each of the interval variables on proxy reporting status directly across waves with grade of membership and other techniques which employ less information (Hill, 1988), evaluating the accuracy of reports is now and has been a difficult research problem to overcome. In the case of SIPP and the operation of the programs it monitors, accuracy of reporting for both self and proxy may be evaluated against a very good standard for a limited number of means tested and nonmeans tested programs associated

with payment for health services. A model for predicting a coverage event in the survey context, e.g., unemployment insurance is well developed (Heckman and Singer). Medicare is a program that pay for health services for disabled persons and persons 65 years of age and older. Based on coverage under the Social Security Act, the Health Care Financing Administration Office of the Actuary estimates that over 96% of aged persons resident in the United States were covered by Medicare. Once covered by Medicare, few if any persons leave the program except through death. An additional check for this variable is receipt of Social Security income. Medicaid is a state run program with federal participation which pays medical costs for persons in need. It is a means tested recipiency program. The principal means whereby aged persons come to be covered by Medicaid both concern their current income but one of them also concerns their assets. Until the Medicare Catastrophic Coverage Act of 1988, States had substantial discretion in determining how low income was defined. Thus, in 1984, substantial interstate variation in coverage occurred among low income aged persons. Reports and nonreports of coverage are subject to knowledge constraints and a clear negative response set. The aged covered by Medicaid in the noninstitutional population secured that coverage in one of two ways: 1) securing supplement security income at age sixty-five or older consequent to possession of little income and few, if any, assets, 2) spending down to Medicaid, e.g., disposition of assets to pay for living and medical expenses until few, if any, are left and a low income level is achieved. Short et al. reports that few persons report securing Medicaid coverage in this way and remain in the SIPP study universe, e.g., the civilian noninstitutionalized population. Thus, SSI benefit receipt, a program few persons depart from, except through death is, along with poverty level reports, a very good indicator of accuracy of Medicaid coverage reports among the aged. In a subsequent paper we will detail what proxy response reports bias exists among these important health policy indicators in a grade of membership context. In the subsequent methods and results sections, we employ the 1984 SIPP panel, to characterize the determinants of proxy response by identifying pure types in the social, economic, demographic and interview based variables available in each SIPP wave with grade of membership analysis in an event history story context.

## DATA PREPARATION AND PROCESSING

1984 Panel SIPP data was secured from the University of Wisconsin SIPP Access Project. A wide range of social, demographic, economic and interview-related variables were secured from the SIPP Access database in order to avoid main frame processing costs associated with using the large rectangular data files which are part of any complex, longitudinal study. Further, the SIPP Access data file offered the additional advantage of data that had been repeatedly examined edited and corrected for consistency beyond the available public use data files stored in an environment (relational data base structure) where we believed it would be inexpensive and straightforward to extract a machine readable data file for the aged population.

Our expectations proved both true and false. Use of SIPP Access proved inexpensive but communicating at a distance and the size of our data file request (6,182 cases and 2,200 variables) proved to be difficult for the small mainframe used for the SIPP project at Wisconsin to accommodate. In addition, the transaction with Wisconsin required the purchase and installation of software which is not available for operation in a UNIX environment, the primary Center for Demographic Studies operating environment. As the transaction with the Wisconsin relational database was our initial contact, each of the above elements took substantial amounts of time to resolve as each production component was put in place.

The remainder of file processing, including statistical analysis, was carried out without the benefit of a computer center or mainframe. The Center for Demographic Studies has developed a dedicated super-micro computer system which carried out the SIPP file processing, analytical tasks, and associated machine readable data file transfers and storage once a data tape containing the SIPP file arrived from SIPP Access in Wisconsin. When ideally configured, our super-micro-computer network operates in a UNIX environment with an MS/DOS window available to run programs like PC-SAS, which we installed. Several very large (330 mg) hard discs are available with these machines and one is equipped with 800 mg worm technology (Write On Read Mostly) optical storage device. After receipt of the file from Wisconsin, it was read into a series of raw data files for processing as a SAS dataset. We found disc partitioning and file transfer to be time consuming problems in our computing environment and found the necessity to transfer files manually via floppy disc from machine to machine until new software could be installed. The data were stored in a SAS data set on a worm drive or optical disc and processed in an MS/DOS window in a UNIX operating environment as part of a micro computer network. All Grade of Membership analyses were prepared and conducted from this data file which will serve as an essentially no incremental or unit cost environment for unlimited analyses of this data base conducted at speeds (33 mhz) equivalent to a last generation super computer.

Such an approach has proven to be a cost effective way to analyze large survey data files and may be applied to many complex data files (Survey of Income and Program Participation, Long Term Care Survey, Health Interview Survey, National Medical Expenditure Survey) as well as linked secondary data when it is available (Medicare, Medicaid medical use files, and death records) without spending large proportions of project budgets to pay mainframe processing charges.

## METHODS

The analytic procedure to be employed in the proposed analyses is the Grade of Membership procedure adopted to analyze event history data. The principal distinction between the GoM model and other analytic procedures is that the GoM procedure assigns persons to K latent groups where within group heterogeneity can be represented. Let us assume that J discrete response variables, each with $L_j$ response levels, are measured for I individuals. For each response we can define a binary variable, $x_{ijl}$, which adopts the value 1.0 when the *l*th response to the jth variable is observed for the ith person.

We assume that the probability that $x_{ijl} = 1.0$ can be predicted as a function of two types of coefficients. The first type of coefficient represents the probability that a person in the Kth group has the *l*th response to the jth variable. These coefficients, $\lambda_{kjl}$, are the probabilities that help define the nature of the Kth group. The second type of coefficients distinguish this model from standard discrete classification models. This coefficient represents the degree to which the $\lambda_{kjl}$ for the Kth group contributes to the probability that $x_{ijl} = 1.0$. This coefficient, $g_{ik}$, is estimated under the constraints that $0 \leq g_{ik} \leq 1$ and $\sum_k g_{ik} = 1.0$. This determines that the $\lambda_{kjl}$ represent a K-1 dimensional simplex that, except for stochastic error, will bound all observed responses. With these definitions the model can be written

$$\text{PROB}\,(x_{ijl} = 1.0) = \left(\sum_k g_{ik}\,\lambda_{kjl}\right). \qquad (1)$$

The coefficients in (1) are estimated by maximum likelihood procedures. Most simply they may be estimated assuming a multi-nomial model, or

$$L = \prod_i \prod_j \prod_l \left(\sum_k g_{ik}\,\lambda_{kjl}\right)^{x_{ijl}}. \qquad (2)$$

The properties of the coefficients are established by extensions of the Kiefer-Wolfowitz conditions. With these conditions (or their appropriate modification, identifiability is demonstrated differently than in Kiefer-Wolfowitz,Tolley and Manton (1989a)) the $\lambda_{kjl}$ can be shown to be consistent as can up to the Jth order moments of the $g_{ik}$ distribution. Furthermore, with consistency, the likelihood ratio for models of order K and K+ can be shown to yield a critical region of size $\alpha$ from a $\chi^2$ distribution (Tolley and Manton, 1989b).

To use this model for event history analysis, we need to modify the metric of the likelihood function to represent terms. This can be done, for the current case, by defining all measurements made at each wave, at occurring at time t (t = 1, 2, ..., T; where the intervals between measurement is about 4 months). Additionally, we could represent monthly measurements of payment of benefits or income by e. The general likelihood in this case is,

$$L = \prod_i \prod_j \prod_e \prod_t \prod_e \left(\sum_k g_{ik\cdot t\cdot e}\,\lambda_{kjl\cdot t\cdot e}\right)^{x_{ijl\cdot t\cdot e}}. \qquad (3)$$

This function can be used to describe a wide range of multivariate event history models each distinguished by a particular pattern of constraints imposed on t or e. In the current analyses we set $e_1 = e_2 = e_3 = e_4$ and allow $g_{ik\cdot t}$ to vary by wave. The $\lambda_{kjl\cdot e}$ are assumed fixed for t (so that the basic dimension are fixed over time) but that one allow the estimation of certain $\lambda_{kjl\cdot e}$ on a monthly basis (the variables used to define the K variables are assumed fixed over t and e).

The likelihood function can be shown to have optimal weighting for representing sample design effects on parameter variance estimates (Woodbury and Manton, 1985). To recapture the population structure reweighting by sample weights of the MLE of parameter is required.

## RESULTS

The analysis is designed to address two specific areas and provide the framework for several subsequent related research activities. First, we examine the pattern of proxy reporting among the aged noninstitutionalized population in the context of their social, demographic, health and interview related characteristics. No previous multivariate event history analysis with grade of membership techniques has been conducted with this data set. While we first estimate the GoM model for the four main groups of variables, we are particularly concerned with interview related variables (e.g., proxy and wave status and coverage under health insurance programs). When wave/person is employed as the unit of analysis, 57,195 unique observations are obtained. We identified 36 variables that are consistent across waves and describe social, demographic, health care coverage, and interview characteristics germane to our interest in the pattern of proxy reporting. These 36 measures were available for the 6,355 persons 65 and over at wave one. It should be noted that an analysis of this type essentially follows a cohort defined at wave one and is designed to examine the correlates of the events occurring within the cohort, including exit from the population according to reason for exit. Such groups contribute to estimation of GoM statistics and definition of pure types.

Analyses of the 36 social, demographic, health care coverage, and interviewer characteristics were conducted with 4 and 5 pure types. Differences in the likelihood statistics = $(\sqrt{2x2} - \sqrt{2\,n\text{-}1})$ between models each specifying and increasing number of pure types was examined. Beginning with 4 pure types we conducted 6 analyses with up to 10 pure types. The difference between 9 and 10 pure types was not significant while earlier comparisons were highly significant and we thus chose the nine pure type solution for evaluation.

In Table 1 the first column describes the variable and its response level. Then in the second column we present the marginal frequencies of these variables in the integrated samples. The frequencies represent the unweighted counts from the survey. They are used to identify a set of K dimensions that describe the variation of nursing home residents for all survey years, i..e, the K types will define a set of health and functional dimensions general enough to describe the characteristics of individuals in all years. Below, we will use several indicator variables to decompose that variation for the different survey years. In columns 3 to 7 we present the probabilities that each pure type have the specific response to a given variable.

The nine dimensions or pure types provide a clear portrait of what may be considered complex attribute and behavior in the SIPP study. Extreme concentration of program participation, income, and interview related variables in particular pure types is the norm. Demographic variable concentrations consistent with the above findings reinforce the view that the observed pure types are substantively meaningful in and of themselves. Both health care coverage and interview status/characteristics are represented by more than one variable. 98.06% of observations report Medicare. Only pure type VI contains any number of observations where Medicare coverage was not affirmed. In a similar view, Group V contains all observations (wave/person) reporting Medicaid coverage. Next, Groups VII and VIII concentrates all of the observations occurring with a report of private health insurance.[1] When interview status and characteristics are examined, Groups I and VII are both composed of married persons, but no group I persons had proxy interviews while all Group VII persons had a proxy interview conducted for them. 14% of available observations were conducted with a proxy. Two-thirds of all these were conducted with a spouse and all those reside in Group VII. The remaining proxy interviews reside in Group VI and were conducted for widowers and widows. Groups I, II, and VI contain disabled individuals (12% of the total) with Group I entirely composed of such persons. Groups I, II, III, VI and IX did not contain any one who completed all nine waves of the survey but Groups IV, V and VIII contain persons who uniformly completed all the survey rounds. When we examined Group VII, it was split on level of completion, unlike the other Group (VI) composed entirely of proxy interviews. When we examined the variable, reason for leaving the sample universe, it is interesting to note that all deaths fell into Group I while all institutionalized fell into Group IX. This initial description illustrates the complexity of the GoM groupings of pure types for understanding the SIPP data. One could construct a complex multidimensional matrix or Hesse diagram which would place these and other variables in a lattice structure clearly defining the distances between and among pure types and the variables associated with the primary event of interest.

Of the nine pure types, Groups II, VI, and IX had high proportions of the oldest old population. Each of these groups is composed of females. Groups I an IX are very

clearly differentiated on social and demographic groups. Group I is all male and quite young (65-69 years of age). Group IX is all female and 85 years of age or older. Black persons are represented exclusively in Groups IV and V, the groups, which on the whole, have lower levels of educational attainment than the younger whites in Groups I and the older whites in Group IX. In summary, we have identified the demographic characteristics that are most characteristic of each profile or type by comparing and contrasting the response patterns. Group one is white, male, disabled, and very, very, young. Group II is female, tends to be disabled, and maybe young or old. Group III is female, white, has high levels of education, and is in the seventh decade of life. Group IV is black and male as well as poorly educated and in the seventh decade. Group V is female, black, more highly educated, and is also in the seventh decade on the whole. Groups VI is female, proxy reporting mostly Asian, and either very young or very old. Group VII is female proxy reporting, well educated, and in the seventh decade. Group VIII is younger, well educated, white and female while Group IX is very old female and white with a ninth grade education or less.

The nine groups or profiles present the combined results across the nine survey waves and certain characteristics patterns of missingness which are most often excluded from the calculation of the statistics of interest.[2] One item of concern should be noted. Given our definition of the population of interest, observations are evenly distributed across waves.[3] Profiles VI and VII, the groups where all the proxy reports reside, shows concentration of all of the responses in waves 1 to 7. This is most unusual. Further, when we examine the number of persons completing all nine interviews, Group VII, in particular, is not that dissimilar from several other profiles even though the wave/proxy initial analysis would make such a result possible. Our view is that this data is probably not pathological but continued examination of this finding is certainly necessary.

In interpreting these profiles, it is also useful to examine external variables that describe the receipt of cash and noncash benefits from an extensive set of sources as well as coverage under means tested and nonmeans tested programs. Here we can represent the effect of this set of important core SIPP variables on the probability space. We calculate the $\lambda_{kjl}$s for the pure types for external variables not included in the likelihood functions. While most of the income receipt and program coverage variables are organized by month within wave, e.g., m = 1 equates to the source of data across waves in that month of reference. Several are organized by wave. Further, it is clear that the greatest number of these variables are relevant to very few aged persons indeed. In fact, several variables record no survey affirmative responses in any wave. We can discuss accuracy of reports of social security payment, Medicare coverage, supplemental security payment and Medicaid coverage. Briefly, all SSI payment observations are in Group V as are all reports of Medicaid coverage across each of the waves, a remarkable concurrence. In Groups V and VI uniform reports of Medicare coverage do match reports of social security receipt while in Group IV splits occur in each group among these proxy reports. Other variables of interest include pension reports of coverage in Groups IV, VII and VIII with the highest level in Group IV while private health insurance is concentrated on Groups VII and VIII. We should also note food stamp receipt is most heavily represented in Group V.

## DISCUSSION

Here, we present an initial event history based grade of Membership analysis that 1) strongly differentiates aged civilian noninstitutionalized persons according to their health care coverage, interview and demographic characteristics. We find strong patterns of association between proxy reports and indications of the accuracy of Medicare and Medicaid coverage reports as well as an indication that further investigation of patterns of wave nonresponse by profiles generated herein is most certainly merited. We further confirm the strength of the profiles to differentiate this aged population over time by reference to external variables which, while on the whole are not useful based on low levels of response, specifically continue over accuracy of reporting issues as well as reinforce the specific profiles (e.g., food stamps).

On the basis of these results we can confirm that the noninstitutionalized population is quite heterogeneous in a manner consistent with other longitudinal analyses that had health and disability measures available for information (Manton, 1988). Our findings indicate that this heterogeneity extends to the circumstances of the interviews and indicates the potential for estimating reduction in bias associated with reducing nonresponse with proxy interviews as well as estimating increase in accuracy due to such rules.

## FOOTNOTES

1. Wave nonresponse was routinely excluded, however, a small number of variables with very high levels of missing values were included in the preliminary data analysis step. Also, following the file documentation, appropriate not applicables according to skip pattern were converted to no responses.
2. Adjustment for well known rotation group problems were made as the file was being constructed.
3. It should be noted that substantial amounts of wave nonresponse and inputation occurred with the SIPP data set. Given the use of wave/person in the event history context, both wave and item nonresponse patterns may be identified.

## REFERENCES


Heckman JJ and B Singer. 1984. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica* 52:271-320.

Hertzog AR and WL Rodger. 1988. "Age and Response Rates to Interview Sample Surveys." *Journal of Gerontology* 43(6):5200-5205.

Hill DH. 1987. "Response Errors in Labor Surveys: Comparisons of Self and Proxy." SIPP Working Paper No. 8706, Bureau of the Census, Washington, DC.

Horvitz DG et al. 1987. "The Use of Standards in Survey Estimation." Proceedings of the Section on Survey Research Methods, American Statistical Association, Washington, DC, pp. 546-551.

Kasprzyk D. 1987. "Survey of Income and Program Participation Update." SIPP Working Paper No. 8722. Bureau of the Census, Washington, DC.

Kish L. 1987. *Statistical Design in Research.* New York: John Wiley.

Manton KG. 1988. "A Longitudinal Study of Functional Change and Mortality in the United States." *Journal of Gerontology: Social Sciences* 43(5):153-161.

Manton KG et al. 1987. "Grade of Membership Techniques for Studying Complex Event History Processes with Unobserved Covariates." In *Sociological Methodology, 1987*, C Clogg, Ed., pp. 309-346. San Francisco: Jossey-Bass.

Marquis KH et al. 1981. "Response Errors in Sensitive Topic Surveys." Santa Monica: Rand Corporation.

Moore JC. 1988. "Self/Proxy Response Status and Survey Response Quality: A Review of the Literature." *Journal of Official Statistics* 4(2):155-172.

Norris FH. 1985. "Characteristics of Older Nonrespondents over Five Waves of a Panel Study." *Journal of Gerontology* 40(5):627-636, 1985.

Short PFS et al. 1988. "The Dynamics of Medicaid Enrollment." SIPP Working Paper No. 8809, Bureau of the Census, Washington, DC.

Tolley HD and KG Manton. 1989a. "Distribution of test statistics for fuzzy partition models." In review at *Biometrika.*

Tolley HD and KG Manton. 1989b. "Large sample properties of a fuzzy partition." In review at *Journal of the American Statistical Association.*

Woodbury MA and KG Manton. 1985. "A continuous-time multivariate Gaussian stochastic process model of change in discrete and continuous state variables." *Sociological Methodology 1985* (Tuma N, Ed.). Jossey-Bass, pp. 277-315

Table 1: Nine Pure Types (K=9) Grade of Membership Analysis of 36 Social, Demographic, Health Care Coverage, and Interview Characteristics Dimensions in the 1984 SIPP Panel (Internal Variables)

| | Frequency | I | II | III | IV | V | VI | VII | VIII | IX |
|---|---|---|---|---|---|---|---|---|---|---|
| Sex | | | | | | | | | | |
| Male | 40.03 | 100.00 | 0.00 | 0.00 | 100.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Female | 59.70 | 0.00 | 100.00 | 100.00 | 0.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| Race | | | | | | | | | | |
| White | 89.91 | 99.47 | 60.56 | 89.45 | 37.67 | 100.00 | 20.15 | 100.00 | 100.00 | 100.00 |
| Black | 8.78 | 0.00 | 38.79 | 10.44 | 41.08 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Indian | 0.17 | 0.53 | 0.65 | 0.11 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Asian | 1.13 | 0.00 | 0.00 | 0.00 | 21.24 | 0.00 | 79.85 | 0.00 | 0.00 | 0.00 |
| Highest Grade Completed | | | | | | | | | | |
| 0 | 2.14 | 0.00 | 0.00 | 0.00 | 14.19 | 17.37 | 0.00 | 0.00 | 0.00 | 0.00 |
| 1 | 0.78 | 0.00 | 0.00 | 0.00 | 5.55 | 4.72 | 0.00 | 0.00 | 0.00 | 0.00 |
| 2 | 1.62 | 0.00 | 0.00 | 0.00 | 12.00 | 7.71 | 0.00 | 0.00 | 0.00 | 0.00 |
| 3 | 2.28 | 0.00 | 0.00 | 0.00 | 16.11 | 14.23 | 0.00 | 0.00 | 0.00 | 0.00 |
| 4 | 3.00 | 0.00 | 0.00 | 0.00 | 22.98 | 11.33 | 0.00 | 0.00 | 0.00 | 0.00 |
| 5 | 3.47 | 0.00 | 0.00 | 0.00 | 29.17 | 1.94 | 0.00 | 0.00 | 0.00 | 0.00 |
| 6 | 5.74 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 100.00 | 0.00 | 0.00 | 0.00 |
| 7 | 7.22 | 22.60 | 19.55 | 0.00 | 0.00 | 1.40 | 0.00 | 0.00 | 0.00 | 18.20 |
| 8 | 20.21 | 55.53 | 47.04 | 0.00 | 0.00 | 0.00 | 0.00 | 13.32 | 0.00 | 59.30 |
| 9 | 8.70 | 21.87 | 30.15 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 2.11 | 22.77 |
| 10 | 7.59 | 0.00 | 3.26 | 17.79 | 0.00 | 10.07 | 0.00 | 12.87 | 15.34 | 0.00 |
| 11 | 5.65 | 0.00 | 0.00 | 12.36 | 0.00 | 9.97 | 0.00 | 9.70 | 12.20 | 0.00 |
| 12 | 31.61 | 0.00 | 0.00 | 69.86 | 0.00 | 21.27 | 0.00 | 64.11 | 70.35 | 0.00 |
| Disabled | 12.16 | 100.00 | 55.75 | 0.00 | 0.00 | 9.46 | 0.00 | 0.00 | 0.00 | 0.00 |
| Age | | | | | | | | | | |
| 65-69 | 35.45 | 100.00 | 38.27 | 0.00 | 0.00 | 11.68 | 54.54 | 0.00 | 24.16 | 0.00 |
| 70-74 | 27.38 | 0.00 | 0.00 | 60.20 | 65.29 | 52.75 | 0.00 | 59.49 | 47.72 | 0.00 |
| 75-79 | 18.69 | 0.00 | 12.00 | 39.80 | 34.71 | 34.37 | 0.00 | 40.51 | 28.12 | 5.87 |
| 80-84 | 18.47 | 0.00 | 49.73 | 0.00 | 0.00 | 1.20 | 45.46 | 0.00 | 0.00 | 94.13 |
| Valid | 27.29 | 0.00 | 0.00 | 0.00 | 100.00 | 99.13 | 0.00 | 55.67 | 100.00 | 0.00 |
| Proxy | 21.14 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 100.00 | 100.00 | 0.00 | 0.00 |
| Covered by Medicare | 12.94 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 100.00 | 0.00 | 0.00 |
| Reason for Leaving Prev. Wave Add | | | | | | | | | | |
| Deceased | 29.43 | 91.60 | 87.32 | 103.95 | 88.57 | 101.70 | 89.05 | 103.25 | 100.19 | 94.25 |
| Institution | 7.57 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 51.63 |
| Moved | 0.48 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 5.90 | 0.00 | 0.00 | 0.00 |
| Separate | 0.48 | 0.00 | 0.00 | 0.00 | 9.72 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 201+ | 7.09 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 48.37 |
| Other | 54.94 | 0.00 | 100.00 | 0.00 | 90.28 | 0.00 | 94.10 | 100.00 | 100.00 | 0.00 |
| Number of Persons In Family | | | | | | | | | | |
| 1-5 | 98.49 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 38.86 | 100.00 | 100.00 | 100.00 |
| 6-10 | 1.44 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 58.09 | 0.00 | 0.00 | 0.00 |
| 11-15 | 0.08 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 3.06 | 0.00 | 0.00 | 0.00 |
| Kind of Family - Headship | | | | | | | | | | |
| Husband/Wife | 54.89 | 100.00 | 0.00 | 0.00 | 0.00 | 0.00 | 100.00 | 0.00 | 0.00 | 0.00 |
| Male | 10.04 | 0.00 | 0.01 | 100.00 | 100.00 | 100.00 | 0.00 | 100.00 | 0.00 | 0.01 |
| Female | 35.06 | 0.00 | 99.99 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 99.99 |
| # of Own Children in Family | | | | | | | | | | |
| 1 | 6.85 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 80.07 | 0.00 | 0.00 | 0.00 |
| 2 | 1.71 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 19.93 | 0.00 | 0.00 | 0.00 |
| Type of Housing Unit | | | | | | | | | | |
| House | 93.83 | 99.10 | 66.26 | 99.94 | 100.00 | 99.91 | 97.96 | 100.00 | 100.00 | 0.00 |
| Nontrans | 0.26 | 0.00 | 13.34 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Perman | 0.12 | 0.00 | 1.47 | 0.00 | 0.00 | 0.09 | 2.04 | 0.00 | 0.00 | 0.61 |
| Rooming House | 0.14 | 0.00 | 7.36 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Trailer 1 | 4.21 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 77.69 |
| Trailer 2 | 0.98 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 18.14 |
| Other 1 | 0.08 | 0.90 | 2.47 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Quarters | 0.15 | 0.00 | 7.61 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Unit | 0.01 | 0.00 | 0.62 | 0.06 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Unoccupied | 0.02 | 0.00 | 0.87 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Other 2 | 0.19 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 3.56 |
| Access to Housing Unit | | | | | | | | | | |
| Direct | 99.64 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 86.57 | 100.00 | 100.00 | 100.00 |
| Indirect | 0.36 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 13.43 | 0.00 | 0.00 | 0.00 |

Table 1 (cont'd)

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Kitchen Facilities | 0.10 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 4.64 | 0.00 | 0.00 | 0.00 |
| Residence in Public Housing Project | 2.43 | 0.00 | 16.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| # of Units in Structure | | | | | | | | | | |
| Other | 0.54 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 14.47 | 0.00 | 0.00 | 0.00 |
| Trailer | 5.12 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 18.11 |
| One Det | 67.46 | 100.00 | 0.00 | 0.00 | 0.00 | 0.00 | 85.53 | 0.00 | 0.00 | 81.89 |
| One Att | 4.55 | 0.00 | 0.00 | 0.00 | 45.41 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Two | 5.47 | 0.00 | 0.00 | 0.00 | 54.59 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 3-4 | 3.89 | 0.00 | 23.08 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 5-9 | 3.06 | 0.00 | 18.14 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 10-19 | 2.62 | 0.00 | 15.53 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 20-49 | 2.11 | 0.00 | 12.49 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 50+ | 5.18 | 0.00 | 30.75 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Tenure of Living Quarters | | | | | | | | | | |
| Owned | 75.86 | 97.25 | 0.00 | 0.00 | 0.00 | 0.00 | 100.00 | 0.00 | 0.00 | 94.48 |
| Rented | 20.92 | 0.00 | 100.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| No Cash | 3.22 | 2.75 | 0.00 | 100.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 5.52 |
| Public Housing Unit Where Rent is Collected | 0.54 | 0.00 | 100.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Family Type to Which Person Belongs | | | | | | | | | | |
| Primary I | 65.51 | 100.00 | 0.00 | 0.00 | 0.00 | 0.00 | 93.18 | 0.00 | 0.00 | 0.00 |
| Second | 1.32 | 0.00 | 1.92 | 0.00 | 100.00 | 0.00 | 0.00 | 0.00 | 0.00 | 2.97 |
| Unrelate | 0.07 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.58 | 0.00 | 0.00 | 0.00 |
| Relate | 0.72 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 6.23 | 0.00 | 0.00 | 0.00 |
| Primary 2 | 32.37 | 0.00 | 98.08 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 97.03 |
| Covered by Medicare 1 | 98.06 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 0.00 | 100.00 | 100.00 | 100.00 |
| Covered by Medicaid 1 | 6.43 | 0.00 | 0.00 | 0.00 | 0.00 | 98.87 | 0.00 | 0.00 | 0.00 | 0.00 |
| Covered by Health Insurance 1 | 52.04 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 93.43 | 100.00 | 0.00 |
| Covered by Medicare 2 | 98.06 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 0.00 | 100.00 | 100.00 | 100.00 |
| Covered by Medicaid 2 | 6.44 | 0.00 | 0.00 | 0.00 | 0.00 | 99.09 | 0.00 | 0.00 | 0.00 | 0.00 |
| Covered by Health Insurance 2 | 51.96 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 94.23 | 100.00 | 0.00 |
| Covered by Medicare 3 | 98.06 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 0.00 | 100.00 | 100.00 | 100.00 |
| Covered by Medicaid 3 | 6.45 | 0.00 | 0.00 | 0.00 | 0.00 | 99.27 | 0.00 | 0.00 | 0.00 | 0.00 |
| Covered by Health Insurance 3 | 51.88 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 94.80 | 100.00 | 0.00 |
| Covered by Medicare 4 | 98.05 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 0.00 | 100.00 | 100.00 | 100.00 |
| Covered by Medicaid 4 | 6.45 | 0.00 | 0.00 | 0.00 | 0.00 | 99.51 | 0.00 | 0.00 | 0.00 | 0.00 |
| Covered by Health Insurance 4 | 51.90 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 96.53 | 100.00 | 0.00 |
| Covered by Medicaid | 9.09 | 0.00 | 0.00 | 0.00 | 0.00 | 100.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Covered by Health Insurance | 72.74 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 100.00 | 100.00 | 0.00 |
| Marital Status | | | | | | | | | | |
| Married 1 | 51.61 | 100.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Married 2 | 0.76 | 0.00 | 0.00 | 0.00 | 8.65 | 11.17 | 1.01 | 0.00 | 0.00 | 0.99 |
| Widowed | 36.39 | 0.00 | 76.82 | 0.00 | 0.00 | 0.00 | 84.17 | 0.00 | 0.00 | 86.69 |
| Divorced | 4.39 | 0.00 | 10.53 | 0.00 | 39.99 | 36.61 | 0.00 | 0.00 | 0.00 | 3.81 |
| Separate | 0.99 | 0.00 | 0.00 | 0.00 | 18.79 | 25.51 | 0.00 | 0.00 | 0.00 | 0.00 |
| Never | 5.85 | 0.00 | 12.65 | 0.00 | 32.57 | 26.71 | 14.82 | 0.00 | 0.00 | 8.50 |
| If Proxy Interview - Is Spouse Proxy Person | 9.38 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 100.00 | 0.00 | 0.00 |