



U.S. Department of Health and Human Services  
Assistant Secretary for Planning and Evaluation  
Office of Disability, Aging and Long-Term Care Policy

**REPORT ON PROBLEMS AND  
SUGGESTIONS FOR  
IMPROVING THE NLTCS  
FILES FOR 1982 AND 1984**

May 17, 1989

## **Office of the Assistant Secretary for Planning and Evaluation**

The Office of the Assistant Secretary for Planning and Evaluation (ASPE) is the principal advisor to the Secretary of the Department of Health and Human Services (HHS) on policy development issues, and is responsible for major activities in the areas of legislative and budget development, strategic planning, policy research and evaluation, and economic analysis.

ASPE develops or reviews issues from the viewpoint of the Secretary, providing a perspective that is broader in scope than the specific focus of the various operating agencies. ASPE also works closely with the HHS operating divisions. It assists these agencies in developing policies, and planning policy research, evaluation and data collection within broad HHS and administration initiatives. ASPE often serves a coordinating role for crosscutting policy and administrative activities.

ASPE plans and conducts evaluations and research--both in-house and through support of projects by external researchers--of current and proposed programs and topics of particular interest to the Secretary, the Administration and the Congress.

## **Office of Disability, Aging and Long-Term Care Policy**

The Office of Disability, Aging and Long-Term Care Policy (DALTCP), within ASPE, is responsible for the development, coordination, analysis, research and evaluation of HHS policies and programs which support the independence, health and long-term care of persons with disabilities--children, working aging adults, and older persons. DALTCP is also responsible for policy coordination and research to promote the economic and social well-being of the elderly.

In particular, DALTCP addresses policies concerning: nursing home and community-based services, informal caregiving, the integration of acute and long-term care, Medicare post-acute services and home care, managed care for people with disabilities, long-term rehabilitation services, children's disability, and linkages between employment and health policies. These activities are carried out through policy planning, policy and program analysis, regulatory reviews, formulation of legislative proposals, policy research, evaluation and data planning.

This report was prepared under contract between HHS's Office of Social Services Policy (now DALTCP) and Social and Scientific Systems, Inc. For additional information about this subject, you can visit the DALTCP home page at [http://aspe.hhs.gov/\\_/office\\_specific/daltcp.cfm](http://aspe.hhs.gov/_/office_specific/daltcp.cfm) or contact the office at HHS/ASPE/DALTCP, Room 424E, H.H. Humphrey Building, 200 Independence Avenue, S.W., Washington, D.C. 20201. The e-mail address is: [webmaster.DALTCP@hhs.gov](mailto:webmaster.DALTCP@hhs.gov). The Project Officer was Floyd Brown.

**REPORT ON PROBLEMS AND SUGGESTIONS FOR  
IMPROVING THE NLTCS FILES  
FOR 1982 AND 1984**

Social and Scientific Systems, Inc.

May 17, 1989

Prepared for  
Office of the Assistant Secretary for Planning and Evaluation  
U.S. Department of Health and Human Services

The opinions and views expressed in this report are those of the authors. They do not necessarily reflect the views of the Department of Health and Human Services, the contractor or any other funding organization.

# TABLE OF CONTENTS

<b>INTRODUCTION</b> .....	1
Background.....	1
The National Long-Term Care Surveys for 1982 and 1984 .....	1
The NLTCS User’s Forum .....	2
<b>DIFFICULTIES AND COMPLEXITIES OF THE 1982 AND 1984 NLTCS DATA</b> .....	4
<b>COMPLEXITIES AND INADEQUACIES IN THE WEIGHTS</b> .....	5
Background.....	5
Proposed Improvements.....	5
<b>DATA CLEANING AND EDITING</b> .....	7
Key Variables.....	7
Flap Variables.....	7
Screener ADL/IADL Definitions Versus Detailed Definitions.....	7
Ambiguous “Missing Data” Definitions .....	7
Inconsistent Helper/Household Member/Children Data .....	8
Summary and Check Variables.....	8
Proposed Improvements.....	8
<b>IMPUTATIONS AND THE CREATION OF DERIVED VARIABLES</b> .....	10
Background.....	10
Proposed Improvements.....	10
<b>IMPROVED DOCUMENTATION</b> .....	11
Background.....	11
Proposed Improvements.....	11
<b>TECHNICAL ASSISTANCE</b> .....	13
Background.....	13
Proposed Improvements.....	13

# INTRODUCTION

## Background

Understanding the characteristics and status of the elderly who receive longterm care is an important and growing area of research, as is interest in the extent of their disabilities and their care requirements. As Americans live longer, the extent and types of disabilities and required services are broadening. Anticipating resource requirements of elderly and disabled persons will be even more critical as members of the "baby boom" generation become older. Greater numbers and types of services may be needed because of the larger size of this portion of the population, and public policy will have to provide the necessary framework for meeting these needs.

Various government agencies, the Congress, academics, advocacy groups, and the general public have a pressing need for credible data about the elderly population on which to base policy alternatives and costs. Currently, survey data are available from many sources: the National Health Interview Survey (NHIS) Supplement on Aging (SOA) and the related Longitudinal Survey on Aging (LSOA), the National Nursing Home Survey (NNHS), and the Epidemiologic Followup Study to the National Health and Nutrition Examination Survey (NHANES1), to name a few.

These surveys generally target either the elderly population who live in the community regardless of their health status, or the disabled elderly who live in institutions. None of these data specifically target the disabled elderly who live in the community, and who are in need of a variety of services because of their disabilities.

Medicare data are also available for the elderly-who participate in that program. These data can be linked to many of the surveys mentioned above, and they provide detailed information on expenditures for medical care. They provide little, however, to enhance analyses of the extent and types of disabilities from which Medicare recipients may suffer.

## The National Long-term Care Surveys for 1982 and 1984

The 1982 and 1984 National Long-term Care Surveys (NLTCS) are unique among surveys of the elderly. They represent the elderly population in the U.S., targeting those who are disabled and living in the community. The survey samples were drawn from the Medicare eligibility data collected by the Social Security Administration. By using the weights provided with the survey files, the samples represent that universe of the elderly. The detailed components of the NLTCS focus on persons identified as community based and disabled. These surveys provide rich and detailed data on this important subsegment of the elderly, while the context of the general elderly population is maintained.

Three additional supplements to the NLTCs add to its richness and coverage:

- the Informal Caregivers Survey for 1982;
- the Institutional Component for 1984; and
- the Deceased Component for 1984.

Included in the 1982 survey were detailed questions asked of persons identified as "informal" (unpaid) caregivers to the community-based disabled. These data make it possible to identify characteristics of those helpers who may or may not be relatives of the disabled person.

The 1984 survey also includes an institutional component for the elderly living in facilities like nursing or personal care homes. Combining this supplement with the data for the community-based elderly, it is possible to produce national estimates on the full disabled elderly population. Further, the 1984 survey includes detailed data about those people in the 1982 sample who died between the two survey periods. These data on the deceased were collected from the next-of-kin.

The longitudinal aspects of the 1982 and 1984 surveys permit examination of changes in the health and functional status of the elderly, as well as analyses of their use of various health care services over time. The inclusion of data on the deceased allows description of the medical services and costs associated with these deaths. Taken together, these rich data on the chronically disabled elderly population are comprehensive, and they comprise a unique data source for public policy research.

The NLTCs data are not without their problems, however. Numerous constraints and complications have kept them from being as accessible as they should be. Sponsorship has migrated through several governmental agencies, causing some redefinition of the focus of the effort. A shift in the survey design, from cross-sectional in 1982 to longitudinal in 1984, introduced analytic complexities. Changes in definitions, as well as improvements and enhancements to individual questions, made the use of prior period data more difficult. Attempts to simplify the physical and logical file structures have resulted in a data file that is inefficient to process. Limited budgets and time constraints precluded sufficient data editing and documentation. Although these problems are not insurmountable, individual researchers and users have been left to grapple with them.

## **The NLTCs User's Forum**

Increased interest in, and broader use of the NLTCs files led ASPE to sponsor the National Long-Term Care User's Forum, held in January 1989. At this meeting, representatives from various government agencies involved in the development of the NLTCs, as well as experienced users from the research community, joined together

with other users and potential users to share their experiences with the data. The various strengths, weaknesses, and difficulties with the files were summarized.

The Forum served as an opportunity for the users to ask questions and share information of general interest, and to collectively identify those areas where technical support is most urgently needed. Sessions on file content and design, data editing, data linkage, weighting, documentation, and users' concerns provided a useful setting for interaction among those most interested and knowledgeable about the NLTCs data. The Forum also provided a vehicle for summarizing the most knowledgeable users' positive experiences with these data.

Further, participants were able to ask pressing questions about the 1988/1989 data, which are currently being collected.

Written proceedings from the Forum were developed, which summarized the sessions in terms of the topics discussed, questions asked, answers provided, and concerns expressed. Those proceedings will record what users hope will be the start of a dialogue among various researchers, system developers, and statisticians who are interested in the NLTCs data and their use.

This paper summarizes the problems with and suggestions for improving the NLTCs files. It incorporates many of the concerns and ideas users stated at the Forum. It outlines concrete areas where improvements and increased technical support are needed so that the research community can conduct the most useful and credible studies possible. Eliminating as many of these problems as possible will go a long way toward freeing analysts to do more research with the data, rather than having to do research about the data.

# DIFFICULTIES AND COMPLEXITIES OF THE 1982 AND 1984 NLTCS DATA

Many structural complexities of the NLTCS data base for 1982 and 1984 make the files difficult to use. Some of these are inherent to any longitudinal data, but are more complex because elderly persons are being surveyed. In particular, the elderly are frequently lost to survey due to death or entry into an institution. Some become incompetent, and then a proxy must supply (perhaps less reliable) information in their place.

Changes in the questionnaire itself add to its complexity. Varying definitions of "disabled" (ex. screener, interview) can be confusing to users of the data. The components of the survey were not administered consistently in both periods. On the one hand, enhancements, additions, and deletions to the survey instrument were essential for keeping the data base analytically current. On the other hand, budget cuts have diminished the continuity of certain components of the survey.

These complications have been, for the most part, unavoidable. They do not detract substantially from the richness and breadth of these data, but they add to the difficulty in using the data. Generally, the complexities and difficulties in using the 1982 and 1984 NLTCS fall into five categories:

1. Inaccuracies in, and inadequate documentation about, the weights provided on the files.
2. Inconsistencies among individual data items within instruments, across instruments, and between survey years.
3. Ambiguous non-response data codes for many variables, and missing data for some key variables.
4. Documentation that needs cross-indexing, better accessibility, better readability, and completeness; i.e., better organization.
5. Inadequate technical support, which leaves users unable to receive timely assistance, answers to questions, and access to any current file information.

Each of these categories will be discussed below.



# COMPLEXITIES AND INADEQUACIES IN THE WEIGHTS

## Background

The complex nature of the NLTCS files is compounded by the various weights that are available on the files. Multiple weights are provided because of the multipurpose nature of the files. Weights were carefully calculated to allow cross-sectional analyses for both 1982 and 1984, as well as for longitudinal analyses across the period. Unfortunately, because of the way that the sample for the 1984 file was constructed, the sample weights reflect a population definition that is inconsistent with that represented on the 1982 file.

The 1984 sample comprised the 1982 community-based group without further screening, supplemented by persons who became disabled between 1982 and 1984 and by others who reached age 65 between 1982 and 1984. The 1982 community-based cohort was not rescreened for disability; rather its surviving members were automatically included in the 1984 community-based component.

Further complications occurred because of the institutional component on the 1984 file. The control totals used for determining the sample weights were based on a broad definition of "institution," and thus the institutional population as defined for the 1984 NLTCS may be overstated. Conversely, the 1982 population may be understated because several hundred people identified for inclusion in the community sample entered institutions between the time of sample selection and survey execution. They were therefore excluded from the 1982 survey but the weights did not compensate for this exclusion.

For the deceased component of the 1984 NLTCS, discrepancies arose when linking the deceased population to Medicare records.

## Proposed Improvements

The user community wants a set of weights that can be reliably used with little or no corrections or ambiguity. Users need explicit documentation relating to the construction and proper use of the weights, as well as any associated effect on error variance estimation. Control totals for benchmarking analyses are essential to insure their proper use.

This does not mean that users want to give up detail or control over the types of analyses they can do. Nor does it mean that they are unwilling to deal with complex data. Many sophisticated users are capable of making their own adjustments to

appropriate weights to do hypothesis testing, as well as to account for possible bias due to non-response. Such users must be provided with the tools to do so.

To begin with, the weights for the 1982 NLTCS need to be corrected, to change the reference point of the file from cross-sectional to longitudinal. Then the longitudinal weights for the 1982-84 period need to be reexamined and adjusted to give the user a complete set of consistent, usable weights. The components (i.e. the raw selection probabilities) needed to make adjustments to the weights for particular research applications must be made available. Control totals need to be generated so that users can verify their own analyses. All aspects of the weighting process should be properly documented.

# DATA CLEANING AND EDITING

## Key Variables

Inherent in the use of data from both the 1982 and 1984 files for longitudinal studies is the need for consistency among the identifying and key variables of each file period: the key screener variables for 1982, and the key detailed interview variables for 1982 and 1984. For example, the reported age of a respondent should have a consistent and predictable value between file periods. Similarly, key demographic variables, (e.g., race and sex), should not change between file periods. These basic demographic variables on the NLTCS files need to be examined for consistency.

## Flap Variables

Another general area of users' concerns centers on the detailed questions regarding activities of daily living (ADL) and instrumental activities of daily living (IADL), and the associated check items and "flap" variables. The flap variables were not meant to be used for analysis, but rather as an aid to the interviewers to determine skip patterns in the questions. These variables are not really part of the survey data, and they were intended to provide observational information (by the interviewer). Indeed, portions of the data were initially edited by the Bureau of the Census based on the response patterns reflected by these flap variables. Yet most researchers need ADL and IADL summary variables much like those provided on the flap. Many analysts use the flap items in spite of their weaknesses. Great benefit would derive from constructing a set of functionally equivalent summary variables based on the detailed ADL and IADL questions.

## Screener ADL/IADL Definitions Versus Detailed Definitions

Problems arise because of ambiguity between the screener ADL/IADL questions and the detailed community ADL/IADL questions. Such inconsistencies of definition are not immediately apparent from the documentation. Careful analysis is necessary to define consistent and useful variables since much of the detailed survey is based on responses to these ADL and IADL questions.

## Ambiguous "Missing Data" Definition

Many of the variables on the NLTCS files are difficult to use individually because the "missing data" values associated with them are ambiguous. When a response is "missing," it generally might be one of four responses: not ascertained, don't know, refused to answer, or not applicable. For many research purposes, the "don't know" and

"refused to answer" categories of nonresponse are actually responses, to be distinguished from the other two categories. Currently, users are forced to investigate, understand, and reproduce the so-called "skip logic" themselves, in an effort to untangle the "don't know" from the "not applicable" response categories. Many users want these ambiguities removed, and do not have the resources to perform the complicated, but necessary, preprocessing of the data.

## **Inconsistent Helper/Household Member/Children Date**

Several thousand variables providing information on helpers, household members, and children of respondents are available on the NLTCs data base. Only a limited amount of information exists for any one such person, but the file structure allows for up to 15 persons in each of the three categories for each of the 2 years. Further, a helper can also be a household member. In this case, information is repeated but has not necessarily been checked for consistency. Researchers would greatly benefit from having these variables systematically examined, cleaned, and edited for intra-record consistency. Rules for such editing would be included in expanded and organized documentation, as described below.

## **Summary and Check Variables**

Summary and check variables, which direct the user through the various segments of the survey, need to be checked and possibly edited for each sample person. There should be no confusion about which segment a respondent (or non-respondent) properly belongs in: the community, institutional, or deceased component. The current variables on the files which were meant to identify the various subpopulations, must be aligned for consistency. The documentation must be modified to accurately describe these variables.

## **Proposed Improvements**

In summary, users of the NLTCs files have a need for systematic examination and adjustment of many of the key and general variables on the files. As pointed out at the User's Forum, a great deal of time has been spent by some; users in cleaning the disability data, reconciling the detailed interview data with the check variables, and creating a screener equivalent definition of "disability" from the detailed interview questions. Other researchers will have to go through the same or similar exercises to use these data effectively.

Data editing and cleaning of the NLTCs files would be welcomed by all who use the data, and will go a long way to allowing researchers to return to their research. The suggestions offered in the preceding paragraphs are summarized as follows:

- Design and implement consistency checking of key and general demographic variables.
- Construct a set of index or summary variables to functionally replace the flap variables.
- Refine and redefine, if necessary, the screener ADL/IADL questions versus the detailed ADL/IADL questions.
- Define unambiguous missing data values by systematically going through the skip logic to untangle ambiguous responses.
- Ensure consistency among the helper, household member, and children sections of sample persons records.
- Define strict intra-record consistency-checks for that subset of variables (or all variables) deemed crucial to research efforts.

# IMPUTATIONS AND THE CREATION OF DERIVED VARIABLES

## Background

Several users expressed concern at the extent of missing data for certain types of variables on the NLTCS files. As is true for most surveys, financial data are particularly difficult to collect because of imprecise knowledge by the respondent or proxy, or a general reluctance to share financial information. Imputation of financial data in a rigorous, systematic manner would be a great service to users. Many researchers do some imputation themselves, but lack the necessary resources and background information to produce the best possible estimates.

## Proposed Improvements

There are a number of areas where imputations could be carried out using various techniques. One such technique would use control numbers from other surveys and published sources as the basis for the estimates. Another technique, hot-decking, uses reported information from responding sample persons to serve as proxy information for persons not answering particular questions. Inclusion of any estimated values on the file must be thoroughly documented, and the imputed values must be easily identifiable. Users would then have the option of not using the estimated values if they did not agree with assumptions made, or procedures used, for the estimation.

Another area of discussion and consideration would be the addition of new variables to the file. The availability of a standard set of constructed variables would be welcomed by many in the NLTCS user community. Variables that have come to be commonly used among the research community, and that should be defined consistently for most research applications, would be included in this type of file enhancement. Basic variables like one or more indices of overall disability would be welcomed by users who seek guidance on their definition. Again, the component data would continue to be available for variations of definition.

# IMPROVED DOCUMENTATION

## Background

Users agree that there is a need for more complete, more usable documentation of both the 1982 and 1984 NLTCS files. The problem stems not from too little documentation, but rather from the organization and accessibility of what is available and the incompleteness of the commentary on certain topics.

Unfortunately, the current documentation consists of voluminous technical memoranda, codebooks, appendices, explanatory notes, etc., which have been produced as both primary documentation and in response to user's questions and needs. Technical memos between the Health Care Financing Administration (HCFA) and the Census Bureau give most of the details on sampling and weighting, but these memos are not summarized, organized, or indexed. Also, editing rules that apparently changed between the preparation of the 1982 and 1984 data are not documented at all.

## Proposed Improvements

As expressed at the User's Forum, concerns about documentation are quite varied. There is interest in having distributions and control totals on key variables, both weighted and unweighted, as part of the standard documentation. More graphic presentations would be beneficial. These could include flowcharts of the population bases for the 1982 and 1984 files, diagrams summarizing imputation procedures, and so forth. Summaries, tables of contents, and indices to help organize and use the documentation are needed. Users indicated that other models of good survey file documentation (like the Current Population Survey) should be considered for future or additional documentation of these data.

At the very least, a technical overview with a detailed table of contents and cross-referenced index is needed. This would identify for the user what is available and where to go to find it. An overview of the entire survey, including methodologies, instruments, and procedures, as well as differences between the 1982 and 1984 data periods, would further enhance the documentation.

Looking at particular subjects, a detailed section describing the weights, how they were derived, how to use them, and how to make adjustments for the nonresponse in the samples should be included in any documentation package. Adjusted, estimated, and constructed variables would be identified, and the adjustment procedures would be explained.

Ideally, if funds were available, the documentation would be totally "redone." It would be restructured to conform to accepted documentation standards, including what

is best about the documentation of other public use files. Elements would include data dictionaries with frequency distributions for all variables, with cautions and limitations on their use. Variables with values estimated by imputation procedures would be clearly identified, and the procedures would be adequately explained. Constructed variables would be similarly described.

Users expressed concern about the upcoming 1988/89 file and its documentation. In addition to the detailed documentation that accompanies a new public use data file, users want to have an adequate understanding of what changes have occurred since the original design of the survey. They also want assurance that these changes will be reflected in the distributed file documentation.



# TECHNICAL ASSISTANCE

## Background

Many users of the NLTCS files complain that no systematic mechanism is available for getting questions answered in a timely manner. They also noted that there is a need for centralized and systematic distribution of summaries of important technical notes so that users can judge whether further detailed information is needed. A technical assistance unit to support the user community would help fill this void.

A number of users had given thought to what elements they would like to see in such a unit. Some felt that a quarterly newsletter that detailed problems encountered and possible solutions, as well as the recent research efforts of other users of the data would be very useful. Others felt that periodic seminars on key topics relating to their research efforts would be beneficial. Generally, the technical assistance component would act as a clearinghouse for questions, directing users to appropriate people or documentation for information.

## Proposed Improvements

It is important to design a technical assistance unit for the users of the NLTCS files. It should be easy to use, well advertised, and available on a regular basis. It should provide users with timely answers, and could serve as a forum for interaction among the users. Such a unit should coordinate distribution of file documentation and updates, a regular newsletter, and other information of general user interest. Telephone technical support, with perhaps an electronic mail/bulletin board system for obtaining information and asking questions could be designed. Other technical assistance units should be analyzed in designing the unit.

To obtain a printed copy of this report, send the full report title and your mailing information to:

U.S. Department of Health and Human Services  
Office of Disability, Aging and Long-Term Care Policy  
Room 424E, H.H. Humphrey Building  
200 Independence Avenue, S.W.  
Washington, D.C. 20201  
FAX: 202-401-7733  
Email: [webmaster.DALTCP@hhs.gov](mailto:webmaster.DALTCP@hhs.gov)

---

---

**RETURN TO:**

Office of Disability, Aging and Long-Term Care Policy (DALTCP) Home  
[\[http://aspe.hhs.gov/office\\_specific/daltcp.cfm\]](http://aspe.hhs.gov/office_specific/daltcp.cfm)

Assistant Secretary for Planning and Evaluation (ASPE) Home  
[\[http://aspe.hhs.gov\]](http://aspe.hhs.gov)

U.S. Department of Health and Human Services Home  
[\[http://www.hhs.gov\]](http://www.hhs.gov)