

**Tracking Software for Longitudinal Studies
at the National Center for Health Statistics**

Sandra T. Rothwell, National Center for Health Statistics, Hyattsville, MD
Aileen D. Bennett, The Orkand Corporation, Silver Spring, MD

ABSTRACT

The mission of the National Center for Health Statistics (NCHS) is to provide statistical information that will guide actions and policies to improve the health of the American people. To support this mission NCHS began to field an increasing number of studies which followed the health status of individuals over time. These longitudinal studies can continue for many years and require the constant update of vital status data and information necessary to re-contact living survey subjects in person, by mail, or over the telephone. As the number of followup studies grew, it became clear that this tracing activity was best supported by a single system which would be applicable agency-wide.

This system, written using SAS/AF™ software, currently provides support for nine of the most often used methods for collecting changes in addresses, telephone numbers, vital status and dates of death. Information may be obtained from the U.S. Postal Service, the telephone companies, selected Government and private databases, and through direct contact with the subjects themselves. The system is designed to support any NCHS survey, to meet strict confidentiality requirements, to be almost infinitely expandable, and to be compatible across mainframes, micro-computers and LAN servers.

INTRODUCTION

Over the past decade, a number of studies within the National Center for Health Statistics (NCHS) have added longitudinal components. For all such studies it is necessary to "track" the study's subjects and proxy respondents over time (20 years in one case). The purpose of tracking is to maintain current address and telephone information and to determine accurate vital status data. A high followup rate is essential in order to obtain reliable results. While all tracking and tracing activities have much in common, leaving each survey to plan and execute its own tracking effort would result in a lack of consistency in tracking methodology from survey to survey. In addition, each longitudinal data collection effort

would incur the cost of developing tracking strategies and supporting software.

Hence, the decision was made to develop a software system which would facilitate tracking and would be applicable to any survey conducted by NCHS. The system would automate the clerical functions of a number of tracking activities frequently used for NCHS surveys. These functions include creating both paper request forms and tape or diskette files; capturing, reviewing, and processing the information returned from the tracking source; and updating a tracking database.

Not only had the past approach to tracking been costly and inconsistent, it was also labor intensive and often required custom programming at almost every step of the way. Each of these programming efforts provided an opportunity to introduce errors and often did. It was not unusual for statistical staff to review volumes of computer listings for the sole purpose of assuring the accuracy of a study subject's vital status. A potentially more serious issue was the possibility of incomplete records from tracking activities. The system was to be designed to address all of these issues.

DESIGN CRITERIA

A steering committee including representatives from several NCHS data divisions determines design criteria. There were five major design goals set by this committee. In order to provide the "right size" system for any size cohort, the system was to be portable - eventually operating on the agency mainframe, the local area network server, and stand alone micro-computers. The system was to be easy for clerical staff to use, providing entry time data validation. It was to be secure, protecting sensitive information such as names, addresses, telephone numbers, Social Security Numbers, and Medicare claim numbers. The system was to provide a complete audit trail allowing the user to trace all the changes for any person in the database. And finally, in order to reduce the number of careless errors, the system was designed to update off-line except in tightly

defined circumstances.

In order to create the system, we needed developmental software that would support menu and screen development as well as be easy to modify. The software was required to support database functionality and be compatible across Center's for Disease Control and Prevention (CDC) mainframes and micro-computers. At the time work started in 1989, the choice of database systems that ran on all our mainframe and micro-computers was not large. In order to keep the project affordable, we chose to use SAS/AF software. On the whole, this has worked well. There have been no instances where design requirements could not be implemented, in some fashion, with this product and the choice is currently facilitating the move to micro-computers.

SYSTEM FUNCTIONS

The current system runs on the NCHS IBM™ 3081 mainframe computer. It prepares forms, mailing labels, tapes and diskettes; controls data entry; checks for consistent information (for example, persons are not allowed to have a death date prior to their last live interview); and handles all updates including the creation of necessary backup files. The system provides a variety of reports. In order to be sure that tracking details can be reconstructed for questionable cases, a complete audit file of all updates to the database is maintained.

Each tracking activity implemented is a separate system module with a separate set of programs except where shared functions, such as security control, use common program code. For each survey, there is a separate database all, of which are identically structured but secured separately from each other. Use of any database by those outside a particular survey staff is denied.

The following figures are simulations of screens that are typical of the system. The system main menu shown in Figure 1 lists the identifiers of surveys that are using or plan to use the system. Current databases include the NHANES I Epidemiologic Followup Study (NHEFS), the Third National Health and Nutrition Examination Survey (NHANES III), and a family of National Health Interview Survey Mortality Surveillance projects (NHIS). The test file is used for demonstrations and for system testing. A future Longitudinal Study on Aging is planned.

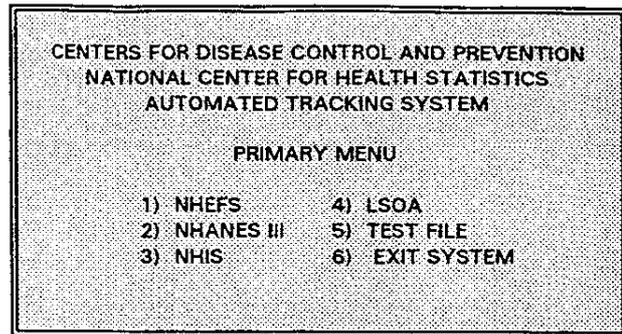


Figure 1. Survey selection menu

Once you have selected a survey, a tracking activity module is selected. Figure 2 shows the modules that have been completed thus far.

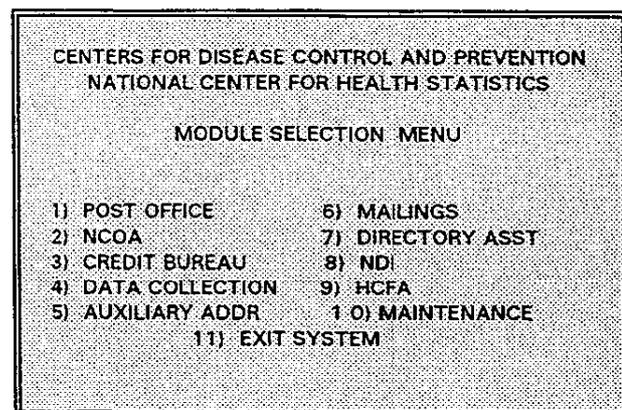


Figure 2. Module selection

Post Office checks obtain corrected address information directly from local postmasters and are paper driven while the NCOA accesses the postal service's National Change of Address Registry database and uses files on diskette rather than paper. Credit bureau searches, another paper driven application, provide marginally reliable information on addresses and vital status. The data collection module provides the interface between data collected from survey respondents and the tracking master files. The module is designed to handle data collected using a variety of methods including mail questionnaires and personal and telephone interviews. A historical file of auxiliary addresses is maintained and can be used as input to other modules for further address verification. General mailings, such as for birthday greetings, can be facilitated using the system. Directory assistance is a direct data entry module where information is presented in area code sequence for collection from the

telephone companies. The National Death Index (NDI) module facilitates matching to a database containing information on all deaths in the United States since 1979. The module is tape driven and a special application procedure is required to use the database. Also under special circumstances and with special permission, the Medicare Enrollment file can be searched using the HCFA module. Finally, the maintenance module allows you to do many things that are not really part of another module, such as browsing the files or building a flat file for analytical purposes. Subjects occasionally contact NCHS with new address information. The maintenance module allows you to correct a record without invoking any of the tracking activity modules. As illustrated, the system supports tracking activities using a variety of databases and several modes of information collection and update including paper, tape, diskette and direct data entry.

Figure 3 is a typical module initiation screen. For any module there are a variety of ways to select the persons for whom you want to invoke the activity.

```

CENTERS FOR DISEASE CONTROL AND PREVENTION
NATIONAL CENTER FOR HEALTH STATISTICS

INITIATE NDI SEARCH

1) SUBJECTS BY AGE
2) SELECTED ID
3) OTHER MODULE STATUS
4) FILE OF IDS
5) MAKE TAPE
6) EXIT SYSTEM
  
```

Figure 3. NDI initiation choices

For an NDI search we allow you to build a submission file of subjects within a given age range, usually above a specific age. You may type in up to 50 selected tracking identifiers, an option useful for working on a few hard to locate cases. You can build an NDI submission including persons for whom the results of other modules indicate that they may be deceased, for example, anyone whom the postmaster indicated was deceased. Additionally you can also submit an external file of tracking identifiers, thus allowing interface

between the tracking system and other software systems such as Computer Automated Telephone Interviews. Persons can be selected using any or all of these options. The last option, "MAKE TAPE", will build a list of all the selected identifiers, un-duplicate the list, and create a correctly formatted NDI submission tape. Other tracking modules have different initiation screens as applicable to the activity.

```

ID CODE: _____
TYPE: _____
NAME: _____
STATUS:  **UPDATED STATUS CODE**
COMMENT: _____

OLD ADDRESS:
STREET: _____
CITY: _____ STATE: _____ ZIP: _____

NEW ADDRESS:
STREET: _____
CITY: _____ STATE: _____ ZIP: _____
SPECIAL PLACE: _____
  
```

Figure 4. Online update screen

An example update screen is shown in Figure 4. For changes of address indicated by the post office, you can view the old address (to be sure the update is for the correct person) and enter the new address. Another option will bring up this screen with the old address already copied to the new address field thus facilitating minor changes.

In the future, modules will be added to handle matches to the Social Security mortality file and requests to state agencies such as vital registrars, motor vehicle and marriage bureaus, voter rolls, driver's license listings and so forth.

SPECIAL CONSIDERATIONS

Security is of primary concern in the system. The Government provides a guarantee of confidentiality to all participants in its health surveys. It is imperative that address and telephone information and identifying numbers such as Social Security and Medicare numbers are protected from disclosure. All system files are protected using the Resource Access Control Facility™. Entry into each database and module of the system is controlled with internal authorization tables that only tracking

supervisors can modify. Even the system programmers cannot see data for the application surveys.

Archival information is also of considerable importance. The system maintains a record of all updates to the database. These updates can be used for studies of effectiveness, as a trail to follow when questionable information is found in the master files, and as the backup of last resort. They are often used for tracing of persons who become difficult to locate. In addition, their existence makes it possible to "undo" updates for specific cases without resorting to the use of backup files. In a case where tracking activities may have picked up on the wrong "John Smith" for example, the current system will "undo" previous changes and start tracking as if from an earlier point.

ADVANTAGES GAINED USING SAS™ SYSTEM

The SAS software has allowed us, one way or another, to implement almost every function required. The use of SAS software has given us the ability to move parts of the system from a mainframe platform to a personal computer with minimal modification. While some changes to security measures were required and occasional compilation errors have been encountered, overall conversion efforts have been less difficult than anticipated.

Several of the options available in the SAS Screen Control Language have proved extremely useful in this application. Functions that support looking for character strings that "sound alike" allow searching through the database for names that have been obtained over the telephone when exact spellings are often not known. These functions also support more complete elimination of duplicates. Variable selection lists were employed in the maintenance module to ease the job of creating flat files. And much of the data verification, such as validating ZIP codes and telephone area codes, is made possible by using SAS data sets for table lookup right in the data entry screens.

The SAS System has provided very useful advantages for documentation. The SAS/AF software makes it easy to download the system's screen displays into WordPerfect™ files to be included in the user manual. Likewise, selected text from the user manual has been uploaded into the "help" screens. This capability has not only decreases the effort required for documentation, but

helps insure consistency between the system itself and the external documentation.

Finally, because the tracking databases are actually SAS data sets, it is easy for staff from the participating surveys to build them and to use them directly in analytical projects or for reports that are not provided directly from the system.

DISADVANTAGES OF USING THE SAS SYSTEM

Disadvantages have been few. There are instances where the implementation staff would have liked greater control over the screen displays, especially during the execution of submit blocks. For certain paths in the system, it would be desirable to be able to select a conditional return path from a screen.

FUTURE DIRECTIONS

NCHS is currently experimenting with a micro-computer version to be used for small cohorts. It would be ideal to put the tracking system on a micro-computer, load a tracking database and turn the data, hardware and system over to the agency responsible for tracking. The Government would then have excellent control over this important activity. Likewise, this kind of "turnkey" tracking system could be set up in any unit of NCHS that needed to track a reasonably small number of subjects without tapping the mainframe computer. The survey selection, module selection, and the maintenance functions are complete. An NCOA module is under development. The micro-computer presents special problems, not the least of which is security. In addition, the time required to complete some functions can be long enough that the user needs to plan ahead carefully. Checking all existing addresses against the entire post office zip code database, for example, requires a minimum of 45 minutes on 486 style computer.

PRODUCTION HISTORY

Table 1 displays results of three post office checks conducted with and without the tracking system. A 1989 effort cost \$25,000 and took 7 months. The first version of the tracking system processed not quite twice the number of postal requests in less time and for about 2/3 the cost. While the cost for the latest postal check is included in a larger data collection contract and cannot be determined as yet, a tracking system programmed in Version 6.06 of the SAS

software with improved interfaces, was used to complete the job in about 1/4 the time of the 1989 check. The system has also been used for more than half a dozen submissions to the NDI with great success.

Table 1. Production History

System and date	Number of requests	Time to complete	Cost
1989 non-govt.	12,000	7-8 months	\$25,000
1990 NCHS 5.18	20,000	4-5 months	\$17,000
1991 NCHS 6.06	12,000	7 weeks	Unknown

statistician/systems analyst with programming support from both Government and non-government personnel. The individual databases are built and maintained by technical staff of the Division responsible for the survey. This three year project has truly been a collaborative effort which has paid off in an easy to use, accurate, secure system with excellent potential for future use and development.

SAS and SAS/AF are registered trademarks or trademarks of SAS Institute, Inc. in the USA and other countries. IBM is a registered trademark or trademark of International Business Machines Corporation. TM indicates USA registration. Other product names are registered trademarks or trademarks of their respective companies.

SUMMARY

The SAS/AF software was chosen for the system because it provided database functions for the tracking information; because it was compatible across all the hardware products that could be used for tracking in the foreseeable future; because it provided a flexible developmental and maintenance base; and because it was affordable. In addition we have discovered that the SAS software offers a structure that allows approved survey staff the flexibility to load their own databases and to produce special reports or subsets of the data outside of the tracking system itself. All new users are pleased that they do not have to master new software skills to support their tracking databases. The developers have been able to implement all capabilities requested at some level. The SAS/AF software offers impressive user interfaces and aids for documentation. A great deal of data editing is possible and the current tracking databases are practically error free.

This system is unique, not only in its task, but in its developmental and staffing configuration. Major design decisions are made by a steering committee which includes one member from each of the data collection units responsible for the followup surveys. Technical design is directed by a