Session 8

LONGITUDINAL SURVEYS

PANEL DESIGN AND ESTIMATION STRATEGIES IN THE NATIONAL MEDICAL EXPENDITURE SURVEY

Steven B. Cohen

Agency for Health Care Policy and Research
2101 E. Jefferson Street
Rockville, Maryland 20852

Presented at the COPAFS Seminar on New Directions in Statistical Methodology
May 25-26, Bethesda Holiday Inn

The views expressed in this paper are those of the author and no official endorsement by the Department of Health and Human Services or the Agency for Health Care Policy and Research is intended or should be inferred. The author wishes to acknowledge the contributions made by the following individuals who were co-authors of papers whose findings have been synthesized in this presentation: D.E.B, Potter, Barbara L. Carlson, and Ayah Johnson.

PANEL DESIGN AND ESTIMATION STRATEGIES IN THE NATIONAL MEDICAL EXPENDITURE SURVEY

Steven B. Cohen, Agency for Health Care Policy and Research

1. Introduction

The National Medical Expenditure Survey (NMES-2) was established to provide an assessment of the health care utilization, expenditures, sources of payment and health insurance coverage of the U.S. civilian noninstitutional population and the population using nursing and personal care homes and facilities for the mentally retarded for calendar year 1987. The core of the data collection effort for the non-institutionalized population was a series of interviews with a household sample that collected detailed information on health status, use of health care services, expenditures and sources of payments, insurance coverage, employment, income and assets, and demographic characteristics for calendar year 1987. The institutional population component was designed to obtain similar types of information for the institutionalized population residing in nursing and personal care homes (NH), and in facilities for the mentally retarded (MR) for calendar year 1987.

The NMES household survey was designed as a panel survey with an initial screening interview conducted in the fall of 1986 for a sample of approximately 35,600 addresses, to obtain information required for oversampling of specific policy relevant population subgroups. A subsample of about 15,000 households was selected for the detailed interviews. To reduce the deleterious impact of long recall periods on measurement error, data collection specifications required four separate interviews conducted with selected households at three to four month

intervals over a fifteen month period to obtain the required heath care data for calendar year 1987. The adoption of a panel design was also motivated by the analytical goal of measuring changes in health insurance coverage over the course of calendar year 1987.

The NMES Institutional Population Component was designed to yield unbiased national and regional estimates at the facility level and for the overall institutional user population. More specifically, the primary objective of the survey was to estimate the use of and expenses for health care services for all persons residing in institutions at any time during calendar year 1987. To obtain a nationally representative sample of the 1987 institutional user population, the survey included a sample of residents residing in selected facilities as of January 1, 1987, in addition to a representative sample of admissions to the selected facilities over the course of 1987. The union of these samples served to represent the 1987 institutional user population. The operational implications of a selection of both residents as of 1/1/87 and admissions over the course of 1987 required the adoption of a panel design for the survey. Interviewers made four distinct visits to each cooperating facility at approximately four month intervals to facilitate sample selection and data collection in the institutions.

This paper provides a summary of the sample design and estimation strategies adopted in the National Medical Expenditure Survey associated with the longitudinal features of the survey. For the Household Survey, particular attention is given to the consequences of adopting an address sample design in meeting specified survey design goals. The NMES Household survey also benefitted from a design strategy that included the selection of a sample of

households in the first round of data collection for the main study that were non-respondents to the screener interview. The impact of this design strategy on the overall survey response rate is also summarized. The paper includes a discussion of the estimation strategy adopted to adjusted for part-year nonresponse. With respect to the Institutional Population Component, both the sample design and the estimation strategy used to correct for the representation of individuals with multiple opportunities for sample selection over the course of 1987 are described in detail.

2. Address Sample Design in the NMES Household Survey

The Household Component of the National Medical Expenditure Survey (NMES) is a panel survey with 1987 as the reference period, collecting measures of health status, use of health care services, expenditures and sources of payment, insurance coverage, employment, income and assets, as well as demographic information for the U. S. civilian noninstitutional population. To meet analytical objectives, the survey included an oversample of the following policy relevant population subgroups: blacks, Hispanics, the poor and near poor, the elderly and persons with functional limitations. A separate screening interview was conducted in the fall of 1986 to facilitate the identification of these population subgroups. The screened households were selected for the main NMES household survey on the basis of the characteristics of the persons they included at the time of the screening interview. However, for the purposes of cost-efficiency and to maximize the response rate, the actual NMES Round One sample was characterized by an address sample design. An address sample design requires the interviewer to go back to the originally sampled address, whether the same household resides there or not. Consequently, the Round One sample consisted of all households living at the sampled addresses

at the time of the Round One interview, whether or not the screened households were still present.

The sample design considerations inherent in the selection of an address sample as an alternative to a household sample design are clearly not limited to the NMES, but generally applicable to all national household surveys that consider non-concurrent screening interviews to identify particular population subgroups targeted for oversampling. When the basic sampling units are defined to be the sample addresses themselves, rather than the households residing at the addresses at the time of the screening interview, a reduction in survey costs should be realized for the subsequent interview since movers would not have to be traced and interviewed at their new location. However, when the address sample design is used, each sample listing (or a representative subsample) should be returned to in the round of data collection following the screening interview, to interview all eligible respondents residing at the address (Cohen and Johnson, 1992). Addresses that were vacant at the time of the screening interview would have to be checked to determine their occupancy status. Dwelling units whose occupants refused to participate in the screener interview would need to be recontacted in addition to those units in which no eligible respondents were at home or available for being interviewed for the screening interview (Cox and Cohen, 1985; Cox et al., 1979).

The adoption of a household sample design would define households residing within the sampled addresses at the time of the screening interview as the basic sampling units. Individuals or families who moved out of the sample dwelling units would have to be traced and followed for interview. This design imposes greater control over the sample to insure that sample size targets are satisfied for the oversampled population subgroups. Furthermore, this design would

generally yield more precise survey estimates for the oversampled population subgroups than an address sample design, which is more vulnerable to the adverse effects of greater sampling weight variation. However, the advantages of the household sample design are less likely to be realized as the time period between the screening and follow-up interviews increases. Additional interviewers would have to be hired to conduct the Round One interview for the selected screener respondents that moved outside the PSUs that comprised the NMES screener sample. Furthermore, the field period would have to be extended to accommodate the time needed to locate the movers and conduct their interview.

The choice of design requires an evaluation of the potential loss in the precision of surveys estimates for the oversampled population subgroups, further contrasted with the potential savings in survey costs and higher response rate achieved by the address sample design. In this study, the consequences of adopting an address sample design for the National Medical Expenditure Survey are evaluated in terms of the realization of specified survey design goals.

The adopted NMES household survey sample design is a stratified area probability design with the following stages of selection: (1) selection of 165 primary sampling units (PSU's) which are counties, parts of counties or groups of contiguous counties; (2) selection of 2,317 segments within PSU's; (3) selection and screening of dwelling units within segments; and (4) selection of dwelling units based on demographic characteristics from the set of screened dwelling units (Cohen, DiGaetano and Waksberg,1991). The survey was sponsored by the Agency for Health Care Policy and Research and the NMES sample represents a union of the national sample frames of Westat, Inc., the prime contractor, and NORC, a subcontractor.

The NMES screener interview was conducted in the fall of 1986. The final NMES screener sample consisted of 35,634 addresses, of which 3,091 were identified as vacant, another 1,085 identified as not a dwelling unit and 250 addresses were determined to be ineligible for the survey. Of the 31,208 remaining dwelling units sampled, 28,458 responded to the screener interview, achieving a 91.2 percent response rate.

The Round One household sample of dwelling units was then selected from the screened households, using sampling rates that achieved the desired sample for specified population subgroups (DiGaetano,1987). The 15,130 dwelling units that constituted the targeted Round One sample, based on the demographic and health status profiles of their members at the time of the NMES screener interview, initially consisted of 16,615 responding reporting units. Reporting units were defined as individuals related by blood, marriage or adoption within a dwelling unit. These reporting units contained 39,885 individuals that represented the civilian non-institutionalized population as of Fall, 1986.

Figure 1 illustrates the transitions in sample composition that would occur between the screener and Round One interviews, as a consequence of the address sample design. The expected transitions include the movement of targeted households out of sampled addresses, changes in household composition, and the inclusion of replacement households that have moved into sampled addresses.

FIGURE 1 ABOUT HERE

It was also expected that some of the addresses contacted which were vacant in the screener field period would become occupied at the time of the Round One interview. Excluding

them from the sample would understate the number of recent moves in the sample. Consequently, a sample of 1,464 vacant addresses was selected to supplement the occupied addresses sampled from all screened households.

o Field Results

Field results indicated that the majority of the 15,130 sampled addresses consisted of the same reporting units that completed the screener interview. Transitions occurred as a consequence of the following situations:

1. movement by screener identified families out of sampled dwelling units,
2. creation of new reporting units, consisting of individuals related by blood, marriage or adoption, that have moved into sampled addresses since the time of the screener interview, and
3. changes in the Round One composition of reporting units that were initially identified at the time of the screener interview.

At the end of the Round One field period, it was determined that 722 (4.5 percent) of the 15,130 sampled addresses responding to the NMES screener interview were vacant. Another 253 sampled addresses were determined to be ineligible for the interview as a consequence of the death or institutionalization of targeted respondents, or due to a change in the original household composition to all military or student members. Furthermore, 847 reporting units that responded to the screener moved out of the sampled addresses, and were replaced by new reporting units at the time of the Round One interview. Overall, more than ten percent of the sampled addresses with screener respondents experienced at least one household move during the five to six month period that passed between interviews..

Within the 15,130 sampled addresses that constituted the targeted Round One sample, there were 15,590 reporting units identified as eligible or NMES interviews during the Round One field period out of 17,412 reporting units linked to the addresses during the screener or Round One interview. Of these, 14,060 responded (90.2 percent) to the Round One interview. The remaining 1,530 eligible RUs were classified as nonrespondents (9.8 percent) due to a refusal to complete the interview, unavailability during the field period, illness, or other nonresponse. Relative to the 1,464 sample addresses vacant at the time of the screener, 1,016 (69.4 percent) remained vacant at the time of the Round One interview. Of the 479 reporting units new to the sample, 408 (85.2 percent) responded to the first household interview. The joint screener-round one response rate for these targeted Round One addresses, including the screener vacant sample, was 82.1 percent. This was derived by multiplying the screener response rate (.912) with the combined Round One response rate for targeted sample and the vacant sample ((14,060 + 408)/(15,590 + 479) = .90).

At the person level, it was noted that 32,205 (80.7 percent) of the targeted 39,885 screener respondents also responded to the Round One interview. Of the 7,680 screener only-respondents, 3,150 individuals (41 percent) were in targeted reporting units that refused to complete the Round One interview. Another 224 individuals (2.9 percent) resided in reporting units that were determined to be ineligible at the time of the Round One interview. Consequently, 43.9 percent of the screener only- respondents would not have participated in the Round One interview, independent of the address sample design.

It was also determined that 1,673 of the screener only- respondents (21.8 percent) departed from the sample as a consequence of the movement of their reporting units (847 RUs)

away from the sampled address, which gained a new replacement household available for the Round One interview. Another 1,529 individuals (19.9 percent) left the sample due to the movement of their reporting units out of the sampled address, which was vacant at the time of the Round One interview. The remaining 1,104 screener-only respondents that departed from their sampled addresses (14.4 percent), were associated with reporting units for which at least one respondent completed both the screener and Round One interview at the sampled address. Consequently, 56.1 percent of the screener only- respondents (4,306) were not sampled in Round One as a function of the address sample design. The final set of respondents to the first round of the NMES household survey consisted of 36,259 individuals. Of the 4,054 individuals that were Round One only-respondents, 841 (20.7 percent) were associated with initially sampled reporting units that did not respond to the NMES screener interview, but were refielded to improved the overall NMES response rate. Independent of an address sample design, a decision to refield these cases could have been incorporated in the NMES sample design. The remaining 3,213 Round One only-respondents (79.3 percent) were added to the NMES sample as a function of the address sample design.

o A COMPARISON OF THE DEMOGRAPHIC CHARACTERISTICS OF NMES MOVERS

Of the 7,680 screener only-respondents, 43.9 percent would not have completed a Round One interview, independent of the address sample design for NMES. Under an alternative sample design that tracked all screener respondents that moved, efforts would have been made to locate or determine the status of 4,306 individuals linked to more than 1,500 reporting units.

The demographic characteristics of these targeted sample movers are presented in Table 1, and contrasted with the characteristics of the 3,213 replacement individuals that are new to the sample as a function of the address sample design (excluding the 841 new Round One respondents associated with RUs that were screener nonrespondents).

TABLE 1 ABOUT HERE

According to Current Population Reports, (U.S. Bureau of the Census, 1987) 18 percent of the population experienced a move during the course of the year. The NMES national estimate of population transition (10 percent) during the five to six month period that transpired between the screener and Round One interviews, compares well with census figures.

Overall, the 3,213 individuals that were new to the sample in Round One represented a 25 percent shortfall in targeted sample size, relative to the movers. The population subgroup that consistently experienced the greatest proportionate differential from sample targets were the elderly. In addition, a greater shortfall in targeted sample size characterized the black and Hispanic population subgroups, relative to the domain that represented whites and other races. It was recognized, however, that even if the targeted sample of movers were to be traced, the expected Round One yield would need to reflect nonresponse and loss in sample due to death, institutionalization, or inability to locate.

Under the assumption of a ten percent loss in sample due to nonresponse (which was the experience for non-movers), and a conservative assumption of an additional five percent loss due to death, institutionalization, or inability to locate, the expected Round One yield for the targeted NMES movers was 85 percent. Relative to the expected number of movers completing the Round One interview, the effective overall shortfall in sample size was only 12 percent. When

focused on specific population subgroups, a more pronounced shortfall is sample size was noted, particularly for the elderly. The address sample design was more effective for subpopulation subgroups defined by race or ethnicity, where black or Hispanic households were likely to be replaced by other black or Hispanic households.

The additional measures of poverty status and functional impairment considered in the sampling scheme were not examined, as a consequence of transitions in classification over time that characterize these measures. Screening for these measures is further complicated by the considerable degree of movement into and out of poverty in any two years (Moeller and Mathiowetz, 1990) and the potential movement of elderly individuals from a state of good health to that of disability over time.

Although addresses that experienced a sample movement out were characterized by a shortfall in expected sample size by virtue of the address sample design, a more important consideration was the effect on sample yields for the overall Round One sample. A comparison of the targeted overall Round One sample (36,511) and the resultant Round One sample (35,418) revealed a three percent shortfall. Relative to the targeted Round One sample, a greater shortfall in sample size characterized the black and Hispanic population subgroups (four percent), when contrasted with the white and other races population subgroup (2 percent). Clearly, the overall effect of the address sample design on the Round One sample yields was minimal.

o EFFECT OF ADDRESS SAMPLE DESIGN ON PRECISION

The precision of the NMES survey estimates for the population subgroups of analytical

interest was expected to decrease only marginally, as a function of the small loss in sample size attributable to the address sample design. Even for the population subgroup characterized by the largest relative sample loss (4.3 percent for Hispanics), the expected increment in the standard errors of their associated survey estimates was only two percent. The major cause for concern, however, was the greater variation in sampling weights for the population subgroups of analytical interest (Kish. 1965). Table 2 presents a summary of the variation in the sampling weights that characterized the targeted and actual Round One samples. More specifically, the distributions of the unadjusted sampling weights, and their coefficient of variation are compared across the alternative sample designs for Round One.

TABLE 2 ABOUT HERE

The actual Round One sample was characterized by a higher level of variation in sampling weights across all population subgroups targeted for oversampling. This was most obvious for the black and Hispanic population subgroups. The greater variation in their sampling weights was primarily a function of the movement of new individuals, with characteristics targeted for oversampling, into sampling addresses selected in NMES at a lower sampling rate based on their composition at the time of the screener interview.

The impact of the address sample design on the precision of survey estimates was directly examined by a comparison of the standard errors for a representative set of demographic and health status measures that were available from the screener interview. It was necessary to restrict this comparison to these measures, since the screener only-respondents did not have any

of the information that was collected only in the Round One interview.

The person level demographic measures under investigation included region, size of city, marital status and veteran status. In addition, the health status measures under investigation included all questions in the NMES screener data base that identified functionally impaired individuals for the purposes of oversampling. For each of the survey measures under consideration, the ratio of the standard errors derived from the address sample design and from the potential Round One sample was computed. Study findings indicate only a slight increase in the standard errors of the survey estimates that characterize the address sample design (ratios generally ranged from 1.01 to 1.04). Even for the population subgroup characterized by the greatest loss in precision (white/other race, 65+), the average increment in the standard errors of their associated survey estimates was only six percent.

The cost savings achieved by the address sample design were estimated to be $175,000, and were attributable to not having to locate the targeted individuals that had a change of address since responding to the NMES screener interview. The availability of these additional funds allowed for the implementation of more intensive survey nonresponse conversion techniques to enhance the overall NMES household survey response rate. As a consequence of the address sample design, the response rate that characterized the NMES Round One interview was not affected by a component of nonresponse due to inability to locate.

Under a NMES design that attempted to include the movers in the sample, the field period for the Round One interview would have had to be extended. This would be required in order to accommodate the time needed to locate the movers and conduct the interview. This

extension in the field period would have a potentially adverse effect on the quality of data obtained from the movers, due to the extended length of the recall period that characterized their interview. Use of an address sample design for NMES eliminated the component of response error that would be attributable to this particular extension in recall period. Alternatively, the problems inherent with tracing movers in a NMES survey that considered a household design could be noticeably reduced by a field redesign that facilitated a significant reduction in the average lag period between the screener and Round One interviews.

The NMES address sample design was cost effective and achieved an acceptable response rate relative to a design that would have tracked movers. In addition, the quality of the Round One data was expected to be improved, as a consequence of the shorter length of recall period (where respondents were required to provide health care information at the time of the Round One interview relative to 1/1/87) that characterized the respondents of the address sample design. The major potential limitation of the adopted design was with respect to the precision of survey estimates. Even on this sensitive dimension, study findings indicated only a slight loss in the precision of survey estimates derived from the address sample design. Consequently, the overall benefits of the address sample design were realized in the National Medical Expenditure Survey, with only minimal effects on the precision of survey estimates.

3. Refielding A Sample of Screener Refusals and Other Nonresponding Dwelling Units in the First Round of the NMES Household Survey.

A supplemental sample of refusals to the screener and other nonresponding dwelling units was

also included in the round 1 sample, to improve the overall NMES response rate over the four rounds of data collection. In addition to 1,600 dwelling units classified as screener interview refusals, another 916 dwelling units were categorized as "other nonresponse". The other nonresponse classification included dwelling units where no one was home after four calls, where the potential household respondents were unavailable during the screener field period, and where respondents were considered too ill to complete the interview.

Refusals and other nonresponding dwelling units were classified into the same four strata used for sampling vacant addresses. A sample of 645 addresses with dwelling units categorized as refusals to the screener was then selected from the 1,600 addresses with such eligible units and a sample of 376 addresses with dwelling units categorized as other nonresponding units were selected from 916 addresses with eligible units with this classification. Consequently, the only set of screener nonrespondents that were not considered for selection in the first round of the NMES household survey consisted of dwelling units with a language problem, dwelling units where the interviewer was unable to enter the structure, and those classified in the residual nonresponse category. These 234 dwelling units were not targeted for selection in round 1 because of the low probability of converting them to participate in the four rounds of the household survey. By recontacting the nonrespondents to the NMES screener interview, 99.26 percent of the sample of all households were given a chance to participate in the NMES round 1 interview. It should be noted, however, that any new dwelling units built during the five month lag time between the screener and the round 1 interview were not included in the round 1 sample frame. The final round 1 sample is summarized in Table 3.

TABLE 3 ABOUT HERE

Round One Field Results

As noted, the joint screener-round one response rate for the targeted round 1 address sample based on responses to the screener interview, including the screener vacant sample, was 82.1 percent. This round 1 response rate was improved by recontacting nonrespondents to the NMES screener interview. In the set of refielded addresses which were characterized by screener nonresponse, 372 (40.1 percent) of the 928 eligible reporting units responded in round 1. This sampling approach contributed an additional 3.3 percent to the joint response rate. Overall, the joint screener round one response rate for the NMES was 85.4 percent.

4. Part-Year Nonresponse in the NMES Household Survey

Panel surveys are subject to wave nonresponse, which occurs when responses are obtained for some but not all waves of the survey. Prior to selecting an adjustment strategy, it is essential that the patterns of nonresponse and potential correlates are examined. When levels of sample attrition are high, it has been suggested that wave nonresponse is related to level of respondent burden (Czajka, 1986). Generally, the process of sample attrition is non-random (Short and McArthur, 1986). Partial respondents are often distinguished from their complete respondent counterparts on a number of dimensions. These distinctions should be considered in the determination of a nonresponse adjustment strategy.

Two general strategies to correct for sample attrition have gained acceptability in the statistical community: sampling weight nonresponse adjustments and imputation (Kalton, 1986;

Czajka, 1986; Herringa and Lepkowski, 1986). Weight adjustments for sample attrition are rather straightforward to implement and avoid the increase in sampling variance expected in survey estimates as a function of imputation. The primary limiting features of the technique are the loss of large amounts of useful data provided by partial respondents, and the deleterious impact of large nonresponse adjustments to sampling weights on the precision of survey estimates. Imputation for panel attrition allows for the inclusion of partial respondents in the derivation of survey estimates, and the use of the data they have provided. Within the imputation framework, there are several general methodologies that are considered to compensate for panel nonresponse: longitudinal hot deck imputation (Herringa and Lepkowski, 1986). Due to the large number of time dependent analytical measures that are directly affected by sample attrition and the sophisticated software requirements to implement the technique, imputation compensation strategies are the more costly of the two types of strategies.

o Characteristics of the NMES Part-Year Respondents

Of 36,753 key participants in the NMES household survey, 2,294, or 6.2 percent, responded for some, but not all, of the time period in 1987 for which they were eligible. Key sample respondents to the household survey consisted of all civilian non-institutionalized individuals who responded to the Round One interview, in addition to individuals who joined responding Round One reporting units and did not have an opportunity for selection during the period of time that spanned the Round One field period (new babies, military returning to civilian status, individuals in institutions or outside the country returning to their primary

residence). To ascertain the potential level of nonresponse bias that was attributable to partial response in NMES, it was necessary to determine whether the part-year respondents differed systematically from their full year counterparts. To facilitate these comparisons, demographic profiles of these two distinct respondent groups were compared for the following measures: gender, age, race/ethnicity, marital status, years of school completed, census region, size of city, and indicators of functional status.

Since the nonresponse adjustment strategies employed to correct for part-year nonresponse would be dependent on a respondent's data profile for his period of participation in the survey, it was necessary to impose a thresshold on what constituted a minimally acceptable time representation of partial data for making annual national health care estimates. In NMES-2, the minimum part-year response requirement of data for more than one-third of a respondent's period of eligibility followed the approach taken in the 1980 National Medical Care Utilization and Expenditure Survey (sponsored by the National Center for Health Statistics). In NMES-2, 48.6 percent of the part-year respondents (1,114 individuals) who constituted 6.2 percent of the person level sample, did not satisfy this criterion. Consequently, they were treated as total nonrespondents and a standard weighting class adjustment for non-response was applied to the sampling weights of the remaining respondents to correct for their exclusion.

The comparisons of the race/ethnicity distributions for the part-year respondents with data for at least one-third of their period of eligibility in 1987 (henceforth referred to as part-year respondents) and full year respondents revealed a significantly greater representation of whites

and other races (non-black, non-Hispanic) among the individuals that provided complete data when compared to the partial respondents (Cohen, Johnson and Carlson, 1989). This pattern was also observed in the 1977 National Medical Care Expenditure Survey (NMCES), which was the predecessor of NMES (Cohen, 1982). Alternatively, the partial respondents had a higher representation of Hispanics than their complete data counterparts. There was also a significantly higher representation of the partial respondents living in the 19 largest Standard Metropolitan Statistical Areas (SMSAs) in the nation, indicating a greater likelihood of sample attrition in NMES in the large urban metropolitan areas. Furthermore, partial respondents were more likely to reside in the Northeast region of the United States than individuals who provided full year response profiles.

No significant gender differentials were noted across respondent groups. With respect to marital status, the full year respondents were more likely to be married than their partial respondent counterparts. Alternatively, the partial respondents had a greater representation of never married individuals, which mirrored the 1977 NMCES experience. Furthermore, a comparison of the age distributions that characterized the respective respondent groups revealed that partial respondents were more likely to be aged 20-29. Since this age group represents a highly mobile population subgroup, this suggests that the sample attrition that they displayed in the NMES was partially a function of migration. Furthermore, complete respondents had a higher representation of elderly individuals between the ages 70-74 than their partial respondent counterparts. With respect to years of education completed, the partial respondents had a higher representation of individuals with some high school training, as indicated by at least 9-11 years

of education.

To minimize the nonresponse bias in survey estimates attributable to partial response, an appropriate estimation strategy is needed to adjust the data for the remaining sample members who did not respond for their entire period of eligibility. In view of the programming time and cost necessary to implement an imputation strategy to correct for partial nonresponse, and the relatively small representation of partial nonrespondents in NMES, the advantages of this technique for NMES application were not obvious. The technique would require the linkage of partial respondents to complete respondents with matching demographic and health status profiles, the extraction of data from the complete respondent which corresponded to the nonresponding time period of the partial respondent, and its imputation to the partial respondent for each time dependent variable in NMES-2, representing a complex and expensive process.

Traditionally, when the level of partial response is low, it is often preferable to treat partial respondents as complete nonrespondents. Using this approach, only those sample participants providing complete data would be used in the analysis. This was the approach taken for the remaining part-year respondents in the NMES-2 household survey. Weighting classes were formed by cross-classifications of the following measures: race/ethnicity, age and gender. The person level sampling weights for the full year respondents were further post-stratified to poverty status estimates derived from the Current Population Survey.

5. An Estimation Strategy to Represent the Institutional User Population in the NMES-2

Institutional Population Component

The Institutional Population Component (IPC) of the National Medical Expenditure Survey (NMES) was established to provide an assessment of the health care utilization, costs, sources of payment and health insurance coverage of the U.S. institutionalized population residing in nursing and personal care homes (NH), and in facilities for the mentally retarded (MR). The primary objective of the survey was to estimate the use of and expenses for health care services for all persons residing in institutions at any time during calendar year 1987. To obtain a nationally representative sample of the 1987 institutional user population, the survey included a sample of residents residing in selected facilities as of January 1, 1987, in addition to a representative sample of admissions to the selected facilities over the course of 1987. The union of these samples served to represent the 1987 institutional user population.

o Sample Design

The adopted NMES institutional population survey is a stratified, two stage probability design with two phases of facility selection. Current residents (residents on January 1, 1987) and admissions (persons admitted between January 1, and December 31, 1987) were sampled within participating facilities at the second stage.

The IPC facility sample consisted of 851 eligible nursing and personal care homes and 730 eligible facilities for the mentally retarded. Facilities were considered to be respondents to the survey when they completed a Facility Questionnaire. Consequently, the IPC facility level response rate was 95.2 percent for nursing and personal care homes, and 94.7 percent for

facilities for the mentally retarded.

The design of the survey required that the institutional use and expenditure data for current residents were to be collected for their entire period(s) of institutionalization in 1987. In contrast, IPC data collection for the admissions sample began with their first admission to a sampled IPC eligible facility, independent of prior institutional stays over the course of 1987. Consequently, their 1987 institutional data collection period was constrained. For estimation purposes, individuals who responded for at least a third of their eligibility period of institutional data collection were considered respondents.

In the nursing and personal care IPC sample, 805 participating facilities (94.6 percent) allowed for the selection of a sample of their residents as of January 1, 1987. Overall, 3,392 eligible residents were selected, representing a national nursing and personal care home population of 1.5 million residents. Similarly, in the IPC sample of facilities for the mentally retarded, 685 participating facilities (93.8 percent) allowed for "current" resident sampling. Overall, 3,738 eligible residents were selected, representing a national population of 212,000 residents in facilities for the mentally retarded.

Overall, the response rate in the IPC for current residents providing data for at least one-third of their period of institutionalization in 1987 was 89.5 percent for residents in nursing and personal care homes (.946 facility level response rate x .946 resident level response rate), and 88.4 percent for residents in facilities for the mentally retarded (.938 facility level response rate x .942 resident level response rate). This data was to be obtained in the IPC through the administration of the Institutional Use and Expenditure Questionnaire (IUEQ), to be completed

by facility staff (Edwards and Edwards, 1989).

The admissions sample consisted of 2,608 eligible sampled admissions to nursing and personal care homes, and 889 eligible sampled admissions to facilities for the mentally retarded. Sampled admissions were defined to be individuals who were admitted to the sampled IPC facility during 1987 and had no prior admissions to that facility during the survey year.

In the nursing and personal care home sample, 758 participating facilities (89.1 percent) allowed for the sample selection of admissions at all rounds of data collection. Similarly, 657 facilities for the mentally retarded (90 percent) allowed for the sample selection of new admissions at all rounds of data collection. .

Overall, the response rate for new admissions providing data for at least one-third of their period of institutionalization in 1987 was 81.2 percent for those sampled in nursing and personal care homes (.891 facility level response rate x .911 admission response), and 81.3 percent for admissions sampled in facilities for the mentally retarded (.900 facility level response rate x .903 admissions response rate).

Data collected from facility respondents included facility level characteristics, physical and mental health status and functional limitations of sampled persons, and their socio-demographic characteristics and residential history in and immediately before admission to sampled facilities. Information collected on health care services use and expenses included facility services provided, charges and sources of payment, hospitalizations during the institutionalized period and associated conditions, number of physician contacts, and contacts with other medical care providers and therapists.

This data collection effort was referred to as the Survey in Institutions (SII). During each visit,

interviewers obtained or constructed lists of residents from each cooperating facility and proceeded to select the sample (Edwards and Edwards, 1989). The current resident sample was selected from a list of all residents in sample facilities as of January 1, 1987. Similarly, the admission samples were selected on three separate occasions in cooperating facilities from separate lists of all admissions that occurred during the following time periods in 1987: January 1 to April 30, May 1 to August 31, and September 1 to December 31. Sampled persons were followed throughout 1987. For those who left the facilities in which they were selected, facility use and expenditure data were collected up to the time of discharge. If a sample person entered another IPC-eligible facility, the institutional data collection procedures were continued in the new facility.

Since study objectives required data that facility staff could not be expected to provide, the IPC also included a Survey of Next of Kin. This survey consisted of a set of questionnaires administered to community respondents who knew about sampled persons and their lives outside of institutions. Data were obtained on use and expenditures linked to specific residence periods, living arrangements outside of sampled institutions, perceptions of health status and functional limitations, and arrangements for informal care.

o The NMES Institutional User Population

The IPC sample design consisted of two distinct selections of 1987 institutional users: the first selection was designed to provide a representative national sample of residents in IPC

eligible facilities as of January, 1, 1987 (current residents); and the other selection was designed to provide a nationally representative sample of 1987 admissions to IPC eligible facilities. The strict requirement of a single day of sample eligibility for the current resident sample resulted in a single opportunity of selection for each sampled current resident as of 1/1/87. Imposition of a similar restriction for the selection of admissions, requiring the selection of individuals experiencing their first institutional stay in 1987, would have simplified the sample design by allowing each sampled institutional user a single opportunity of selection. Since this information regarding an individual's prior periods of institutionalization was not available at the time of sample selection, and often unavailable from facility records, such a restriction could not be imposed. Resident history information for sampled admissions was often obtained through the IPC Survey of Next of Kin, whereby community respondents who knew about sampled persons would be the primary source for information regarding prior institutional stays.

As a consequence of the sample selection scheme that was employed, an individual who experienced more than one institutional stay over the course of 1987 had multiple chances of selection into the IPC sample. Furthermore, a subset of sampled admissions was determined to have also resided in an IPC eligible facility on 1/1/87, indicating an overlap with the independent sample of January 1 residents. In order to identify the sample of institutional users that had multiple opportunities of selection in the IPC sample, it was necessary to further classify the IPC sample of institutional users according to their institutional experience over the course of 1987.

o Classification of Current Resident Sample

With respect to the population of institutional users that resided in an IPC eligible facility on January 1, 1987 (referred to as current residents), four mutually exclusive and exhaustive classifications are specified in order to characterize their institutional experience over the course of 1987 (Figure 1). More specifically, the first group of current residents consists of institutional users who remained in the same facility over the course of 1987. Residents in this class are referred to as **static full year residents (Group 1)**. The next class of current residents consists of institutional users who remained in the same facility for only part of calendar year 1987, with no subsequent admissions to IPC eligible facilities (i.e., nursing and personal care homes and facilities for the mentally retarded that met the definition for eligibility in the NMES IPC) over the course of 1987. These institutional users are referred to as **single stay part-year residents (Group 2)**. Current residents with this classification could have returned to the community as a member of the civilian non-institutionalized population, been transferred to an out-of scope facility or institution (e.g., acute care hospital, psychiatric institution), or died while in the institutional setting.

FIGURE 2 ABOUT HERE

The remaining current residents experienced at least one subsequent admission to an IPC eligible facility over the course of 1987. They are distinguished in the following manner. The first group consists of current residents who were formally discharged from the facility they resided in as of January 1, 1987 and subsequently were readmitted to the same facility over the course of 1987. These institutional users are referred to as **current residents with re-admissions to same facility (Group 3)**, and consist of residents with one or more re-admissions restricted to the same facility over the course of 1987. Alternatively, the remaining group consists of

current residents with admissions to different facilities (Group 4). This classification also includes institutional users who were re-admitted to the same facility they resided in on 1/1/87 **and** who also experienced at least one admission to a different IPC eligible facility over the course of 1987.

Classification of Individuals Sampled as Admissions

In a complementary manner, four mutually exclusive and exhaustive classifications are specified in order to characterize individuals who experienced at least one admission to an IPC eligible facility over the course of 1987: these individuals are referred to as sampled admissions (Figure 1). The first class of affected institutional users consists of individuals who were not residents in IPC eligible facilities as of 1/1/87 and whose first institutional admission in 1987 was in a sampled IPC facility (Primary Sample Facility (PSF)). Institutional users in this class are referred to as **sampled admissions with initial 1987 admission to an IPC Primary Sample Facility (Group 5)**. This group includes individuals with one or more unique admissions to eligible institutions over the course of 1987 (Groups 5a or 5b, and Groups 5c or 5d, respectively) .

The next classification identifies individuals with a 1987 institutional admission to an IPC Primary Sample Facility, who were also residents in the same facility as of 1/1/87. Institutional users in this group are referred to as **residents sampled as admissions with 1987 admission(s) to the same IPC Primary Sample Facility (Group 6)**. This group of institutional users was already represented in the NMES Institutional Population Component Survey by current residents classified as **residents with re-admissions to same facility (Group 3)**, and by a subset of the

residents with admissions to different facilities (Group 4) who were also re-admitted to the Primary Sample Facility. As a consequence of the ease in identifying these current residents who were re-admitted to the same facility over the course of 1987, they were not considered eligible for IPC data collection. A related group of institutional users consists of individuals with an admission to an IPC Primary Sample Facility who were also residents in an non-sampled IPC eligible facility as of 1/1/87. Such individuals are referred to as **residents in non-sampled facilities with 1987 admission(s) to an IPC Primary Sample Facility (Group 7)**. This class of institutional users was also dually represented in the NMES IPC sample by a subset of the **current residents with admissions to different facilities (Group 4)**. These sampled admissions were not excluded from IPC data collection as a consequence of being unable to determine, at the time of sampling, whether they were institutionalized in some other IPC eligible facility on January 1, 1987.

The remaining group of institutional users with 1987 admissions consists of individuals who did not reside in IPC eligible settings as of 1/1/87, and who were admitted to non-sampled IPC eligible facilities in 1987 prior to an admission to a Primary Sample Facility. Institutional users in this class are referred to as **admissions in IPC Primary Sample Facilities with initial 1987 admission to a non-sampled IPC eligible facility (Group 8)**. This class of institutional users was also dually represented in the NMES IPC sample by a subset of the **new admissions with initial 1987 admission to an IPC Primary Sample Facility (Group 5d)**.

IPC Sample Distribution of Institutional Users

A summary of the IPC sample distribution of institutional users, further distinguished by

facility setting, is presented in Table 4. These tabulations include individuals with response profiles for utilization and expenditure data for at least a third of their period(s) of institutionalization in 1987, beginning with their **sampled stay.** For current residents, this translated to their entire period(s) of institutionalization in 1987. Alternatively, individuals with 1987 admissions were classified as respondents when response profiles were obtained for at least a third of their institutional experience in 1987, beginning with their **sampled admission.** Inclusion of these partial respondents in the derivation of national health care utilization and expenditure estimates for the institutional user population requires implementation of an imputation procedure to adjust for missing time dependent data (Cohen and Potter, 1990). Since IPC data collection for the admission sample began with their first admission to a sampled IPC facility, their 1987 institutional data collection period was constrained. When resident history information was not obtained either through the IPC Survey of Next of Kin or the IPC Survey in Institutions for periods in 1987 prior to their **sampled admission,** resident history profiles were imputed for the missing time gaps in 1987 (Potter and Cunningham, 1990). Inclusion of these sampled admissions with prior periods of institutionalization in the derivation of national health care utilization and expenditure estimates for the institutional user population also requires implementation of **additional** imputation procedures to correct for missing time dependent data (Cohen and Potter, 1990).

The sample of institutional users in nursing and personal care homes consisted of 5,585 respondents, with 3,209 (57.5 percent) sampled as current residents and 2,376 (42.5 percent) sampled as 1987 admissions (Table 4). Relative to the current resident sample, 2,586 (80.6 percent) were classified as static full year residents (Group 1), another 150 (4.7 percent) were

single stay part-year residents (Group 2), with the remaining 473 (14.7 percent) experiencing subsequent admissions to eligible facilities in 1987 (Groups 3 and 4). After excluding the 448 sampled admissions that were considered ineligible for IPC data collection (Group 6), the admission sample was dominated by 2,002 (84.3 percent) institutional users who were not institutionalized on 1/1/87 (Groups 5 and 8).

Alternatively, the sample of institutional users in facilities for the mentally retarded consisted of 4,323 respondents, with 3,520 (81.4 percent) sampled as current residents and 803 (18.6 percent) sampled as 1987 admissions (Table 4). As a consequence of the low representation of **sampled admissions** in these types of facilities in any given year, IPC sample size specifications for the admission sample in facilities for the mentally retarded did not assume separate national estimates would be made for the **sampled admissions**. Relative to the current resident sample, 3,089 (87.8 percent) were classified as static full year residents (Group 1), and another 73 (2.1 percent) were single stay part-year residents (Group 2). The remaining 358 experiencing subsequent admissions to eligible facilities in 1987 (Groups 3 and 4) with the majority (316) experiencing admissions to non-sampled facilities (Group 4). After excluding the 76 sampled admissions that were considered ineligible for IPC data collection (Group 6), the admission sample was primarily represented by 432 (53.8 percent) institutional users who were not institutionalized on 1/1/87 (Groups 5 and 8).

TABLE 4 ABOUT HERE

o Estimation Strategy for the Institutional User Population

The stratified, multi-stage probability sample design adopted for the IPC institutional user population allows for the derivation of approximately unbiased estimates of health care parameters at the national level. This is conditioned upon the application of sampling weights to the sample data that properly reflect the sample selection scheme. The sampling weight for a sample member is defined as the reciprocal of a sample unit's probability of selection (Cox and Cohen, 1985). The estimation strategy for the IPC includes additional adjustments for all levels of nonresponse experienced in the survey. Nonresponse adjustments to the sampling weights have been implemented at the facility level and the institutional user level. To further improve the precision of survey estimates that characterize the IPC sample, post-stratification adjustments on facility level and resident level characteristics have also been implemented, using information from the 1986 Inventory of Long Term Care Places (Flyer, 1992).

One estimation strategy under consideration attempted to maximize the precision in survey estimates that characterize the institutional user population by the inclusion of all responding sampled institutional users in the estimation process. However, the implementation of this estimation strategy is not without penalty. In order to derive national estimates of the health care utilization and expenditure experience for the institutional user population, an imputation strategy must be considered to correct for missing time dependent health care data associated with institutional stays in 1987 prior to the sampled admission. Greater programming resources are required to implement the imputation process that corrects for missing time dependent health care utilization and expenditure data. The approach requires a determination of the exact time period for which institutional data is missing, a linkage between the institutional user with missing time dependent data to the best matching donor with a complete

data profile (using minimum distance function techniques), and imputing the appropriate time dependent data from donor to recipient (Cohen, 1992). The inclusion of sampled units that have multiple opportunities of selection in the derivation of national estimates requires additional adjustments to the survey sampling weights that reflect corrections to sample unit selection probabilities. Furthermore, the inclusion of a multiplicity adjustment to the estimation weights of institutional users selected from the IPC admission sample adds greater variability to their sampling weight distribution, partially limiting the expected gain in precision associated with an increase in sample size.

Given the complexities associated with the implementation of this strategy, and the need to provide timely national health care expenditure estimates of the institutional user population, an alternative approach was adopted for the derivation of use and expenditure estimates and other time dependent measures. More specifically, the alternative estimation strategy restricted the admission sample to institutional users whose first institutional stay in 1987 was in a sampled facility (i.e., Group 5). Adoption of this approach obviates the need for an imputation strategy to correct for missing time dependent data associated with institutional stays in 1987 prior to an institutional user's sampled admission. Furthermore, the restriction of the IPC admission sample to a sample of first institutional stays in 1987 obviates the need for a multiplicity adjustment to estimation weights and an adjustment for dual frame representation of residents in facilities as of 1/1/87. This is a consequence of limiting the sample of institutional users to a single opportunity of selection. Institutional users determined to have experienced institutional stays prior to their sampled admission (Groups 7 and 8) would be defined as ineligible for the

purposes of estimation.

Implementation of this approach results in a sample diminution of 513 out of 5585 respondents for the sample in nursing and personal care homes (9.2 percent reduction) and a comparable loss of 403 out 4,323 respondents in the sample of facilities for the mentally retarded (9.3 percent reduction). The impact of this sample restriction on the precision of survey estimates translates to a 4.9 percent increase in the standard errors that characterize the survey estimates of all institutional users in nursing and personal care homes and a corresponding 5 percent increment for the sample in facilities for the mentally retarded. Greater increments in standard errors are to be noted for the subset of institutional users that experience admissions during 1987. Although the magnitude of this loss in precision is non-negligible, it falls within acceptable levels when contrasted with the time and resource demands inherent in the implementation of an imputation strategy to correct for missing time dependent data associated with institutional stays in 1987 prior to an institutional user's sampled admission. Furthermore, consideration of the restricted first admission sample for the purposes of estimation eliminates exposure to a component of nonresponse bias due to missing time dependent data for prior institutional stays. This component of bias is often only partially reduced through application of imputation techniques. Implementation of imputation strategies to correct for missing time dependent health care utilization and expenditure data associated with prior institutional stays will inform future methodological investigations regarding the impact on survey estimates and their precision due to the inclusion of individuals whose sampled admission was not their first institutional stay in 1987.

6. Summary

The next cycle of the National Medical Expenditure Survey (NMES-3) will be in the field from 1996 through 1997, in order to obtain health care utilization, expenditure and insurance coverage information that characterizes the health care experience of the civilian non-institutionalized population and the population in nursing home for calendar year 1996. A number of design strategies that have been adopted in NMES-2 will also be considered for the NMES-3. One design feature of the NMES-2 Household Survey that will not be adopted for NMES-2 is the address sample design. This decision was not based on any design limitations that were identified in NMES-2 with the adoption of the address sample design, but driven by additional analytical demands placed on the survey. One of the primary motivations for the choice of 1996 as the time period for data collection was the need to have baseline data to assess the impact of health care reform on the nation's health care experience. A number of fast-track states are already in the process of implement health care reform initiatives. Consequently, the households identified for sample selection through the administration of a screening interview will also be subject to an additional interview in the fall of 1995. to gauge their satisfaction with the health care delivery system, their perceptions regarding access, and their current level of health insurance coverage. Since one of the analytical objectives of the NMES-3 household survey will be to assess the longitudinal changes in insurance coverage and access to health health care system over time, including data from the fall of 1995, it will be necessary to follow the same individuals that complete the screening and baseline interviews in the fall of 1995.

The NMES-2 was successful with its strategy to re-field a supplemental sample of

refusals to the screener and other nonresponding dwelling units in the first core round of data collection for the Household Survey. This strategy resulted in a three percent improvement to the overall NMES-2 response rate over the four rounds of data collection. As response rates are monitored for both the NMES-3 household and nursing home surveys, this design strategy will again be given serious consideration.

As a consequence of the panel designs of the NMES-3 household and nursing home surveys, wave nonresponse will remain a concern. As in NMES-2, methodological investigations will be conducted to determine whether the part-year respondents differ systematically from their full year counterparts. Both the level of nonresponse encountered in NMES-3 and the results of the evaluation of the patterns of nonresponse will guide the choice of the nonresponse adjustment strategy that is to be implemented.

Based on the estimation strategy considered in NMES-2 to represent the institutional user population, the NMES-3 nursing home survey design will employ a data collection scheme that limits the likelihood of multiple opportunities of selection into the survey. The planned design will restrict the admission sample for estimation purposes to a sample of individuals experiencing their first institutional stay in a sampled facility for the targeted survey year. Results from the NMES Institutional Population Component Feasibility Study revealed that facility respondents are able to provide accurate information regarding an individual's prior period of institutionalization in a given year. It is recognized that the restriction of the sample of admissions to the individuals experiencing their first institutional stay in a sampled facility for

the targeted survey year cannot be completely implemented at the time of sample selection. However, the ability to use facility information on prior admissions for the selected sample will result in significant cost savings to the survey, based on a reduction in unnecessary data collection activities associated with cases that are not eligible for estimation purposes.

7. References

Cohen, S. B. (1982). "An Analysis of Alternative Imputation Strategies for Individuals with Partial Data in the National Medical Care Expenditure Survey", Review of Public Data Use, 10:153-165.

Cohen, S. B., P. Flyer and D.E.B. Potter (1987). "Sample Design of the Institutional Population Component of the National Medical Expenditure Survey". Presented at the Annual Meeting of the American Public Health Association, New Orleans, LA.

Cohen, S. B., R. DiGaetano and J. Waksberg (1991). Sample Design of the National Medical Expenditure Survey Household Survey, Methods 3. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research.

Cohen, S. B., D.E.B. Potter and P. Flyer (1993). Sample Design of the National Medical Expenditure Survey Institutional Population Component, Methods 6. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research.

Cohen, S.B., A.E. Johnson and B.L. Carlson (1989). "An Analysis of Part-Year Nonresponse in the Household Component of the National Medical Expenditure Survey". Journal of Economic and Social Measurement, 15:281-299.

Cohen, S.B., and D.E.B. Potter (1993). "An Estimation Strategy to Represent the Institutional User Population in the 1987 National Medical Expenditure Survey". Journal of Economic and Social Measurement, 19:159-177.

Cohen, S.B. and A.E. Johnson (1992). "Address Sample Effects the National Medical Expenditure Survey". The American Statistician 46: 100-105.

Cohen, S.B.and D.E.B. Potter (1990). "An Analysis of Part-Year Nonresponse in the Institutional Population Component of the National Medical Expenditure Survey". American Statistical Association, Proceedings of the Section on Survey Research Methods.

Cox, B.G., and S.B. Cohen (1985). Methodological Issues for Health Care Surveys. New York, Basel: Marcel Dekker, Inc.

Cox, B., R. Folsom, B. Moser and T. Virag (1979). Housing Unit Versus Household as the Sampling Unit, Report No.RTI/1725/01-01S, Research Triangle Park, North Carolina.

Czajka, J.L. (1986). "Discussion of Papers on Nonresponse in the Survey of Income and Program Participation". Proceedings of the Section on Survey Research Methods.

DiGaetano, R.G. (1987). Sampling Report for the Selection of Dwelling Units for the Household Survey of the NMES. NMES Report No. 1.00. Westat, Inc., Rockville, Maryland.

DiGaetano, R.G., and J. Waksberg (1986). Sample Design of the Household Survey (HHS) and the Survey of American Indians and Alaska Natives (SAIAN). NMES Report No. 1.46. Westat, Inc., Rockville, Maryland.

Edwards, W. S. and B. Edwards (1989). National Medical Expenditure Survey: Questionnaires and Data Collection Methods for the Institutional Population Component. Methods 1. Department of Health and Human Services, Public Health Service, National Center for Health Services and Health Care Technology Assessment, DHHS Pub. No. (PHS) 89-3440.

Flyer, P. (1992). Sample Design for the Institutional Population Component (IPC) of the National Medical Expenditure Survey. NMES Internal Report. Westat, Inc., Rockville, Maryland.

Herringa, S.G. and J.M. Lepkowski (1986). "Longitudinal Imputation for SIPP". ASA.

Proceedings of the Section on Survey Research Methods.

Kalton, G. "Handling Wave Nonresponse in Panel Surveys". Journal of Official Statistics, 2:303-314.

Kish, L. (1965). Survey Sampling. New York: New York, John Wiley and Sons.

Moeller, J.F. and N. A. Mathiowetz (1990). Problems of Screening for Poverty Status. American Statistical Association: Proceedings of the Section on Survey Research Methods.

Potter, D.E.B. and P.J. Cunningham (1990). "Constructing Year-Long Profiles of Residence Histories for a National Sample of Nursing Home Users." American Statistical Association, Proceedings of the Section on Survey Research Methods.

Sirken, M. G. (1970). "Household Surveys with Multiplicity." Journal of the American Statistical Association, 65,257-266.

Short, K. and E. McArthur (1986). "Life Event and Sample Attrition in the Survey of Income and Program Participation". ASA, Proceedings of the Section on Survey Research Methods.

U.S. Bureau of the Census (1987). Current Population Reports, Series P-25, No. 957.

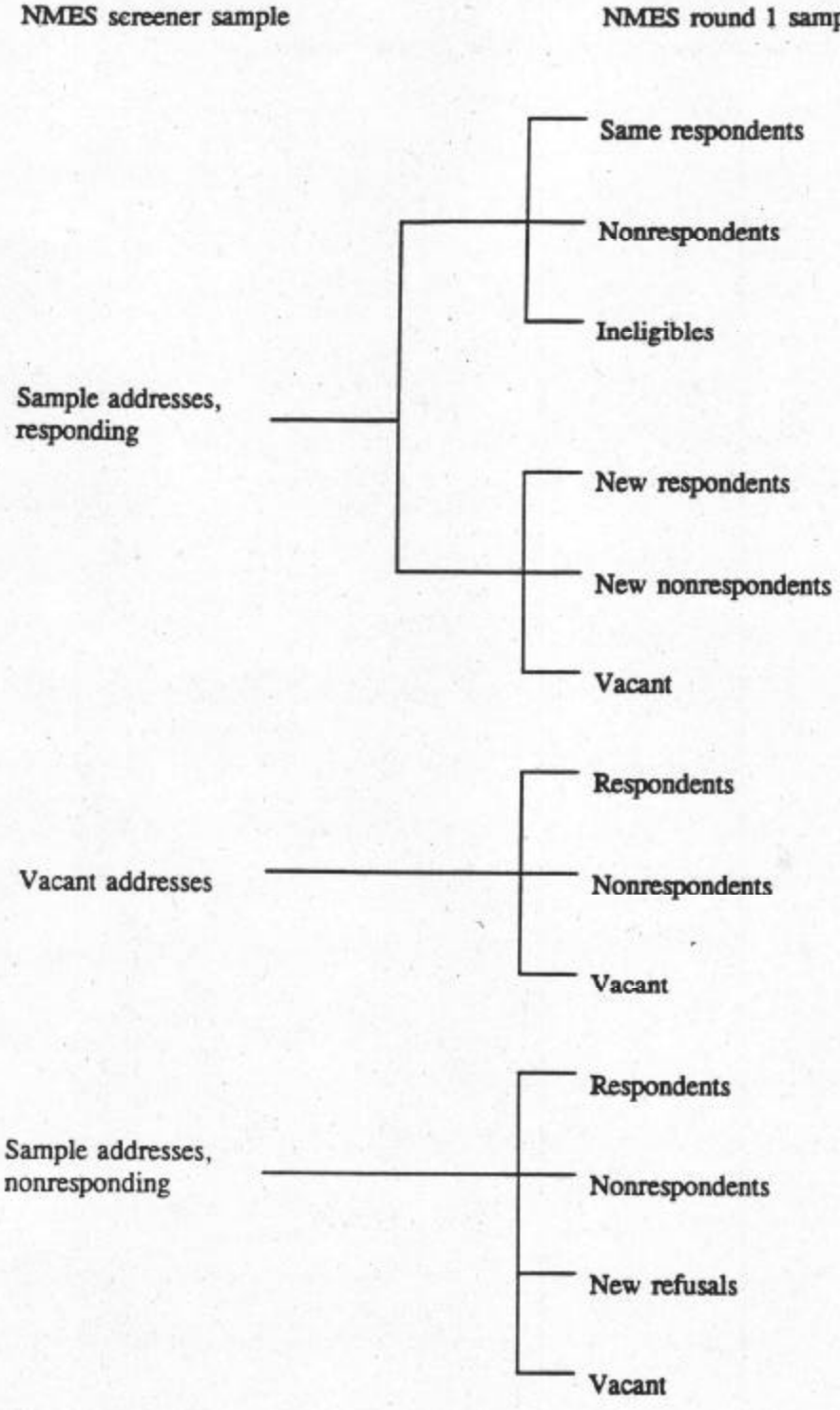


Figure 1. Results of NMES Address Sample Design: Transitions from Screener to Round 1.

Table 1. Comparison of Demographic Characteristics of NMES Movers

Race/ethnicity	Age	Screener-only respondents[a]	Expected[b] Round 1 response	Round 1-only respondents[c]	Round 1-only respondents relative to: Targeted sample (%)	Round 1-only respondents relative to: Expected sample (%)
Hispanic	<65	615	523	447	72.7	85.5
	65+	13	11	6	46.2	54.6
Subtotal		628	534	453	72.1	84.8
Black/non-Hispanic	<65	978	831	686	70.1	82.6
	65+	33	28	20	60.6	71.4
Subtotal	<65	1,011	859	706	69.8	82.2
White/other	65+	2,400	2,040	1,926	80.3	94.4
		267	227	128	47.9	56.4
Subtotal		2,667	2,267	2,054	77.0	90.6
Total		4,306	3,660	3,213	74.6	87.8
National estimates (in thousands)		(24,060)		(21,500)		

[a] Screener-only respondents are those who moved away from sampled addresses prior to the administration of the NMES Round 1 interview.

[b] Under assumption of 85% response rate.

[c] Round 1-only respondents are individuals new to the NMES sample as a function of the address sample design.

Source: Agency for Health Care Policy and Research: National Medical Expenditure Survey, United States, 1987.

Table 2. Variation in Sampling Weights and Mean Standard Error Ratios

Race/ethnicity	Age	Distribution of weights						
		Potential Round 1 sample			Actual Round 1 sample			
		Coefficient of variation	Percentiles		Coefficient of variation	Percentiles		
			1st	99th		1st	95th	99th
Hispanic	<65	.08	2562	3420	.19	2562	3420	7935
	65+	.08	2562	3925	.19	2562	3420	7935
Black/non-Hispanic	<65	.07	2562	2659	.21	2562	2959	5572
	65+	.05	2562	2949	.12	2562	2949	7935
White/other	<65	.31	2659	8235	.25	2863	8235	8235
	65+	.21	2863	8235	.36	2659	5193	8235
Total		.46	2562	8235	.46	2562	8215	8235

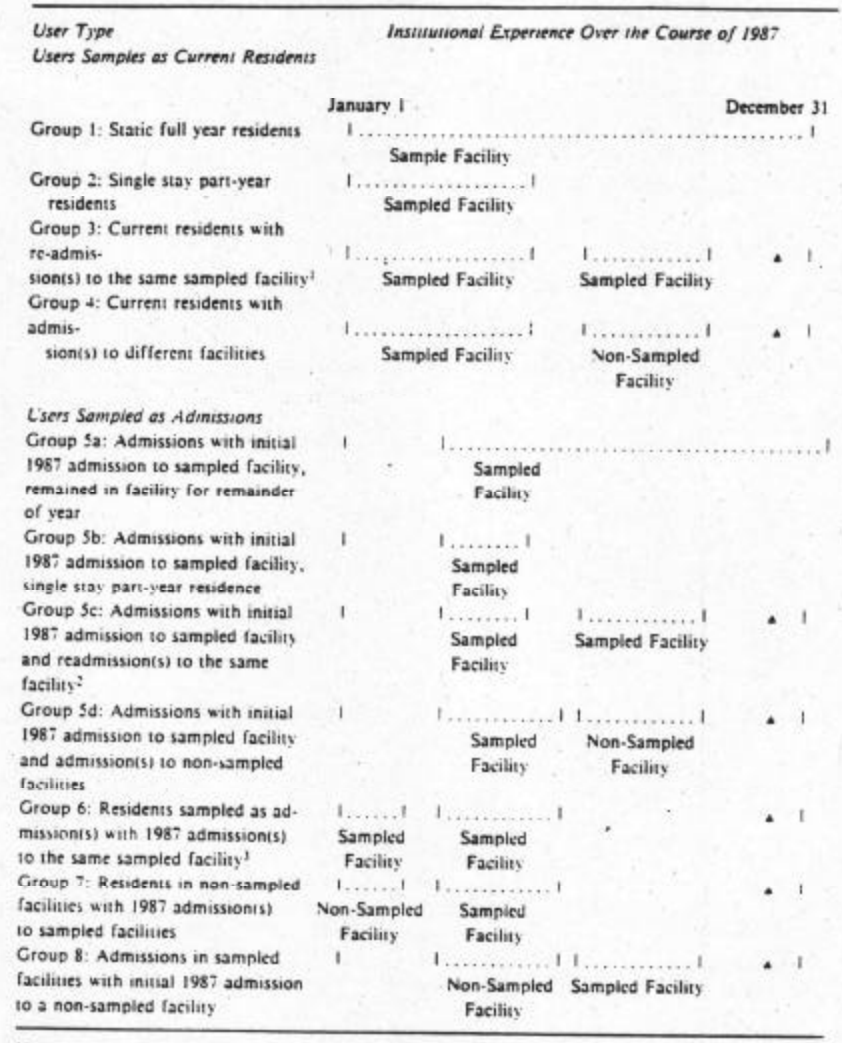
* Mean of standard error ratios for address sample design relative to potential Round 1 sample from household sample design.
Source: Agency for Health Care Policy and Research. National Medical Expenditure Survey, United States, 1987.

Table 3. NMES Household Survey field results at the RU level, round 1

Screener disposition	Number of addresses targeted for round 1	Round 1 responding	Round 1 refusal	Round 1 other non-response	Movers	Other ineligible	Vacant, not a dwelling unit	Total	Overall response rate
Responding	15,130	14,060	1,234	296	847[a]	253[a]	722[a]	17,412	
Response rate	0.912	0.902	0.079	0.019					0.821[b]
Vacant	1,464	408	47	24	0[a]	15[a]	1,016[a]	1,510	
Response rate	--	0.852	0.098	0.050					
Nonresponding	1,021	372	470	86	2[a]	11[a]	111[a]	1,052	
Response rate	0.081	0.401	0.506	0.093					0.033
Nonresponding not fielded	234								
Response rate	0.007	–	–	–	–	–	–	–	0.000
Total		14,840	1,751	406	849	279	1,849	19,974	0.854

[a]Ineligible for round 1.
[b]Combined round 1 response rate for responding and vacant addresses based on screener interview.
Source: Agency for Health Care Policy and Research. National Medical Expenditure Survey.

Figure 2.
Graphic Representation of Institutional User Population in the 1987 National Medical Expenditure Survey



[1]Sampled persons meeting the criterion for User Type 3 and User Type 4 were classified as User Type 4.
[2]Sampled persons meeting the criterion for User Type 5c and User Type 5d were classified as User Type 5d.
[3]Sampled persons meeting criterion for User Type 6 were classified as ineligible.
▲ Potential additional admissions.

Table 4
Distribution of the Institutional User Population, by User and Facility Type, Unweighted (National Medical Expenditure Survey – Institutional Population Component: United States, 1987)

User type	Nursing and personal care homes		Facilities for the mentally retarded	
	Number	Percent	Number	Percent
Users Sampled as Current Residents				
Group 1: Static full year residents	2,586	80.6	3,089	87.8
Group 2: Single stay part-year residents	150	4.7	73	2.1
Group 3: Current residents with re-admission(s) to the same sampled facility	269	8.4	42	1.2
Group 4: Current residents with admission(s) to different facilities	204	6.3	316	8.9
Total current residents	3,209	100.0	3,520	100.0
Users Sampled as Admissions				
Group 5a: Admissions with initial 1987 admission to sampled facility, remained in facility for remainder of year	1,106	46.6	283	35.2
Group 5b: Admissions with initial 1987 admission to sampled facility, single stay part-year residence	432	18.2	68	8.5
Group 5c: Admissions with initial 1987 admission to sampled facility and readmission(s) to the same facility	159	6.7	21	2.6
Group 5d: Admissions with initial 1987 admission to sampled facility and admission(s) to non-sampled facilities	166	6.9	28	3.5
Group 6: Residents sampled as admission(s) with 1987 admission(s) to the same sampled facility	448[a]	0.0[a]	76[a]	0.0[a]
Group 7: Residents in non-sampled facilities with 1987 admission(s) to sampled facilities	374	15.7	371	46.2
Group 8: Admissions in sampled facilities with initial 1987 admission to a non-sampled facility	139	5.9	32	4.0
Total sampled as admissions	2,376[a]	100.0[a]	803[a]	100.0[a]
Overall Total	5,585[a]		4,323[a]	

[a]Users classified into Group 6 were ineligible for data collection and were thus excluded from the totals and percent distributions.
Source: Agency for Health Care Policy and Research.

Council of Professional Associations on Federal Statistics

SEMINAR ON NEW DIRECTIONS IN STATISTICAL METHODOLOGY
Session on Longitudinal Surveys

Methodological Issues Encountered in Following a Cohort of Eighth Graders

Steven J. Ingels, NORC; Jeffrey A. Owings, NCES

Bethesda, Maryland
May 26, 1994

This paper is based on the National Education Longitudinal Study of 1988 (NELS:88). NELS:88 is a longitudinal micro-database sponsored by the National Center for Education Statistics, U.S. Department of Education. The ideas and opinions expressed in this paper are those of the authors and do not necessarily represent positions or conclusions endorsed by NCES.

Introduction. Longitudinal studies typically employ a probability sample of a unit (for example, individuals, institutions [e.g., schools], groups [e.g., families] or inanimate objects [e.g., dwelling units]) that is drawn at one point in time, then repeatedly observed, so that change in units can be measured over time. Longitudinal designs provide a powerful vehicle for reliably[1] measuring individual-level change and development as well as for describing the dynamics of change and the processes that are associated with it. At the same time, longitudinal studies have both inherent and potential limitations (see, for example, Pearson 1989 for a discussion of advantages and disadvantages of longitudinal surveys).

Some of the most important of these limitations can in large measures be overcome if the study is properly executed or if potential limitations are explicitly addressed in the design. This is the case, in particular, for three specific threats to longitudinal sample representativeness.[2]

(1) *Undercoverage*. Potentially biasing undercoverage[3] may arise from any of several sources. It may arise from deliberate or inadvertent exclusion of part of the baseline "target" sample, or may arise if baseline nonrespondents are not pursued in subsequent waves. Undercoverage problems may also arise if the eligibility of ineligible baseline students whose eligibility status is subject to change is not reassessed in succeeding rounds.

(2) *Need for Freshening*. The unit or cohort being studied may become less statistically representative of the target population (or less policy-relevant) over time. For example, a sample of individuals in a given geographical area may become less representative of that area as sample members disperse and other individuals move into the area. A sample of eighth graders two years later is not fully representative of the nation's tenth graders at the second point in time.

(3) *Attrition and Nonresponse*. Sample attrition poses substantial risks for a longitudinal study's representativeness. This danger can be overcome if high response rates are maintained across all rounds, and may be partially compensated for in weighting.

In this paper, the National Education Longitudinal Study of 1988 (NELS:88) serves as an example of how these three specific problems of representativeness in a longitudinal study may be approached and overcome. We more briefly comment on some additional potential sources of survey error.

1. Description of NELS:88

As a point of entry into our topic, we briefly describe NELS:88 by summarizing its goals, surveys administered, response rates, and analysis potential.

1.1 *Goals of NELS:88*

Beginning in 1988 with a cohort of 26,432 eighth graders attending 1,052 public and private schools across the nation, NELS:88 was designed to provide longitudinal data about critical transitions experienced by students as they leave eighth grade school settings, progress through high school (or drop out), enter and leave postsecondary institutions, and enter the work force. The 1988 eighth grade cohort has been followed at two-year intervals (specifically, first follow-up – 1990; second follow-up – 1992) with a third follow-up currently (spring 1994) underway.

Major features of NELS:88 include:

- the integration of student, dropout, parent, teacher, school administrator and school records (transcript) surveys;
- the initial concentration on an eighth grade student cohort with follow-ups at two year intervals;

- the inclusion of supplementary components to support analyses of geographically or demographically distinct subgroups (for example, selected state supplements; oversamples of Asians and Hispanics, and of students in private schools); and

- the design linkages to previous longitudinal studies (High School and Beyond [HS&B], the National Longitudinal Study of the High School Class of 1972 [NLS-1972]) and other current studies (for example, the National Assessment of Educational Progress [NAEP] testing program and high school transcript data collections).

The longitudinal design of NELS:88 permits the examination of change in young people's lives and the role of schools, teachers, community, and family in promoting growth and positive life outcomes. In particular, data from NELS:88 can be used to investigate issues in the context of the family, community, school, and classroom including:

- Students' academic growth over time;

- The transition from eighth grade to high school and the transition from high school to the labor market or postsecondary education;

- The process of dropping out of school, as it occurs from the end of eighth grade on;

- The role of schools in helping the disadvantaged;

- The school experiences and academic performance of language minority students;

- The influence of ability grouping, program type, and coursetaking patterns on future educational persistence and achievement; and

- The features of effective schools.

1.2 Surveys administered

NELS:88 components, by wave, are summarized in Figure 1.

Figure 1: Base Year Through Fourth Follow-Up – NELS:88 Components

BASE YEAR	FIRST FOLLOW-UP	SECOND FOLLOW-UP	THIRD FOLLOW-UP	FOURTH FOLLOW-UP
spring term 1988 GRADE 8 **Students:** Questionnaire, Tests* **Parents:** Questionnaire **Principals:** Questionnaire **Two Teachers per student:** (taken from English, social studies, mathematics, science)	spring term 1990 MODAL GRADE = SOPHOMORE **Dropouts, Students:** Questionnaire, Tests **Principals:** Questionnaire **Two Teachers per student:** (taken from English, social studie mathematics, or science)	spring term 1992 MODAL GRADE = SENIOR **Dropouts, Students:** Questionnaire, Tests, H.S. Transcripts **Parents:** Questionnaire **Principals:** Questionnaire **One Teacher per student:** (taken from mathematics or science)	spring 1994 MODE = H.S. + 2 YEARS **All Individuals:** Questionnaire	spring 1997 HS + 5 YEARS **All Individuals:** Questionnaire

* Reading, social studies, math and science tests are administered in the three in-school rounds.

1.3 *Analysis potential - longitudinal vs. cross-sectional applications.*

Analytic Levels. The NELS:88 design enables researchers to conduct analyses on three principal levels: (1) within-wave (or cross-sectional) analysis at a single time point, (2) cross-cohort analysis (by comparing cross-sectional NELS:88 findings to those of comparable populations studied earlier in HS&B and NLS-72) and (3) cross-wave (or longitudinal[4]) analysis.

The first analytic level within NELS:88 is cross-sectional. By beginning with a cross-section of 1988 eighth graders, following a substantial subsample of these students at two-year intervals, and freshening the 1990 and 1992 samples to obtain representative national cross-sections of tenth and twelfth graders, the study also provides a statistical profile of America's eighth graders, high school sophomores, and high school seniors.

A second analytic level extends representative cross-sections to intercohort comparisons. NELS:88 provides researchers with data for drawing comparisons with previous NCES longitudinal studies. After the release of NELS:88 first follow-up data, researchers were able to conduct trend analyses with the 1980 sophomore cohort of HS&B. With completion of the NELS:88 second follow-up, comparisons may be made among NELS:88, HS&B, and NLS-72 senior cohorts. To facilitate cross-cohort comparisons, some of the questionnaire items used in the NLS-72 and HS&B high school surveys were repeated in NELS:88, and data processing and file conventions were kept consistent, to the maximum possible extent, with HS&B and NLS-72.[5]

The third analytic level is longitudinal, and utilizes repeated measurements on the same individuals over time. However, because NELS:88 comprises three nationally representative grade- and year-defined cross-sections, it supports multiple panels:

1988	eighth graders two, four, six, and nine years later
1990	sophomores two, four, and seven years later
1992	seniors two and five years later[6]

Change Analysis. Cross-sectional analysis provides a snapshot at a single point in time. Repeated cross-sectional analysis, and longitudinal analysis, permit the measurement of change over time. Change (and stability) over time can be measured at the group or individual level:

(1) *At the group level*, change can be measured across the successive cross-sections--eighth graders in 1988, sophomores in 1990, and seniors in 1992. In the same way, multicohort assessments such as NAEP can estimate overall and subgroup gains in specific subject matter proficiency across selected points in the school career (e.g., between fourth, eighth, and twelfth grade). In addition, NELS:88 and comparable studies (e.g., NLS-72 and HS&B) can be analyzed as repeated cross-sections (e.g. of seniors in 1972, 1980/82, and 1992) to measure trends. A cross-sectional time-series such as NAEP also measures trends (e.g. in math achievement for 17 year olds from 1973 to 1990 for the nation and subgroups).

A principal weakness of change measurement at the group level--whether one is looking at rolling (e.g., eighth graders in 1988, sophomores in 1990, seniors in 1992) or repeated (e.g., eighth graders in 1988, 1990, and 1992) static cross-sections[7] is that it sometimes masks individual change; high levels of individual change are not incompatible with stability at the aggregate level. Thus, for example, looking at the proportion of 1988 eighth graders in 1988 who were out of school in 1990 (6.8%) and comparing this to the proportion out of school in 1992 (11.6%) masks the cumulative number of individuals who were

1990 or 1992 spring term dropouts, since some 1990 dropouts had returned to school by 1992.[8] A *locus classicus* of this phenomenon is found in studies of poverty and welfare recipience. While the proportion of adolescent mothers receiving AFDC over time is relatively constant, the AFDC population is not. Mobility onto, and off, the AFDC's rolls is demonstrated by longitudinal data provided by the NLSY, but would not be apparent from repeated cross-sectional results. Likewise, PSID data show that while poverty rates may be roughly stable over time, poverty spells for individuals and households tend to be relatively brief.[9]

(2) Change can also be analyzed *at the individual level* over time. The latter possibility–true longitudinal measurement–represents, for most purposes, the unique strength of the NELS:88 design. Following individual educational histories generally provides the best basis for drawing causal inferences about educational processes and their effects. Two broad kinds of analysis scenarios are possible. Longitudinal analysis can involve repeated measures of the same outcome–for example, test data can be used to measure growth in academic achievement over time. Or longitudinal analysis can show how conditions at an earlier time point are predictive of outcomes at a later time point. For example, one might examine how eighth graders with single or clustered "risk factors" (for example, such status risk factors as coming from a low-income home, having parents who did not finish high school, and so on; or such behavioral risk factors as cutting classes, lack of participation in extracurricular activities, and so on) fared two years later (for example, what proportion had dropped out, repeated a grade, and so on).

While longitudinal studies are prospective, in that they offer the opportunity to record new events, longitudinal analysis may be either retrospective or prospective. In NELS:88, priority in the baseline was given to questions predictive of future behavior. However, while questions that asked for *reasons* for past behavior were deliberately avoided, some retrospective questions were posed, when their focus was on simple descriptions of salient past events. For example, parents were asked whether their eighth grader had attended a Head Start program[10] or kindergarten or preschool, whether other of their children (respondent's elder siblings) had dropped out of school, and so on.

2. Sample Representativeness

This section discusses three key issues. *First*, eligibility and exclusion rules, particularly as applied in the NELS:88 base year, and the measures taken in later rounds of the study to deal with the potential for undercoverage biases that might result from these exclusions. *Second*, the need for sample freshening to ensure representative sophomore and senior cohorts in 1990 and 1992, and the procedures undertaken to bring that sample freshening about. And *third*, attempts to minimize sample attrition and nonresponse error.

2.1 Eligibility: Excluded Students and Undercoverage Bias.

In the base year of NELS:88, students were sampled through a two-stage process. First, stratified random sampling and school contacting resulted in the identification of the school sample; second, students were randomly selected (with oversampling of Hispanics and Asians) from within cooperating schools.

The target population for the base year comprised all public and private schools containing eighth grades in the fifty states and the District of Columbia. Excluded from the NELS:88 school sample are Bureau of Indian Affairs (BIA) schools, special education schools for the handicapped, area vocational schools that do not enroll students directly, and schools for dependents of U.S. personnel overseas.[11] The student population excludes students with severe mental handicaps, students whose command of

the English language was not sufficient for understanding the survey materials (especially the cognitive tests), and students with physical or emotional problems that would make it unduly difficult for them to participate in the survey. This chapter discusses (1) the consequences of student exclusion for the research design and results, and (2) the special measures that have been undertaken in NELS:88 to compensate or correct for the effects of exclusion. Before either of these two topics is pursued in detail, however, it will be desirable to say more about student exclusion in the NELS:88 base year–the 1987-88 school year during which the eighth grade cohort was selected and surveyed.

To better understand how excluding students with mental handicaps, language barriers, and severe physical and emotional problems affects population inferences, data were obtained on the numbers of students excluded as a result of these restrictions.

Seven ineligibility codes defining categories of excluded students were employed at the time of student sample selection:

A - attended sampled school only on a part-time basis, primary enrollment at another school.

B - physical disability precluded student from filling out questionnaires and taking tests.

C - mental disability precluded student from filling out questionnaires and taking tests.

D - dropout: absent or truant for 20 consecutive days, and was not expected to return to school.

E - did not have English as the mother tongue AND had insufficient command of English to complete the NELS:88 questionnaires and tests.

F - transferred out of the school since roster was compiled.

G - was deceased.

Before sampling, school coordinators–members of the school staff, typically an assistant principal or guidance counselor who acted as liaison between the school and the study–were asked to examine the school sampling roster and annotate each excluded student's entry by assigning one of the exclusion codes. Because eligibility decisions were to be made on an individual basis, special education and Limited English Proficiency (LEP) students were not to be excluded categorically. Rather, each student's case was to be reviewed to determine the extent of limitation in relation to the prospect for meaningful survey participation. Each individual student, including LEPs and physically or mentally handicapped students, was to be designated eligible for the survey if school staff deemed the student capable of completing the NELS:88 instruments, and excluded if school staff judged the student to be incapable of doing so. School coordinators were told that when there was doubt, they should consider the student capable of participation in the survey. Exclusion of students after sampling ("post-roster ineligibles") occurred either during the sample update just prior to survey day, or on survey day itself. Such exclusion after sampling normally occurred because of a change in student status (for example, transfer, death). However, in very rare instances such exclusions reflected belated recognition of a student's pre-existing ineligibility–that is, if an annotation error was made and an ineligible student selected for the sample in consequence of such an error, ineligibility became apparent later in the survey, whereupon the student was excluded.

Excluded students were divided into those who were full-time students at the school (categories B, C, and E) and those who were not (categories A, D, F, & G). Our main concern here is with students who were full-time students at the school but who were excluded from the sample. Excluding these students will affect estimates made from the sample.

Students in categories A (n=329), D (n=733), F (n=3,325), and G (n=6) were either not at the school or were present only part time (with primary registration at another school, hence a chance of selection into NELS:88 at another school). Thus excluding students in these categories has no implications for making estimates to the population of eighth grade students.

It should be noted that students in category F, those who had transferred out of the sampled school, had some chance of being selected into the sample if they transferred into another NELS:88 sampled school just as transfers into NELS:88 schools from non-NELS:88 schools had a chance of selection at the time of the sample update. The sampling of transfer-in students associated with the sample update allowed NORC to represent transfer students in the NELS:88 sample.

The total eighth grade enrollment for the NELS:88 sample of schools was 202,996. Of these students, 10,853 were excluded owing to limitations in their language proficiency or to mental or physical disabilities. Thus 5.37 percent of the potential student sample (the students enrolled in the eighth grade in the 1,052 NELS:88 schools from which usable student data were obtained) were excluded. Less than one half of one percent of the potential sample was excluded for reasons of physical or emotional disability (.41 percent), but 3.04 percent was excluded for reasons of mental disability, and 1.90 percent because of limitations in English proficiency.

Put another way, of the 10,853 excluded students, about 57 percent were excluded for mental disability, about 35 percent owing to language problems, and less than 8 percent because of physical or emotional disabilities. Because current characteristics and probable future educational outcomes for these groups may depart from the national norm, the exclusion factor should be taken into consideration in generalizing from the NELS:88 sample to eighth graders in the nation as a whole. This implication for estimation carries to future waves. For example, if the overall propensity to drop out between the eighth and tenth grades is twice as high for excluded students as for non-excluded students, the dropout figures derivable from the NELS:88 first follow-up (1990) study would underestimate early dropouts by about ten percent. (In point of fact, the 1988-90 status dropout rate derivable from the eligible NELS:88 sample representing about 94.6 percent of the cohort is between 6.0 and 6.1 percent, and from the expanded--eligible + ineligible--1988 sample representing [virtually] 100 percent of the cohort, 6.8 percent.)

Undercoverage of course affects the power of a study both to produce national estimates, and, yet more dramatically, to produce estimates for the particular group that is not fully covered.[12] Undercoverage, moreover, poses some special difficulties for the representativeness of a multi-cohort longitudinal study such as NELS:88.

In a school-based longitudinal survey such as NELS:88, baseline excluded students affect the representativeness of freshened grade cohorts in future waves. To achieve a thoroughly representative tenth grade (1990) and twelfth grade (1992) sample comparable to the High School and Beyond 1980 sophomore cohort (or, for 1992, the HS&B 1980 senior cohort and the base year of NLS-72), the NELS:88 follow-up samples must approximate those which would have come into being had a new baseline sample independently been drawn at either of the later

time points. In 1990 (and 1992) one must therefore freshen, to give "out of sequence" students (for example, in 1990, those tenth graders who were not in eighth grade in the spring of 1988) a chance of selection into the study. One must also accommodate excluded students whose eligibility status has changed, for they too (with the exception of those who fell out of sequence in the progression through grades) would potentially have been selected had a sample been independently drawn two years later, and must have a chance of selection if the representativeness and cross-cohort comparability of the follow-up sample is to be maintained. Thus, for example, if a base year student excluded because of a language barrier achieves the level of proficiency in English that is required for completing the NELS:88 instruments in 1990 or 1992, that student should have some chance of re-entering the sample.

A substantial subsample of the base year ineligibles was, accordingly, followed in 1990 and 1992, to reassess eligibility status and gather information about excluded students' demographic characteristics, educational paths, and life outcomes. Data on persistence in school to be obtained from this subsample has been used to derive an adjustment factor for national estimates of the eighth grade cohort's dropout rates between spring of 1988 and spring of 1990, and from 1988 and 1990 to 1992.

The base year ineligibles study largely compensates for population undercoverage. Small populations who remain outside the baseline sampling frame include students who are educated at home or in private tutorial settings, those who are in excluded categories of schools[13] and those who have dropped out of school before reaching the eighth grade.

Table 1 shows that by 1992, a substantial portion of the sample of base year ineligible students had been reclassified as eligible. Excluded students who were later classified as eligible were included in NELS:88 follow-up surveys.

Reclassifications reported in Table 1 reflect multiple phenomena. In some cases--and presumably this is particularly the case for the language exclusions--reclassification reflects change in the eligibility status of the sample member over time. In other cases, change represents the unreliability of exclusion judgments, particularly for exclusion reasons that are more open to interpretation (e.g., mental as opposed to physical handicaps) or that apply to individuals at the margin of the classification--different individuals were asked to assess eligibility at different points in time. Finally, some of the change registered in Table 1 reflects the fact that in the follow-ups we provided more detailed interpretation for the guidelines, so that the validity of exclusion judgments would be enhanced. All in all, however, if any individuals in the target population are to be subject to exclusion from the baseline of a longitudinal study, it is of some importance to reassess their eligibility over time, particularly, in a school-based survey, if the panel is to represent additional grade cohorts.

Table 1: 1992 Status Ns of 1988 Excluded Students

1988 reason for exclusion:	ELIG.	INELIG.	OUT OF SCOPE	N.A.	SAMPLING ERROR
language	125	22	25	30	23
physical	13	9	0	1	1
mental	166	140	5	25	16
unknown	30	15	2	10	16
TOTAL	334	186	32	66	56

* N.A. = status not ascertained.

2.2 Representativeness and New Grade Cohorts: Sample Freshening.

Pearson (1989) notes that a potential limitation of longitudinal samples is that they may provide estimates of the population from which they were originally drawn, but not of the current population. It is of interest to follow a sample of 1988 eighth graders. Nevertheless, an eighth grade panel two years later will not by itself provide a representative sample of the nation's high school sophomores, nor four years later a representative sample of seniors. Representative sophomore and senior samples are analytically desirable at all three levels of NELS:88 analysis *First*, it is desirable to be able to make cross-sectional generalizations about the nation's sophomores in 1990 and seniors in 1992. *Second*, it is desirable to be able to make intercohort comparisons between HS&B 1980 sophomores and 1990 NELS:88 sophomores; between NLS-72 (1972) and HS&B (1980) seniors and NELS:88 (1992) seniors; and between the transcript records of HS&B (1982), NAEP (1987 and 1990), and NELS:88 (1992) seniors. *Third*, it is desirable to be able to conduct longitudinal analyses of 1990 sophomores two, four, and more years later, and of 1992 seniors two and more years later.

Hence a major sampling objective of NELS:88 was to create a valid probability sample of students enrolled in tenth grade in the spring term of the 1989-1990 school year and of students enrolled in the twelfth grade in the spring term of the 1991-92 school year. This goal was achieved by a process we have termed "freshening." The 1990 freshening procedure was carried out in four steps:

1. For each school that contained at least one base year 10th grade student selected for interview in 1990, a complete alphabetical roster of all 10th grade students was obtained.

2. For each base year sample member, we examined the next student on the list; if the base year student was the last one listed on the roster, we examined the first student on the roster (that is, the roster was "circularized").

3. If the student who was examined was enrolled in the 8th grade in the U.S. in 1988, then the freshening process terminated. If the designated student was not enrolled in the 8th grade in the U.S. in 1988, then that student was selected into the freshened sample.

4. Whenever a student was added to the freshened sample in step 3, the next student on the roster was examined and step 3 was repeated. The sequence of steps 3 and 4 was repeated (adding more students to the freshened sample) until a student who was in the 8th grade in the U.S. in 1988 was reached on the roster.

At a given first follow-up school, the freshening process could yield zero, one, or more than one new sample member. Altogether, 1,229 new students were added to the tenth grade sample—on average, just less than one student per school.[14] This procedure was repeated in 1992, to generate a probability sample of the nation's high school seniors.

This freshening procedure is an essentially unbiased method[15] for producing a probability sample of students who were enrolled in the tenth grade in 1990 (or twelfth grade in 1992) but were not enrolled in the eighth grade in the U.S. in 1988. There is a very small bias introduced by the omission of eligible tenth (or twelfth) graders attending schools that included *no* students who were eighth graders in 1988. There is an additional small bias introduced by not freshening on the members of the sample of base year ineligibles. All other 1990 sophomores (or 1992 seniors) who qualify for the freshening sample have some chance of selection. This is because every student who was in the tenth grade in 1990 (or twelfth grade in 1992) but not in the eighth grade in 1988 is linked to exactly one student who was a 1988 eighth grader—this is the 1988 eighth grader who would immediately precede the candidate for the freshening sample on a circularized, alphabetical roster of tenth graders at the school. Because each 1988 eighth grader had a calculable, non-zero probability of selection into the base year and first follow-up samples, we can calculate the selection probabilities for all students eligible for the freshening sample. Thus, the freshening procedure produces a student sample that meets the criterion for a probability sample.

The NELS:88 school sample in 1990 and 1992—the schools to which 1988 eighth graders matriculated—was of course not a nationally representative sample of schools. However, for a select subset of schools, in order to provide a basis for studying school effects, feeder pattern information was collected so that tenth grade school selection probabilities could be approximated, and student samples augmented to make them robust and representative of the school's tenth grade class.[16]

2.3 Nonresponse Error as a Potential Source of Bias: Measures to Maximize Response Rates.

Cumulative nonresponse poses a special threat to longitudinal studies. Some individuals are missed in the baseline measurement, and may enter the study late. Other individuals may be lost, through mobility and the inability to locate them at a later date, or may cease to participate in the study. Still others may participate in the baseline, become temporarily out of scope by leaving the country or become nonrespondents by refusing to participate in the initial follow-up, then re-enter the study in a later follow-up. A longitudinal study must maximize the number of individuals who have data at all data points. Although weighting may help to adjust for nonresponse, the representativeness of the panel depends, in the final analysis, on maintaining high participation rates.

NELS:88 Response Rates. High response rates have been achieved by the study. In the NELS:88 base year (1988) 93.1 percent of selected eighth graders participated. In the NELS:88 first follow-up (1990), 93.9 percent of student and dropout sample members (19,264 of

20,524) took part. In the second follow-up, 90.7 percent of student and dropout sample members took part.

However, from the point of view of longitudinal analysis, a more critical statistic is the proportion of the sample with data at all time points (or, the proportion of baseline participants with data for all follow-ups). Of the 18,261 base year participants retained in the first follow-up, 17,424–or 95.4 percent–were successfully resurveyed. From this base of eighth grade cohort members with both (1988 and 1990) data points, 95.1 percent were resurveyed in the second follow-up.

Table 2A shows overall and subgroup results for the base year-first follow-up respondents for whom a reinterview was attempted in 1992. While, as noted above, around 95 percent were successfully resurveyed (that is, completed a student or dropout questionnaire) in 1992 and thus have data for all three waves, far fewer (72 percent) completed the cognitive test in all three rounds. Table 2B depicts the across-round questionnaire completion status of base year-first follow-up participants who were second follow-up *students*, and the likelihood that school contextual data was available for them for all three rounds. These tables show that completion rates were very similar across different school control types, urbanicity, region, and high and low minority enrollment, and that similar response rates were obtained for members of different racial and ethnic groups.

However, even with these high rates of success in baseline and follow-up data collection, the proportion of the 1988 eighth grade cohort in 1992 with all three data points drops to 84 percent (16,489 of 19,645) when all students missing one or more data point owing to base year, first or second follow-up nonresponse or any other source of sample attrition–being deceased, sample members who suffered grave impairments in the course of the study that did not permit them to be surveyed, individuals out of scope for either follow-up round by virtue of being outside the country–are factored in.

Overall, then, NELS:88 has achieved reasonably high student panel response rates. In addition, final weights have been adjusted for nonresponse, using nonresponse adjustment cells based upon combinations of classification values reflecting race, gender, and data collection status (e.g., dropout; in school in expected grade; in school in another grade; and so on).[17]

Means of Achieving High Response Rates. The means by which these high response rates were achieved may be concisely summarized. Most individuals changed schools, and many changed home addresses, between the base year and the follow-up surveys. About 99 percent of students were successfully traced between the base year and first follow-up, whereupon clusters of students were subsampled to reduce, for cost reasons, the number of high schools to be included in the study. The ability to successfully trace individuals was based upon extensive locating information collected in the base year from both students and parents. This locating information included name, address and telephone number for the student, each parent, and the family's closest relative or friend who did not live in the household. Eighth grade students were also asked to indicate what school they expected to be attending two years later. Tracing was carried out at two levels: first, it was ascertained if the sample member was at the expected school. If not, household information was used to locate the individual. In order to find base year nonrespondents (about 7 percent of the sample did not complete a 1988 student questionnaire and hence did not provide locating information), in addition to conventional survey locating sources, information about the schools matriculated to by the eighth grader's classmates was also utilized. Tracing procedures were repeated in the second follow-up, though between tenth and twelfth grade there is less dispersion to new schools and it was not necessary to further subsample students.

TABLE 2A

NELS:88 Second Follow-Up student survey results for Base Year– First Follow-Up panel participants

	Student/Dropout questionnaire (BY, F1 and F2) Completion rates		Student/Dropout cognitive test[a] (BY, F1 and F2) Completion rates	
	Weighted	Unweighted	Weighted	Unweighted
Total				
Participated	94.7	95.1	69.6	72.2
Selected	16,489[b]		11,902	
School type[c]	17,337		16,489	
Public				
Catholic	94.3	94.7	69.0	71.4
Other private	97.9	97.0	74.1	78.6
Urbanicity[c]	97.4	97.0	73.0	73.7
Urban				
Suburban	93.5	95.1	64.3	69.5
Rural	95.5	95.3	69.1	70.1
Region[c]	94.8	94.9	74.6	77.2
Northeast				
South	94.8	95.1	70.3	71.3
Midwest	94.1	94.5	68.2	73.1
West	95.7	96.0	74.9	76.4
Ethnicity	94.6	95.1	63.7	65.7
Asian/PI				
Hispanic	93.3	95.0	71.5	71.9
Black	93.1	94.4	63.9	65.5
White	92.4	92.6	59.6	67.0
Am. Indian	95.5	95.7	72.1	74.2
Refused/Missing[d]	94.1	91.3	64.8	64.0
Minority schools[c]	81.1	75.0	38.3	55.6
Schools with more than 19% minority students	92.2	93.5	55.1	59.3
Schools with less than 19% minority students	95.0	95.3	71.0	73.5

[a] Cognitive test coverage rate for each sample member who has completed a BY student questionnaire, F1 and F2 student/dropout questionnaire.

[b] Sample members who participated in the BY, F1 and F2.

[c] Refers to 8th grade schools.

[d] Refused/Missing refers only to the status of a sample member's ethnicity. It does not refer to student/dropout nonparticipants.

TABLE 2B

NELS:88 Second Follow-Up data collection results for Base Year – First Follow-Up panel participants

	Student questionnaire (BY, F1 and F2) Completion rates		School questionnaire[a] (BY, F1 and F2) Completion rates	
	Weighted	Unweighted	Weighted	Unweighted
Total	95.7	96.1	95.5	95.6
Participated	14,674[b]		13,182	
Selected	15,269		13,783	
School type[c]				
Public	95.4	95.8	95.8	95.7
Catholic	98.2	97.3	94.3	94.8
Other private	97.5	97.1	93.5	95.8
Urbanicity[c]				
Urban	94.4	96.4	93.7	94.7
Suburban	96.2	96.1	94.4	94.3
Rural	95.8	95.9	98.4	98.2
Region[c]				
Northeast	95.2	95.5	94.9	94.6
South	95.8	96.2	95.6	95.9
Midwest	96.2	96.5	97.3	97.8
West	95.5	96.0	93.1	93.2
Ethnicity				
Asian/PI	94.9	95.8	90.2	93.9
Hispanic	94.2	95.8	89.8	91.3
Black	94.3	95.0	95.1	95.3
White	96.2	96.4	96.5	96.5
Am. Indian	93.8	90.9	97.6	97.3
Refused/Missing[d]	74.2	72.7	100.0	100.0
Minority schools[c]				
Schools with more than 19% minority students	92.5	96.3	90.7	90.0
Schools with less than 19% minority students	96.0	94.4	96.0	96.2

[a] School questionnaire coverage rate for each student who completed a BY, F1, and F2 student questionnaire.

[b] Panel *students* only.

[c] Refers to 8th grade schools.

[d] Refused/Missing refers only to the status of a sample member's ethnicity. It does not refer to student nonparticipants.

In order to survey students, contractor (NORC) staff administered the survey forms at a date agreeable to the school. Make-up sessions were conducted for students who missed the initial survey session. Dropouts and chronic absentees were pursued outside school. Such individuals were invited to group sessions and provided reimbursement for their travel expenses, or were interviewed in their households, over the telephone or in person.

In rare instances, NELS:88 has made use of respondent fees. For example, some dropouts received a monetary incentive, as did some high burden teachers (teachers who had to rate an unusually high number of NELS:88 students such that their burden of questionnaire completion might be two hours or more). School coordinators were given a modest honorarium (normally $25) for assisting with survey activities (for example, supplying annotated rosters, arranging space, and so on), but neither schools nor students were ever paid for their participation

3. Other Sources of Survey Error

When all is said and done, it is the total variable error and bias of a survey estimate that is critical (see Kish, 1965; Andersen, Kasper, and Frankel, 1979; Groves, 1989). From the point of view of total survey error, our discussion thus far is incomplete. It may be useful to identify additional sources of survey error, though space limitations do not permit us to address them.

There are various "repeated measurement" problems in longitudinal surveys. One of these problems is that of panel effects.[12] We do not believe that problems associated with repeated measurements (such as remembering past responses to individual items) are likely to be a difficulty, both because of the sheer number of test and questionnaire items asked, and the two year intervals between data collections. However, participation in a longitudinal study in theory may influence the survey member's subsequent behavior or attitudes.

There are many sources of measurement error. The validity of responses to the NELS:88 eighth grade questionnaire items has been examined in Kaufman, Rasinski, Lee and West (1991), which compares parent and student reports. Transcript and student reports were compared for the HS&B data by Fetters, Stowe and Owings (1984). Psychometric issues in the base year tests are addressed in Rock and Pollack (1991) and in a forthcoming second follow-up psychometric report.

Our earlier discussion dealt with unit nonresponse as a problem of maintaining individual participation across rounds. However, school nonresponse in the base year, and item nonresponse across the survey instruments, also are important nonresponse issues. To the extent that students at noncooperating base year schools may have differed from students at cooperating schools, student level bias is introduced that persists through subsequent waves of observation. Base year school nonresponse is documented and analyzed in the *NELS:88 Base Year Sample Design Report*.

Item nonresponse rates and patterns are documented in the various NELS:88 user's manuals. In general, missing data have not been imputed in the NELS:88 dataset. Although item response rates in NELS:88 are generally high, item nonresponse propensities vary with student characteristics (e.g., race, gender, test quartile), and hence may be a source of bias.

Finally, our discussion has not dealt with the important consideration of sampling error. Design effects for NELS:88 are documented in the various user's manuals. In this respect, dispersion of the student sample after eighth grade has been both a blessing and a curse for NELS:88. The high costs of following dispersed students required that we subsample students in the first follow-up; subsampling increases design effects. At the same time, the general tendency in a longitudinal study is for design effects to decrease over time, as dispersion reduces the original clustering.

Bibliography

Andersen, R., Kasper, J., and Frankel, M.R. (1979). ***Total Survey Error.*** San Francisco: Jossey-Bass.

Babbie, E.R. (1973). ***Survey Research Methods.*** Belmont, CA: Wadsworth.

Bradburn, N.M., Rips, L.J., and Shevell, S.K. (1987). "Answering Autobiographical Questions: The impact of Memory and Inference on Surveys." ***Science***, 239:157-161.

CBO (Congress of the United States, Congressional Budget Office). (1990). ***Services of Support for Adolescent Mothers.*** Washington, D.C.: Author.

Duncan, G.J., Hill, M.S., and Hoffman, S.D. (1988). "Welfare Dependence Within and Across Generations." ***Science***. 239:467-471.

Fetters, W.B, Stowe, P., and Owings, J.A. (1984). ***Quality of Responses of High School Students to Questionnaire Items.*** Washington, D.C.: National Center for Education Statistics.

Groves, R.M. (1989). ***Survey Errors and Survey Costs***. New York: Wiley.

Ingels, S.J., et al. ***NELS:88 Data File User's Manuals***. Base Year: Student, Parent, Teacher, School (NCES, 1990); First Follow Up: Student, Dropout, Teacher, School (NCES 1992); Second Follow-Up: Student, Dropout, Parent, Teacher, School, Transcript (NCES 1994).

Kasprzyk, D., Duncan, G., Kalton, G., & Singh, M.P., eds. (1989). ***Panel Surveys***. New York: Wiley.

Kaufman, P., Rasinski, K.A., Lee, R., and West, J. (1991). ***Quality of Responses of Eighth-Grade Students in NELS:88.*** Washington, D.C.: National Center for Education Statistics.

Kish, L. (1965). ***Survey Sampling***. New York: Wiley.

Kish, L. (1987). ***Statistical Design for Research.*** New York: Wiley.

Lee, V.E., and Loeb, S. (1994). "Where Do Head Start Attendees End Up? One Reason Why Preschool Effects Are Not Sustained." Paper presented to the American Educational Research Association, Annual Meeting, New Orleans, Louisiana, April 7, 1994.

McGrew, K.S., Thurlow, M.L., and Spiegel, A.N. (1993). "An Investigation of the Exclusion of Students With Disabilities in National Data Collection Programs." vol.15. ***Educational Evaluation and Policy Analysis.***

Pearson, R. W. (1989). "The Advantages and Disadvantages of Longitudinal Surveys." ***Research in the Sociology of Education and Socialization***. vol. 8.

Rock, D.A. and Pollack, J.M. (1991). ***Psychometric Report for the NELS:88 Base Year Test Battery.*** Washington, D.C.: National Center for Education Statistics.

Rosenzweig, M.R., and Wolpin, K.I. (1993). "Maternal Expectations and Ex Post Rationalizations: The Usefulness of Survey Information on the Wantedness of Children." ***Journal of Human Resources***, vol. XXVIII.

Spencer, B.D., Frankel, M.R., Ingels, S.J., Rasinski, K.A., and Tourangeau, R. (1990). ***NELS:88 Base Year Sample Design Report.*** Washington, D.C.: National Center for Education Statistics.

Spencer, B.D. (1991). "Eligibility/Exclusion Issues in the 1990 Trial State Assessment." In G. Bohrnstedt et al., ***Assessing Student Achievement in the States: Background Studies for the Evaluation of the NAEP Trial State Assessment.*** National Academy of Education. Stanford, California: Stanford University.

Spencer, B.D., and Foran, W. (1991). "Sampling Probabilities for Aggregations, With Application to NELS:88 and Other Education Longitudinal Surveys." ***Journal of Educational Statistics***, 16(1).

END NOTES

1Longitudinal studies are prospective, in that they record new events for *individual* units under observation more or less as they occur. A repeated cross-sectional study can also be prospective, and measure change over time at the *group* level. A single (not repeated) cross-sectional study can measure change in *individual* units over time by assuming a retrospective focus--for example, by relying on individual memories to reconstruct an historical record of events and statuses. While retrospective studies are appropriate for many purposes, when used for other purposes there may be significant reliability problems. For a concise summary of issues concerning the reliability of retrospective reports, see Bradburn, Rips and Shevell (1987). For a useful comparison of prospective and retrospective studies, see Kish (1987) pp. 178-181. For a recent example (*ex post facto* reports of wantedness of children) of an analysis of the degree to which retrospectively-obtained survey data provides unbiased estimates, see Rosenzweig and Wolpin (1993).

2We take a representative sample to be a probability sample drawn, with known selection probabilities for sample units, from the target population.

3Groves (1989, chapter 3) provides a useful discussion of coverage error and its consequences.

4There are many ways to characterize sample designs that measure change over time, and the term "longitudinal" has both strict and looser usages. Kish (1987, Chapter 6) presents a useful typology of designs for covering time spans across populations, and Babbie (1973, pp.62-66) some standard terminology.

5For individuals interested in conducting trend analyses of NLS-72, HS&B and NELS:88 data, further information on content and design similarities and differences between these three studies is presented in the second follow-up student component data file user's manual. Comparison of sophomore cohort dropouts across HS&B and NELS:88 is discussed in the dropout component user's manual, while high school transcript comparisons (HS&B, NAEP 1987, NAEP 1990, NELS:88) are discussed in the transcript user's manual.

6For each cohort, the timing of the last follow-up assumes that the tentatively scheduled date for the fourth follow-up -- 1997 -- will hold.

7Repeated cross-sections compound sampling error. This is the case because a repeated cross-section is drawn two or more times; change measurement must contend with the fact that differences in multiple sample means will in part be a function of the sampling errors associated with each independent sample. In contrast, a longitudinal sample is drawn but once. However, for a freshened cohort study such as NELS:88, some sampling error may be associated with the freshening process. Hence when NELS:88 data are analyzed in the aggregate as a rolling cross-section, some of this advantage of a longitudinal design is lost.

8The 1988-90 dropout rate for the expanded (eligible + ineligible) NELS:88 eighth grade cohort was 6.8 percent for 1988-90. Excluding students who dropped out between 1988 and 1990 (or left the country), the dropout rate between 1990 and 1992 was 7.6 percent. However, the proportion of 1988 eighth graders who were dropouts in the spring of 1992 was 11.6 percent. (Of course, the number of sample members experiencing brief duration dropout spells or dropout *events* is even further undercounted by virtue of using a cohort status [spring to spring across two years] measurement.)

9On NLSY (the BLS National Longitudinal Survey of Youth which began in 1979), see CBO, 1990. On PSID (Panel Study of Income Dynamics, a nationally-representative sample of families, begun in 1968) results, see Duncan, Hill, and Hoffman, 1988.

10Researchers (see Lee and Loeb, 1994) have used the response to this retrospective item in conjunction with NELS:88 measures of school quality to inquire into whether Head Start participants are more likely than their peers to attend lower quality elementary/middle schools, a possibility that could in part explain why academic gains from Head Start may fade out over time.

11For further details of school-level exclusion, see Spencer, Frankel, Ingels, Rasinski, & Tourangeau, 1990, p.10.

12Recent investigations of the extent to which students with disabilities are allowed to participate in major national data collection programs suggest that 40-50 percent of students with disabilities are typically excluded from major assessments, though students with disabilities are included to a greater degree in data collections that do not require the completion of cognitive tests (McGrew, Thurlow, & Spiegel, 1993). Additional numbers of students are excluded from assessments or other state and national education data collection programs owing to language barriers to participation. For a parallel discussion based on the NAEP trial state assessments, see Spencer in Bohrnstedt, ed., 1991.

13According to Office of Special Education figures reported in the *Digest of Education Statistics, 1992*, Table 51, 5.5 percent of special education students receive services in separate schools or residential facilities, while .8 percent are in a homebound or hospital environment. Not all of these individuals are in graded programs. Separate facilities tend in particular to be available for comparatively rare populations such as individuals with severe visual or hearing impairments, and for emotionally disturbed students whose presence might impede regular classroom activities. Most students who are doubly physically disabled by being both deaf and blind are educated in special facilities.

14Some of these freshened students were dropped in the subsampling process either because they themselves were not included in the subsample or because the base year student to whom they were linked was not included. Some 1,043 students selected through the freshening procedure remained in the final first follow-up sample. In the second follow-up (1992), 244 students were added through freshening.

15See Kish (1965) for a discussion of the half-open interval procedure that underpins this approach.

16A strategy for estimating a school's selection probabilities under these circumstances is sketched in Spencer and Foran, 1991.

17Again, however, while weights can compensate for nonresponse by correcting errors in the population estimates for particular subgroups, they do not correct nonresponse *bias* within subgroups. For example, weighting can adjust for the fact that male eighth graders responded to NELS:88 at a lower rate than did their female classmates, but do not address bias that may be present if male responders and nonresponders differed in the very characteristics inquired into by the base year student questionnaire.

18Discussions of longitudinal conditioning or panel effects (also known as "time in sample bias" or "panel conditioning")--for example, whether strong effects potentially exist or could affect data quality--may be found in Kasprzyk, D., Duncan, G., Kalton, G., & Singh, M.P., eds. *Panel Surveys*, 1989 (New York: Wiley). See especially contributions by B. Bailar; D. Cantor; D. Holt; A. Silberstein and C. Jacobs; L. Corder and D. Horvitz; and J. Waterton and D. Lievesley.

DISCUSSION

Gary M. Shapiro
Abt Associates

The two papers in this session discuss the methodology of two of the most important longitudinal surveys ever conducted, and thus are of considerable interest. This session is on longitudinal surveys, but in fact the papers relate to two topics that are in the Statistical Policy Working Paper Series in addition to longitudinal surveys: nonresponse and survey coverage.

Both of these surveys have been very well designed. The sample design issues for the surveys have been carefully investigated, and are well documented both here and in earlier publications. Specific comments on each paper follow.

I. National Medical Expenditure Survey Paper by Steven Cohen

I will separately discuss the main topics covered in the paper. The first topic was the screening procedure and differential sample selection by demographic characteristics. The interviewers returned to screened-in addresses rather than follow individual persons. The paper provides a careful, detailed discussion of the consequences of following addresses, and contains a good discussion of the pro's and con's of following addresses vs. persons.

I have a couple of side observation on this topic. First, there were 722 units that were occupied at time of screening that became vacant by time of interview. However, there were only 448 vacant units that became occupied by time of interview. In theory, one would expect these two figures to be equal. The difference is probably statistically significant. This most likely signifies some coverage loss - either some of the 722 were not really vacant, or more than 448 of the vacants actually became occupied. I do not find this discrepancy to be alarmingly large, but it is evidence of a minor problem.

My second observation is on the use of screening. There were about 28,000 occupied addresses that were screened, with about 15,000 interviewed the next year. The smallest sampling rate in any demographic group was about 40%. This makes me wonder how cost effective the screening was. Fewer demographic categories and a simpler differential sampling scheme would quite possibly have been about as effective. It might have been preferable to undertake a very simple sub-sampling scheme at the time of interview instead of having entirely separate screening interviews. This has obvious implications for NMES-3.

I have no comments on the paper's second major topic, the return visits to households that were nonresponse for screener interviews, in an effort to improve response rates.

The third topic dealt with persons who were nonrespondents for only some of the waves. The paper discusses the characteristics of these people and estimation considerations for them. About 6.2 % of all participants missed at least one wave but not all waves, and about half of them were missing more than 1/3 of the data. The original paper states "When the level of partial response is low, it is often preferable to treat respondents as complete nonrespondents", as opposed to using imputation for missing waves. I invited audience members to discuss under what circumstances nonresponse adjustment should be done as opposed to imputation. One or two audience members expressed the view that imputation might have been preferable in this survey, and indicated some level of disagreement with the quotation above.

The fourth major topic was on the estimation for the institutional sample, when people move in and out of institutions during the year. I have no significant comments on this topic.

II. National Education Longitudinal Study of 1988(NELS:88) Paper by Steven Ingels and Jeffrey Owings

There are three main topics in this paper, two dealing with coverage issues and one dealing with nonresponse. I'm extremely pleased to see the emphasis here on coverage. Coverage is a major problem in many surveys, but it usually receives much less attention than nonresponse and many other less important topics.

The paper first discussed students that were excluded from the sample, with emphasis on disabled and non-English speaking student exclusions. These are reasonable exclusions that I think many survey practitioners would not have worried at all about excluding - I applaud the authors for their concern about the effects of the exclusions. I very much liked the paper's discussion of the effects of this undercoverage on survey results - more analyses of this type are needed.

The decision to follow students who were ineligible in 1988 in the 1990 and 1992 interviews is exemplary. I think most survey practitioners would probably not have undertaken the expense. It's unfortunate that except for the language exclusion category, the follow-up was not very effective.

The second topic of the paper was on sample freshening, to assure that all tenth graders in 1990 were covered in the survey. This again is a coverage issue that many surveys would not have worried about, and it is a real pleasure to see this level of care in survey implementation. As a side note, the survey estimated the number of schools attended by tenth graders in 1990 using complex methodology in Spencer and Foran(1991). I believe a better approach to this type of estimation issue is one used in the Survey of Income and Program Participation for estimating the number of households when following individuals (see Huang,1984, and Ernst,1989.)

The third topic dealt with two nonresponse issues. I have no comments on this section.

The last section of the paper briefly mentions several other sources of error, including a statement that item nonresponse was rarely imputed for. I note that even when there is not explicit imputation in a published table, there is implicit imputation in the sense that a reader generally assumes that nonrespondents are distributed like respondents. Thus, even a very crude explicit imputation is usually better than no imputation at all.

REFERENCES

Ernst, L. (1989), "Weighting for Longitudinal Household and Family Estimates", in "Panel Surveys", edited by D. Kasprzyk, G. Duncan, G. Kalton, and M.P. Singh, pp. 139-159, Wiley and Sons.

Huang, H. (1984), "Obtaining Cross-sectional Estimates for a Longitudinal Survey: Experience of ISDP", Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 670-675.

Spencer, B. and Foran, W. (1991), "Sampling Probabilities for Aggregations, With Application to NES:88 and Other Educational Longitudinal Surveys", Journal of Educational Statistics, Vol. 16, No.1, pp. 21-33.

DISCUSSION

James M. Lepkowski
University of Michigan

Panel surveys are used to address many scientific and policy issues. The primary focus of many panel surveys is to assess change over time and examine causal mechanisms in various phenomenon. A wide range of designs are employed in these panel surveys, reflecting the diverse uses to which panel survey data may be put.

The presentation of papers on the design of two panel surveys in this session provides an opportunity to compare and contrast features of panel survey design. The two panel surveys described in these papers have very different topics, populations, sampling frames, and sources of survey error. The National Education Longitudinal Study of 1988 (NELS 88) is designed to provide data on education for a cohort of eighth graders followed over a six year period, sampling students from schools. The National Medical Expenditure Survey (NMES) is a survey of health and health care utilization and coverage among the U.S. civilian non-institutionalized population and those residing in nursing homes. The NMES uses a national area sample frame supplemented with a national nursing home list. There are similarities in design features between these surveys, but there are also many interesting differences. Comparing and contrasting these surveys illustrates the value and flexibility of panel survey design.

The two surveys, for example, have different statistical estimation purposes. The NELS 88 provides data that can be used to create cross-sectional estimates at each data collection time period, inter-cohort comparisons over time, and individual change through the repeated measurement of the same individuals. These are typical goals of many panel survey designs, attempting to answer questions about a single time period through cross-sectional estimates and about changes over time in a group or among individuals of the group. The NMES uses the panel survey design for a different purpose, to provide precise and accurate measurement of health and health care utilization and coverage through the accumulation of retrospective reports from a panel of respondents. By limiting the time between observations, NMES reduces the size of recall errors. There is little if, any interest, in assessing change at an individual or group level in the NMES.

Different purposes have lead to different design features. NELS 88 uses a longer period between interviews, following the cohort of eighth graders every two years. Characteristics measured in the NELS 88 are not expected to change substantially in a shorter interval, and thus more frequent data collection does not appear necessary. NMES interviews reporting units (essentially households) every three or four months over a 16 month period.

This frequent interviewing schedule allows NMES to measure continuously the rapidly changing health characteristics of the sample. The longer period between interviews would pose a larger problem for the NELS 88 if it were not for the fact that a large share of the NELS 88 sample each interview is students in school being followed. Information on contact persons is collected to assist in the tracking of students who leave school or move to another location and school. The NMES also collects contact information to assist tracking efforts, but NMES reporting unit members appear to be more difficult to track than the NELS 88 student sample.

An important methodological difference between the two surveys is the need for screening and sampling at varying rates subgroups of the population. NELS 88 does not need to sample any particular subgroups at higher rates than others. NMES attempted to sample poorer persons at higher rates through an interview conducted before the primary data collection began. A substantial share of the NMES paper is devoted to an examination of the screening procedures and their effectiveness.

The two surveys address the issue of the changing nature of the population that the sample represents quite differently. In panel survey design, "representation" of the population by the sample is hindered by loss due to non-response and failure to recruit into the sample persons who are recently joined the population such as immigrants. The NELS 88 sample is "refreshed" every interview period by a selection of additional students who could not have been selected at any of the previous rounds of data collection because they were not members of the U.S. eighth grade student population in 1988. The refreshing process is designed to improve the coverage of the NELS 88 sample of students who were eighth graders in 1988 at any given future interview round. The NMES covers a much shorter time period during which fewer changes to the population that could affect the quality of the NMES data occur. For instance, immigration changes to the U.S. civilian population is not expected to be an important departure from complete coverage during the time period of the NMES.

The NELS 88 and NMES are both subject to non-response, and both surveys make considerable effort to reduce non-response rates. The NELS 88 faced difficulty recruiting school districts and schools, the primary sampling units in the selection, into the sample. When a sample school district or school refused to participate, NELS 88 substituted a school district or school as similar as possible to the non-participating school. Student participation within cooperating schools was quite high, with very low non-response at the initial round and all subsequent rounds of data collection. High response rates were maintained despite the difficulties of tracking school drop-outs. NMES did not lose any primary sampling units to non-response in the household sample portion of the sample, although it did experience losses of reporting units and reporting unit members at first and subsequent interviews. NMES also experienced loss of primary sampling units

in the nursing home survey portion of the sample, as well as difficulty obtaining interview data for persons who were extremely ill.

Many panel surveys compensate for non-response losses through weighting for unit non-response and imputation for item non-response. The NELS 88 analyzed first and subsequent round non-response, but it is not clear from the paper if non-response weights are employed. No imputation has been done to compensate for item non-response rates, which are reported to be low. The NMES discards persons from the sample who provided reports for less than one-third of the target year. Non-response compensation weights are employed to adjust for the these and all other persons who failed to respond to the survey adequately. As in the NELS 88, imputation is not employed to compensate for item non-response. As a result, both surveys implicitly impute for missing items by using available data, effectively an imputation of the mean values of the responses for the missing items.

The NELS 88 and NMES are examples of well-designed and carefully executed panel surveys. Since both have different goals and topics, the panel designs employed by each are quite distinct. Those interested in understanding more fully the features and complexities of panel survey design can learn much from a reading of these two papers.

Reports Available in the Statistical Policy Working Paper Series

1. Report on Statistics for Allocation of Funds (Available through NTIS Document Sales, PB86-211521/AS)
2. Report on Statistical Disclosure and Disclosure-Avoidance Techniques (NTIS Document sales, PB86-211539/AS)
3. An Error Profile: Employment as Measures by the Current Population Survey (NTIS Document Sales PB86-214269/AS)
4. Glossary of Nonsampling Error Terms: An Illustration of a Semantic Problem in Statistics (NTIS Document Sales, PB86-211547/AS)
5. Report on Exact and Statistical Matching Techniques (NTIS Document Sales, PB86-215829/AS)
6. Report on Statistical Uses of Administrative Records (NTIS Document Sales, PB86-214285/AS)
7. An Interagency review of Time-Series Revision Policies (NTIS Document Sales, PB86-232451/AS)
8. Statistical Interagency Agreements (NTIS Documents Sales, PB86-230570/AS)
9. Contracting for Surveys (NTIS Documents Sales, PB83-233148)
10. Approaches to Developing Questionnaires (NTIS Document Sales, PB84-105055/AS)
11. A Review of Industry Coding Systems (NTIS Document Sales, PB84-135276)
12. The Role of Telephone Data Collection in Federal Statistics (NTIS Document Sales, PB85-105971)
13. Federal Longitudinal Surveys (NTIS Documents Sales, PB86-139730)
14. Workshop on Statistical Uses of Microcomputers in Federal Agencies (NTIS Document Sales, PB87-166393)
15. Quality on Establishment Surveys (NTIS Document Sales, PB88-232921)
16. A Comparative Study of Reporting Units in Selected Employer Data Systems (NTIS Document Sales, PB90-205238)
17. Survey Coverage (NTIS Document Sales, PB90-205246)
18. Data Editing in Federal Statistical Agencies (NTIS Document Sales, PB90-205253)
19. Computer Assisted Survey Information Collection (NTIS Document Sales, PB90-205261)
20. Seminar on the Quality of Federal Data (NTIS Document Sales, PB91-142414)
21. Indirect Estimators in Federal Programs (NTIS Document Sales, PB93-209294)
22. Report on Statistical Disclosure Limitation Methodology (NTIS Document Sales, PB94-165305)
23. Seminar on New Directions in Statistical Methodology (NTIS Document Sales, PB95-182978)

Copies of these working papers may be ordered from NTIS Document Sales, 5285 Port Royal Road, Springfield, VA 22161 (703)487-4650