AN AUTOMATED SYSTEM FOR SURVEY OPERATIONS AND DATA COLLECTION IN A COMPLEX HEALTH SURVEY

Meena Khare and Carolyn L. Petty-Martin, National Center for Health Statistics
Meena Khare, 6525 Belcrest Road, # 915, Hyattsville, MD 20782

KEY WORDS: NHANES, sample management, survey information system

1. Introduction

With the ever increasing need to expedite dissemination of public health data of high quality, the automation of survey operations and data collection in large complex health surveys becomes essential. Although it is a major challenge for system designers, database managers, and survey practitioners to develop complicated computer assisted survey information collection (CASIC) systems, the rapidly growing modern information technology (IT) has made it feasible to automate major components of complex surveys. Appropriate selection of the software and methodology are critical to the design of an efficient CASIC system. In the 1990s, a number of survey organizations began to investigate the design and development of user-friendly CASIC systems to potentially enhance the survey data quality, reduce total survey cost, and facilitate timely dissemination of the survey data and results (1-7). A historical review of the CASIC along with its' advantages and disadvantages was presented by Nicholls and de Leeuw (5). CASIC systems can be further enhanced by including automation to integrate the sampling methodology (4) with other survey operations and data collection systems. The purpose of this paper is to describe an automated Integrated Survey Information System (ISIS) for the National Health and Nutrition Examination Survey (NHANES) including a discussion of the advantages and limitations of this system.

The National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), conducts a number of health surveys to collect, process, and disseminate public health data of high quality. A series of NHANES Surveys, conducted by the NCHS, has been collecting health and nutrition data for the U.S. population since the 1960s. For NHANES, the survey operations, data collection, and quality control methodology have evolved from the paper and pencil interview (PAPI) and manual data-keying technology used in the 1960s to highly complex automated data collection systems in the 1990s. For the NHANES III (1988-1994), the contractor, Westat, Inc., developed a customized semi-automated and partially integrated CASIC system (1,4). The collection of household interview data initially started with PAPI and later changed to a CAPI system during the last three years of the survey (1). The examination data were collected using a semi-automated data collection system, developed with the *ORACLE* database management system for personal computers (PCs) to manually enter the interview and measurement data. With the evolution in high speed micro computers, for the first time, a fully automated Integrated Survey Information System is in place for the current on-going NHANES. The ISIS was designed and developed by the NHANES team at the Westat, Inc., in collaboration with the NCHS professional staff. Key features of the ISIS include automated procedures to implement the survey design and sampling methodology, manage survey operations, collect confidential health and nutrition interview data, and record measurements from the medical examination. Recently, Binzer and Berman (7) presented technical details of the ISIS at the National Conference on Health Statistics.

2. Overview of NHANES

Until 1999, NHANES was a periodic survey. In 1999, NHANES became a continuous annual survey with a nationally representative sample of 6,500 persons selected at 15 locations. It is designed to provide national estimates of health and nutritional status of the civilian noninstitutionalized population of the U. S. NHANES is based on a complex, multistage area probability sample design including an oversample of selected minority subpopulations. Of the selected sample, 5,000 persons are expected to be interviewed in their home and then examined in the especially equipped Mobile Examination Centers (MECs). Medical history, and socio-demographic information are collected through personal household interviews, and more than 20 standardized medical examinations or tests are conducted in the MECs to collect data on physical measurements, physiological tests, diet and nutrition, and biochemical measurements from blood and urine specimens.

In addition to the home office staff, there are three MECs, one advance arrangement team, three field office teams, three interviewer teams, and two examination staff teams to administer survey operations and data collection in the current NHANES. At any point in time, two sets of teams collect interview and examination data at two locations while the third set

travels to the next location for set up. Starting in 1999, except for the listing of households (currently a manual hardcopy process), all of the survey activities are coordinated through the ISIS. After selection of the primary sampling units (PSUs) and survey locations, the advance arrangement team collects the local area information, contacts the local health officials, and makes arrangement to set up the MECs and field offices (FOs) at each location. Two teams simultaneously screen households, collect household interview and MEC examination data at two separate locations. The sample management system implements the sampling methodology and releases households for screening. The major components of the field office operations include, but are not limited to, monitoring of the manual listing of dwelling units (DU), loading released households to the Laptops, automated screening of eligible households and selection of one or more persons from each eligible household based on their demographic characteristics, administration of household interviews using a CAPI, making appointments for medical examinations, making transportation arrangements for appointed persons, and managing petty cash, employee/visitor, and start-up/close-down systems. At the MEC, a coordinator, a MEC manager, a data manager, and a group of health technicians conduct the scheduling and examination operations. Each person gets a set of medical examinations or tests, an instant "report of findings" with selected results, and a certificate and remuneration for their participation.

In the previous NHANES, household data were collected using a PAPI and examination data were collected on paper and/or entered in a database using a stand-alone PC with some automation (Table 1). During the NHANES III (Phase 2, 1991-1994), the mode of household interview data collection was changed from a PAPI to a CAPI system using Laptop computers; the examination data were collected using semi-automated and partially integrated PC systems. These PC systems were connected to a centrally located operating system in the MEC. For the NHANES 1999, collaborative research for a fully automated and integrated data collection system was initiated in 1996. The implementation of ISIS in 1999 has significantly improved the efficiency of the data collection and timely access to the usable data within 24 hours by the home office staff at the NCHS and Westat. Also, data quality has been enhanced by using standardized response values or categories and built-in quality control procedures. To further improve the efficiency of the ISIS, it was divided into a number of components. The next two sections provide an overview of the ISIS and some of its major components.

3. Overview of the ISIS

ISIS was designed and developed using the modern information technology (IT) available for the distributed database environment across *Wide Area Network* (WAN) to provide access to the survey data in near real time. The design and contents of the ISIS are the product of an extensive collaborative team effort of expert IT professionals, system designers, database managers, statisticians, survey practitioners, biomedical engineers, clinicians, telecommunication professionals, sociologists, and other field operations and subject-matter experts. The ISIS for the NHANES consists of a complex telecommunication architecture that links the field office, the MEC, the home offices at Westat and NCHS, and other subcontractors (Figure 1). The client-server applications are designed to efficiently manage survey operations and data collection, to provide access to useable data within 24 hours, and to expedite data dissemination.

ISIS uses the latest state-of-the-art hardware platforms with high speed PCs to collect data in the MECs. For the screeners and household CAPI interviews, hand-held pen-based Laptop computers are used. In the MEC, PCs are directly linked to the centrally located server and to complicated biomedical equipment to facilitate direct data collection and data transfer to the database. All *Window NT* based PCS in the MEC also have e-mail capability to send and receive electronic messages from the coordinator. In addition, all hardcopy procedure manuals used in the past have been replaced with an online catalog of manuals on the PCs with a Window based search option. The field office and MEC systems also have Intranet and Internet applications for sharing reports and other information. Finally, the ISIS uses an excellent data security, recovery, and backup system to comply with the confidentiality rules established by CDC.

The design research for the ISIS started in 1996 and two field tests were conducted prior to the final dress rehearsal. The Alpha and Beta tests were done in 1998. The final dress rehearsal training, testing of the ISIS, and data collection were completed in February 1999. The full successful implementation of the ISIS and the data collection for the NHANES started in March 1999.

4. Components of the ISIS

There are several customized components of the ISIS including, but not limited to, sample management, field office management, interview data collection using CAPI, examination data collection in the MEC, and data

processing and transfer using state-of-the-art hardware platforms. These major components are described in the following sections.

4.1 Sample Management

The sample management system, designed primarily in consultation with statisticians and survey practitioners, has components to select and release a sample of the listed dwelling units (DU), screen households for eligibility, and select one or more persons from each eligible household. This system assigns final disposition to each selected case, assigns deselection or desampling to selected cases, monitors the sampling process, and reports the sample yield. It also assigns the basic sampling weights to all selected persons based on their demographic characteristics and probability of selection. For the security of released cases and to control proper sample selection, only selected senior statisticians at the home office have access to this system. All selection and eligibility criteria for screening households and selecting persons are programmed in the automated system with strict quality control measures. For example, after obtaining the demographic information for all household members, the screener CAPI system selects one or more persons based on the self-reported demographic characteristics and displays a message for the interviewer on the PC screen showing which members have been selected. The selection criteria consisting of the pre-defined sampling rates for the target domains are programmed in the screener CAPI.

4.2 Field Office Management

Each of the three field offices (FO) in the NHANES is comprised of a Field Operations Coordinator/ Field Manager, an Office Manager, two Assistant Office Managers, a FO/MEC Data Manager and a number of interviewers (8). The field office management system consists of a number of sub-systems for the Advance Arrangements and community outreach activities, Appointment Management, Case Management, Stand Management, Petty Cash Management, Employee and Visitor Management, Data Management, Editing and Scanning the FO forms, Report Management, and modules for the Validation and the ISIS Security.

Specifically, the Advance Arrangement System collects and maintains the required information to help set up the FO and the MEC at each survey location and provides local community outreach information. The Appointment Management System contains a calender to help FO/MEC managers and interviewers to set up examination appointments, and a subsystem to enter information on participants' special needs for the transportation to the MEC, the wheel chair assistance, or a language interpreter. The Case Management System provides information on every case assigned to the interviewers by the Field Manager. The Stand Management System is used to implement the start-up and close-down procedures of the FO operations at each survey location. The Petty Cash Management System is used to handle all local expenditures and cash transactions between the FO and the MEC and to reconcile the balance on a weekly basis. The Employee and Visitor System maintains a log with contact information on all NHANES field employees and the FO/MEC visitors. The Data Management System, and the Validation module maintain the database with the administrative data, and the original and re-interview data. The Report Management System processes the administrative data to produce a number of routine FO reports in order to monitor survey response rates and other quality control information. The ISIS Security module contains the passwords and identification numbers of all field employees, and accessed by the Data Manager to monitor security violations, if any.

4.3 Household Interview

The CAPI system for the interview includes a screener to select eligible persons from the selected households and to administer household questionnaires with built-in quality control procedures (9). The CAPI contains a number of built-in menus and pull-down lists to select standardized responses in order to reduce data entry error. For example, a pull-down list with valid coded options is used to select gender and race-ethnicity in the screener and an on-line database with alphabetical search options is used to select correctly the brand names, manufacturer names, and potency of the medicines, supplements and vitamins reported in the household interviews. Also, a 10% validation module is designed to collect quality control re-interview data from the interviewed persons; 50% of the validation data is collected using a telephone interview and another 50% using an in-person interview. Validations are done by the Assistant Office Managers. Finally, responses to the selected self-reported health history questions are directly relayed to the MEC system in order to exclude persons from the selected examinations for medical safety. For example, in addition to the responses to the safety questions asked in the MEC, persons who report one or more heart conditions related to the cardiovascular (CV) disease during the household interview, are excluded from the treadmill test of the CV fitness examination in the MEC.

4.4 MEC Examinations

Each MEC team for the data collection consists of a MEC Manager, a MEC Coordinator, a local helper, a Physician, a Dentist, two Dietary Interviewers, two MEC Interviewers, five Health Technicians, one Phlebotomist, three Laboratory Technicians, and two TB (tuberculosis) Readers. A password-protected automated Coordinator System in the MEC controls all scheduling, examination, "Report of Findings", and remuneration activities (10). It assigns participants and examiners to the examination rooms based on the participants' age and is integrated with more than 20 examination components to monitor the progress of each examination in real-time. The MEC system is designed to operate in the *Windows NT* environment with the utmost security to collect data from various examinations and tests, and maintain confidentiality. The collected examination data for each participant are automatically and instantaneously transferred to a *Sybase* database residing in the main server and a copy is dynamically transmitted to the home office at Westat.

In addition to the coordinator system, the automated MEC data collection system consists of an Appointment Management System to access the participants' appointment information, a Stand Management System to start-up and close-down MECs, and a data collection system with built-in quality control procedures for each examination component. In the MEC, a bar code reader is installed in every room to read the bar codes on the bracelet and automatically enter the participant's identification number in the database; an AUDIO-CASI system (where a participant hears a question using a headphone and enters the response by touching the monitor screen) is used with a *KIOSK* screen to collect highly sensitive interview data (e.g., response to questions on sexual behavior and drug use) in private rooms; a shipping and tracking system is installed for laboratory specimens and administrative records. Again, in the laboratory a bar code reader reads the identification numbers from the vials and then transfers to the shipping system. The PCs in each examination room are password protected, have e-mail capability, and directly integrated with complicated biomedical equipment to facilitate accelerated data collection and data transfer to the database. Because data are transferred in real-time from the examinations, the "Report of Findings" system immediately prints the reports with results from selected examinations so that they can be given to the participants before they leave. Furthermore, the home office has an integrated quality control system to periodically produce response rate reports and to monitor data quality at each survey location.

5. Advantages of the ISIS

The automated data collection using ISIS based on modern information technology has significantly improved the efficiency and quality of the NHANES. In particular, ISIS has facilitated more rapid data collection, and processing, thus resulting in more timely release of the survey data. By comparison, in the NHANES III, it took several months to process the data and prepare the "Report of Findings". In addition, by using standard response categories or values, and pull-down menus and lists to enter responses, the magnitude of data entry errors and the amount of post-survey processing are expected to be reduced substantially. Also, by using bar code bracelets and readers, ISIS has significantly reduced the error in entering the identification numbers. In the NHANES III, where sample identification numbers were manually entered in the database, unexpected delays resulted in the data release due to errors in the identification numbers. Additionally, automation and integration of the Audio-CASI system for the sensitive questions are expected to improve the quality and confidentiality of the responses. Furthermore, the automation and integration of the ISIS with the laboratory system is expected to improve the shipping, and receipt and control of the laboratory specimens.

With implementation of an improved coordinator system with e-mail capability in the MEC, the coordinator is able to follow the exact location of the examination staff and participants without leaving the reception area. Efficiency is also gained with the ability to send e-mails to the staff to move participants from one room to another without bringing them back to the waiting room. Moreover, the improved scheduling, built-in quality control procedures, improved data processing, and automated data transfer have significantly improved the efficiency of the MEC operations and substantially eliminated the data entry error. Automated data collection has also eliminated the cost of post-survey data keying. Overall, ISIS has significantly enhanced the efficiency of the sample management, survey operations, and the quality of the final data. Furthermore, since ISIS was designed to integrate various independent modules, new modules can be added and/or existing modules can be eliminated. It is expected that in the long run the total survey quality will be improved significantly and at the same time total survey cost will reduce.

6. Limitations of the ISIS

Although the advantages of using ISIS in the

NHANES are significant, there are also some limitations to this system. Because ISIS is based on a complex architecture and a network of numerous subsystems, it is difficult to make quick *ad hoc* minor changes. All changes or edits have to be carefully planned, applied in batches, and tested for the performance across all linked modules before implementation in the field. Most of the modules or subsystems are designed to work independently, however, data are shared across the various modules, therefore, changes to any single module or data item could significantly impact the other modules or data. ISIS also has implications for the operational costs of the survey. Most surveys plan their annual budget to be equally distributed over time and a slightly higher start-up cost for the first year. With the implementation of a system like ISIS, the start-up cost should be budgeted at a much higher rate to cover the cost of designing, developing, and testing the ISIS. Moreover, the costs for the subsequent years should include additional costs of maintaining, updating, and documenting the ISIS and its components. Periodically, ISIS will also require upgrade of its' software. A CASIC system like ISIS depends considerably on the rapid changes in the modern information technology. Because of the automation, hardcopy records are totally eliminated from the survey resulting in no hardcopy backup to recover missing data or to verify errors in the collected data. Although, extensive efforts have been made to reduce down-time for the ISIS with state-of-the-art data replication, recovery and backup systems, the possibility for missing data exists. For example, if the main operating system crashes or goes down, there is no alternative to recover the missing data. Also, due to the complexity of the design, the FO and MEC systems require frequent intervention by the data managers and the system designers resulting in the increased need for technical staff at the FO and at the home office. Finally, the field staff including managers, interviewers and examiners require intense initial training, periodic retraining, and a longer practice time. Although ISIS is designed with menu driven user-friendly modules, a slower initial learning curve exists for the staff who are not familiar with computers and complex biomedical equipment.

7. Summary

There are significant challenges to the design and development of systems like ISIS for large complex surveys. Nevertheless, substantial gains can be achieved. The ISIS has significantly improved the efficiency of the sampling procedures, sample management, and survey operations. In addition, the automation and integration of the ISIS provided survey data of high quality by significantly reducing the data entry and post-survey data processing errors. Moreover, ISIS has completely eliminated the use of hardcopy forms and the associated cost for data keying. The near real-time access to the data facilitated timely data release for quality control analyses. In summary, ISIS for the NHANES has shown that it is feasible to design and develop such a complicated CASIC system for large complex health surveys to improve efficiency of the survey operations, and facilitate timely dissemination of the public health data with enhanced quality.

References

1. Sperry S, Rhoads M, Shepherd J, and Smith J. (1994), "Strategies for implementing CASIC at Westat", Proceedings of the 1994 Annual Research Conference and CASIC Technologies Interchange, Bureau of the Census, Suitland, MD, pp 647-661
2. Matchett, DS, Creighton, KP, and Landman, CR. (1994), "Building Integrated Systems of CASIC Technologies at the Bureau of the Census", Proceedings of the 1994 Annual Research Conference and CASIC Technologies Interchange, Bureau of the Census, Suitland, MD, pp 573-595
3. Whelen JL, Karlsen RB, and Yost, PA. (1995), Proceedings of the 1995 Annual Research Conference, Bureau of the Census, Suitland, MD, pp 593-603
4. Waksberg J, and Mohadjer, L. (1992), "Automation of Within-Household Sampling", 1991 Proceedings of the Section on Survey Research Methods, ASA, Alexandria, VA, pp 350-355
5. Nicholls, WL. and de Leeuw, ED. (1996), "Factors in Acceptance of Computer-Assisted Interviewing Methods: A Conceptual and Historical Review", 1996 Proceedings of the Section on Survey Research Methods, ASA, Alexandria, VA, pp 758-763
6. DHHS. (1996), NHANES III Reference Manuals and reports: Plan and Operations of the Third National Health and Nutrition Examination Survey, 1988-1994, National Center for Health Statistics, Hyattesville, MD
7. Binzer, GW and Berman, LE. (1999), "The National Health and Nutrition Examination Survey Integrated Survey Information System", presented at the National Conference on Health Statistics, sponsored by the National Center for Health Statistics, Washington, D.C.
8. NHANES. (1999), National Health and Nutrition

Examination Survey: Field Office Operations Manual, National Center for Health Statistics, Hyattesville, MD, internal document,
9. NHANES. (1999), National Health and Nutrition Examinaton Survey: Interviewer Procedures
Manual, National Center for Health Statistics, Hyattesville, MD, internal document
10. NHANES. (1999), National Health and Nutrition Examination Survey: Coordinator Procedures Manual, National Center for Health Statistics, Hyattesville, MD, internal document.

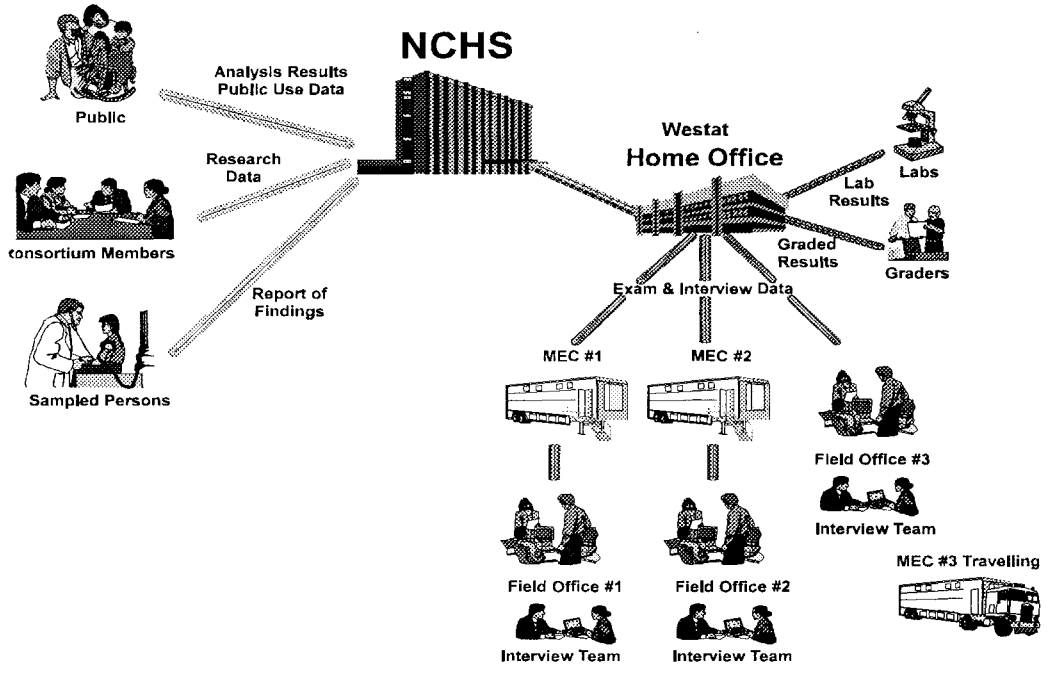
Appendix

Table 1: Automation in the National Health and Nutrition Examination Surveys

Surveys	Dates	Age	Sample Size	*Automation*
NHES I	1959-62	18-79	7,710	*none*
NHESII	1963-65	6-11	7,417	*none*
NHES III	1966-70	12-17	7,514	*none*
NHANES I	1971-75	1-74	28,043	*none*
NHANES II	1976-80	6mos-74	27,801	*none*
HHANES	1982-84	6mos-74	15,105	*some*
NHANES III	1988-94	2mos+	39,695	*moderate*
NHANES 99	1999+	All ages	6,500*	*ISIS*

* Estimated annual sample size covering 15 locations

Figure 1: The NHANES 1999 Network



Source: Westat Inc.