

**MONITORING SURVEY FIELD OPERATIONS:
THE NATIONAL MEDICAL CARE EXPENDITURE SURVEY AUTOMATED CONTROL SYSTEM**

Mary B. Woodside, Barbara A. Moser, and Linda L. Quinn, Research Triangle Institute

I. INTRODUCTION

A. Background and Objectives

The National Center for Health Services Research and National Center for Health Statistics are jointly sponsoring the National Medical Care Expenditure Survey (NMCES). Information collected in this survey will be used to address current health care policy issues. For example, a current important issue is that of National Health Insurance. The NMCES data are very extensive and complete enough to examine the costs and benefits of the various alternative National Health Insurance plans being proposed. The full length of the study, from its inception in July 1976, will be three and one-quarter years.

Although the complete project consists of three distinct surveys, the emphasis in this paper is on the household survey which will be completed in July 1978. In addition to the household survey, record check surveys are currently being conducted of providers of medical care, such as doctors and hospitals, etc., and of insurance firms and employers who provide coverage for medical expenses. These data cover the same reporting period as the household data and will be used to determine the accuracy of information reported from households and as a major component of the overall analysis effort. The household survey, which will collect 1977 medical care data for over 13,000 households, is a panel design including 6 interviews. The provider and insurance surveys, which will be conducted in the last half of 1978, are not panel-designed surveys.

Starting with the first field period in early January 1977, every selected household was interviewed 5 times over a period of 16 months. After the fifth wave of interviewing there was one last follow-up telephone interview. Each wave of interviewing, starting with Wave 2, was regarded as a follow-up to the previous interview. Shortly before each follow-up interview, the respondent received in the mail a computer-generated summary of the utilization and expenditure information collected in the preceding interview, and was asked to review the information with other household members for accuracy and completeness. Hence, what were considered critical data items which were missing on previous waves of the survey could eventually be obtained before the final data base is developed. The instrument also served as a mechanism to prompt the interviewer to recognize inconsistencies in responses across waves and to provide a mechanism for true corrections of the data which had been flagged as inconsistencies by the edit check.

B. The Project Team and Data Processing Facilities

The project team is made up of three corporations: the Research Triangle Institute (RTI) located in the Research Triangle Park in North Carolina as the prime contractor, the New York office of the National Opinion Research Center (NORC), and Abt Associates Inc. (AAI) of Cambridge, Massachusetts, as subcontractors. RTI's responsibilities are the overall project management, coordination with NORC in the field interviewing and training, provision of the data processing for managing field operations, converting the data to machine-readable form, and establishing the preliminary edited data base. NORC's primary responsibilities include instrument development, interviewer training and materials, and conduct

of half of the interviews. AAI developed a summary report update, production, and mailing system to produce the summaries.

The computer facilities used to support this project are the Triangle Universities Computation Center (TUCC), in the Research Triangle Park, and the SDL International computer facility in Ottawa, Canada. Both facilities are IBM 370/165 main frames with associated peripheral devices and system support.

C. Project Size and Variability

The volume and variability of data that required processing for any single wave of interviewing were significant factors in the design of computer software to support NCMES. The complete data base will be approximately one and a half to two billion characters of data covering survey data on approximately 39,000 respondents in over 13,000 reporting units* for the complete year of 1977.

The instruments for data collection included a core set asked each wave (the Questionnaire and Control Card), and other topic-specific questionnaires administered in one or more waves and various continuation pages or multiple forms as needed in each wave. The number of instruments completed in any particular interview varied with the number of persons in the reporting unit (each Questionnaire and Control Card could report on 6 persons or less) and the number of medical provider visits, dental visits or conditions reported by the family.

The respondents were tracked for follow-up surveys if they moved anywhere in the Continental United States. They were allowed to move from an original reporting unit to form a new reporting unit (i.e., husband and wife separate) or to move from an original reporting unit into another existing reporting unit (i.e., a college student moves back home for the summer). Hence, any one respondent could have data reported under more than one reporting unit identifier during the course of the study but never more than one within a given wave.

The study design required processing each wave of data and preparing it for return to the field in the subsequent wave. This meant that almost a quarter billion characters of data had to be converted to machine-readable form, edited and returned to the field for purposes of assignment of follow-up interviews, family-specific probes, and reconciliation. The time frame for completion of these tasks was less than 13 weeks per wave. There were two major computer-produced field instruments for each follow-up wave of interviewing. The Control Card provided assignment and locator information to the current field interviewer. It also contained data on each respondent in the reporting unit and identified any respondents who were no longer in the reporting unit.

The Summary was a computer-generated synopsis of the respondent's visits for medical care and medical expenses, including their associated costs and sources of payment as reported in all previous waves of the survey. The Summary was mailed back to both the respondent and the interviewer prior to each follow-up interview. This instrument was used by the interviewer to prompt the respondent for detailed visit information which may not have been available during a previous

essed through the Household Summary and Control Card generation, the Control System is prepared for the next round of data processing. It now has a participant record for each reporting unit in the survey and is expecting a response from each participant in each subsequent round. The Control System is designed to monitor the response from each participant even if the participant moves to a different geographic location, a different reporting unit, and/or a different reporting unit and then back into the original reporting unit.

Since the Control System must monitor data at both the participant and reporting unit level, appropriate linkages must be maintained. Current linkages which show how reporting units are configured or split in a particular round must be maintained in order to monitor reporting units that split and reunite in subsequent rounds. This linkage is maintained by automated directories in the Control System—the Current Reporting Unit Directory (CRUD), the Split Reporting Unit Directory (SRUD) and the reporting unit to reporting unit file (RURU). This capability of the Control System provides the field interviewers sufficient information to obtain complete reporting unit participation in a given round. It will also assure that duplicate participant information is not collected.

2. Maintain Current Participant Identification and Address Information

Since NMCES must follow respondents for a complete year of data collection, the Control System is designed to maintain the most up-to-date participant identification and address information. This information is required for mailing of Household Summaries and for assignment of interviews to field staff in each round. As indicated in Figure 1, the edited data are input to the control file update system which maintains this information.

3. Monitoring the Data Processing of Each Participant

As discussed in the previous section of this report, the Control System monitors the detailed events associated with the flow of each instrument through the data processing of each instrument. The Control System will not allow incomplete data sets on any participant to proceed through the processing. In addition, the Control System will generate messages at specified time intervals, requesting the documents (or data) necessary to complete a particular set.

4. Scheduling Data Collection

A special routing in the Control System utilizes completeness of information collected in the current round to schedule the data collection effort in the subsequent round. At the same time, the Control System generates requests to the data processing system to produce and mail out the required Control Card and Household Summary for those scheduled interviews.

III. TECHNICAL DISCUSSION OF CONTROL SYSTEM

A. General

An overview of Control System operation is depicted in Figure 2, *Control System Flow Chart*. External monitoring points generate event and update data which are edited and stored in "save" files. While five save files exist for each type of transaction, only one is active during a given run; the others function as backup files. Events failing to pass edit are rejected and only correct entries proceed to actual updating.

Since the event and update data may be input either for

a particular participant or for a reporting unit, the software expands, when necessary, all input data to the participant level. These input data are then concatenated and sorted, and they become input to a general updating program which can add, delete, or change control file records and update key Control System directories.

The software accepts the externally generated event information and, together with input data tables describing events, report requests, and other pertinent information, processes the Control File and History File. The Control File contains a record for each participant in the survey, and every record is sequentially accessed on each run to determine if either external input or internal input described in the general event table affects the given record or if data from a record are required for reports requested from the run. When records have been processed, system summary reports are then produced and a sort step to group the individual diagnostics and notices for printing is executed.

The final step is a file maintenance program which updates the catalog and copies the updated backup files to the main files. This process protects the main files should problems occur during execution of other programs within the system, allows for identical copies of all files, and allows restarts at any job step.

B. User Features

Design of the Control System was initiated with both the user and the programmer in mind. The following sections briefly describe some of the user features incorporated into the Control System software.

1. General Grouping Capability

The general grouping capability allows the user to define and form subsets of the data file as a function of any variables contained in the file. The Boolean operators AND and OR are implied in the system control card formats and the standard set of arithmetic operators (less than, less than or equal, equal, greater than or equal, greater than, not equal) are available to the user for establishing the limits of each grouping. This feature also allows file initialization of new events to be incorporated into the system with ease.

2. Monitoring of Multiple Events

The Control System is designed to monitor multiple activities or documents simultaneously. This capability has been added to the control software through event range control cards. A status associated with each of the ranges is retained in the control record.

3. Historical Information

The System provides a complete time-sequenced history on all of the defined events occurring to each survey participant. As a sorted file, this History File can provide valuable data on response patterns and durations.

4. Alterations to Tables

Since the control software functions from tables, it is an easy matter to remove, change, or add events, diagnostics, or requests. The user simply removes/inserts a new record from/into the appropriate table via control cards.

C. General Technical Description

The Control System is programmed in FORTRAN

interview (such as insurance refunds).

D. Overview of Control System

A survey of this size and complexity requires computer support for monitoring of the data as they are processed through each stage. For this purpose, a previously developed automated survey control system was modified and tailored to the specific needs and requirements of NMCES management. The capabilities required of the NMCES Control System stem from the magnitude and complexity of the survey operation and the demand for accuracy and completeness of data.

The purpose of the Control System is to monitor the operational aspects of a survey from the initial presentation of survey instruments to the establishment of an edited data base. In general, the control program software maintains a record of the current status(es) of each individual in the survey population. To accomplish this task, event codes which are associated with selected survey activities (such as assignment, receipt, and refusal to participate) are maintained in each individual's record. Normally changes in events are externally generated as the survey instruments progress through the various monitoring points. To allow for further definition of activities, sublevels exist within events and are generated internally when a preset amount of time has elapsed.

While the system maintains the current survey status(es) of each participant, it also has the capability to retain information of a historical nature. A file, the History File, contains the total set of time-sequenced transactions which occurred to an individual throughout the survey.

In addition to monitoring numerous survey activities, the Control System provides, via user request, report and general statistics based on either operational categories (events) or other selected variables in the file.

II. OPERATIONAL RELATIONSHIP TO CONTROL SYSTEM

A. Control System Interaction with Data Flow

The general relationship of the Control System to data processing is a series of monitoring stages at specified points in the data flow. Figure 1, *Major Data Processing Components of the NMCES Data Processing System*, depicts these monitoring points. At each point specific event codes assigned to the instrument tell the Control System the status of data collection and processing for a particular reporting unit and/or individual within the reporting unit.

1. Data Receipt

At the data receipt step, each package of instruments (one per reporting unit) is carefully examined for completeness, and an event code describing the extent of completeness is assigned. The Control System contains considerable detail regarding incomplete packages in order to assist project staff in correcting these problems and in training field staff to avoid specific types of problems. Complete and nonproblem packages are event keyed as complete packages which have passed the Data Receipt edit. They are then separated into batches by type of form, in preparation for Data Entry, and the batches are sent to Document Control.

2. Document Control

The first step in the data flow through Document Control is to key the event codes and current batch location number assigned to the Reporting Unit Identification Num-

bers. This informs the Control System which documents are now ready for data entry. From this event forward the Control System maintains the current batch number on every document.

Document Control must maintain a manual master file on the location and storage of each batch. Hence any document can be physically retrieved with a minimum of time and effort.

3. Data Entry

The conversion of data to machine-readable form is done via a programmable key-to-disk system (SYCOR 440 mini-computers) which allow for basic range checking, edit, and verification. The data collection instruments (Questionnaires, Reporting Unit Folders, Control Cards, and supplemental forms) are batched by type for keying. From this point forward in the process, the Control System automatically monitors the data flow for individual documents.

Data keyed each day are transmitted to the computer facilities at TUCC for overnight production and addition to the data base. When data records are transmitted properly, the Control System receives an event code from the transmission program recording that status. If there are problems with the transmission, the proper reject event code is assigned and Document Control is notified to retrieve the nontransmitted documents for rebatching, re-event coding and rekeying. Hence the Control System is informed of the new batch number.

4. Data Edit

After data are transmitted to the data base, the machine-readable data are edited. The machine edit program is capable of examining range values, consistency of responses across data items, and routing patterns designated by specific data items. After the data are edited, the Control System is notified by the Edit program which specific documents have passed edit.

5. Data Preparation for Next Round

The Control System now performs one of its primary functions: scheduling the next round of interviewing. The scheduled assignments are then used to generate requests for Household Summaries and Control Card printing.

Edited data and the request for Household Summaries are sent to AAI. At AAI the summary of data must be assembled by reporting unit and participant. Following this step, the requested Household Summaries will be generated. An event is sent to the Control System for each summary generated.

At the same time, RTI produces the subsequent round Control Cards for each scheduled reporting unit. The Control System is similarly notified of the Control Card production. Each reporting unit for which a Control Card and Summary have been produced are considered "assigned" for the subsequent wave of interviewing.

B. General Capabilities of the Control System

1. Monitor Participant Response in Each Round

The Control System must be capable of monitoring not only the flow of data through the processing in a particular round but also the collection of data from each participant in all rounds. As indicated in Figure 1, after the data are proc-

for an IBM 370/165 with key operations being handled with Assembler routines. Running time for the system is basically a function of the number of records in the Control File and in the History File, the logical record length of the Control File, and the number of event groups to be simultaneously monitored.

For conservation of peripheral storage, packing routines have been employed. Nevertheless, four 2314 disk packs are required for the NMCES application.

Three basic functions are inherent in the design of this system: flexibility, confidentiality, and adequate backup/recovery procedures.

Flexibility is obtained through the use of table-driven operations which rely on user input for many parameters. Thus the input data alterations necessary for new uses (such as new diagnostics, new events, and new event ranges) do not require changes in the software code. Even the addition of new crosstabulation reports is accomplished with relative ease.

Assurance of confidentiality for personal data is often desired by the user. To maintain data in a confidential manner, a system control card parameter is available which notifies the system to encrypt the data file using this parameter as a key. Only individuals having knowledge of this key can unscramble the file.

A key feature of the Control System is the automatic backup/recovery procedures. Multiple backup files are maintained for every file within the system. Furthermore, the selection of backup files is performed automatically by the software as it generates the Job Control Language (JCL) for a given run.

To minimize errors of incorrect identification numbers (perhaps causing the wrong record to be updated), check-digit routines have been utilized. The use of these subroutines will permit detection of all single-digit errors and transpositions of adjacent digits. The probability of errors occurring that cancel out, thus remaining undetected by the check-digit computation, is extremely low.

IV. REPORTS AND MONITORING CAPABILITIES

A. General

The Control System is designed to be exceptionally flexible in its capability for report generation. Currently over 200

specific reports may be generated with each run or by request. Three basic types of reports are generated, as follows: diagnostic reports, listing reports, and system summary reports.

B. Diagnostic Reports

The diagnostic reports are generated with each update of the Control System and provide frequency counts of the entire event list at report time. Figure 3, *Count of Participants by Events*, shows an example of one of the five event report tables produced by each NMCES Control System run. The table shows counts for the participants at events associated with packages received at Data Receipt. This is the main stream of the package for each Interview from its field status to Control Card and Summary Generation for the next round.

C. Listing Reports

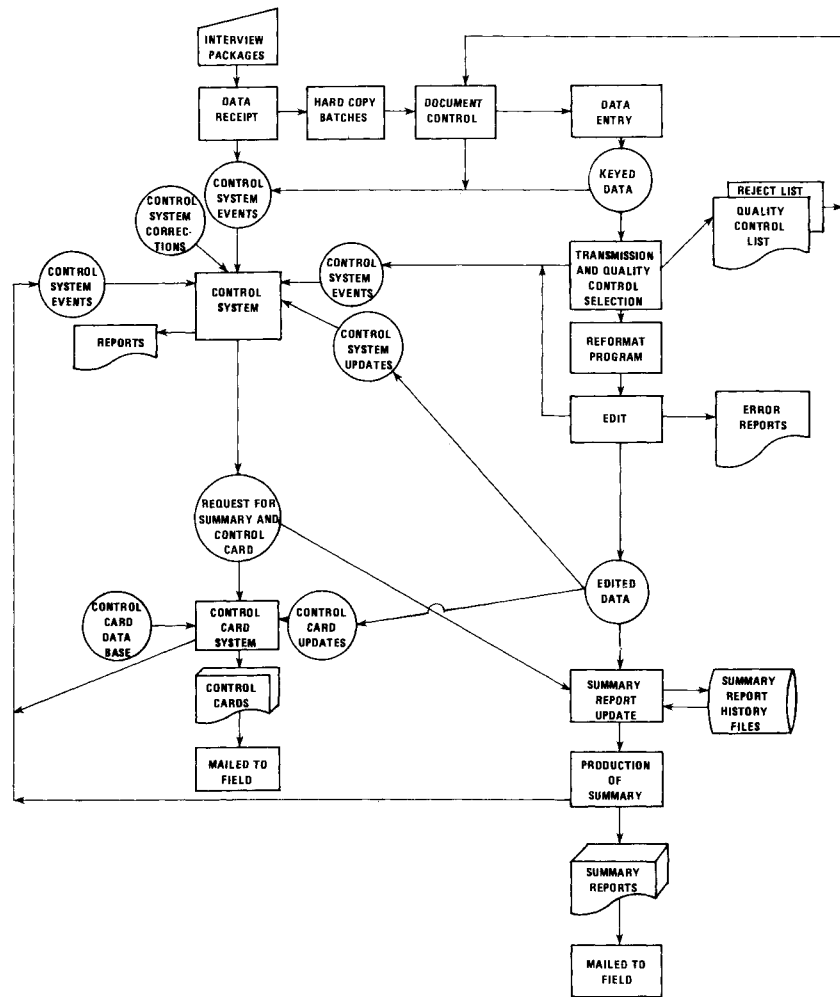
Listing reports may be generated routinely on each Control System update or may be produced to meet a specific one-time request. Listing reports were tailored to be requirements for each round's documents and their control. Figure 4, *Example of Table Headings*, gives a list of some of the report titles currently being used.

The special request listing reports are in the same output format, and the condition being met is shown. These reports list each participant meeting the specified conditions. They may be tailored through the use of Boolean Logic operators to group specified variables contained in the Control File.

D. Summary Reports

System summary reports are one-line summations of all participants which meet specified event status and grouping conditions. Any number of system summary reports may be requested and generated on any Control System update. Specific ones pertaining to data flow are routinely generated and distributed to appropriate project staff.

*Reporting Units consist of all members of a family that live at one address. A family member who is residing at a different address is interviewed in a separate reporting unit. Likewise, even if two distinct families share the same dwelling unit they are viewed as two different reporting units.



MAJOR DATA PROCESSING COMPONENTS OF THE NMCES DATA PROCESSING SYSTEM

Figure 1

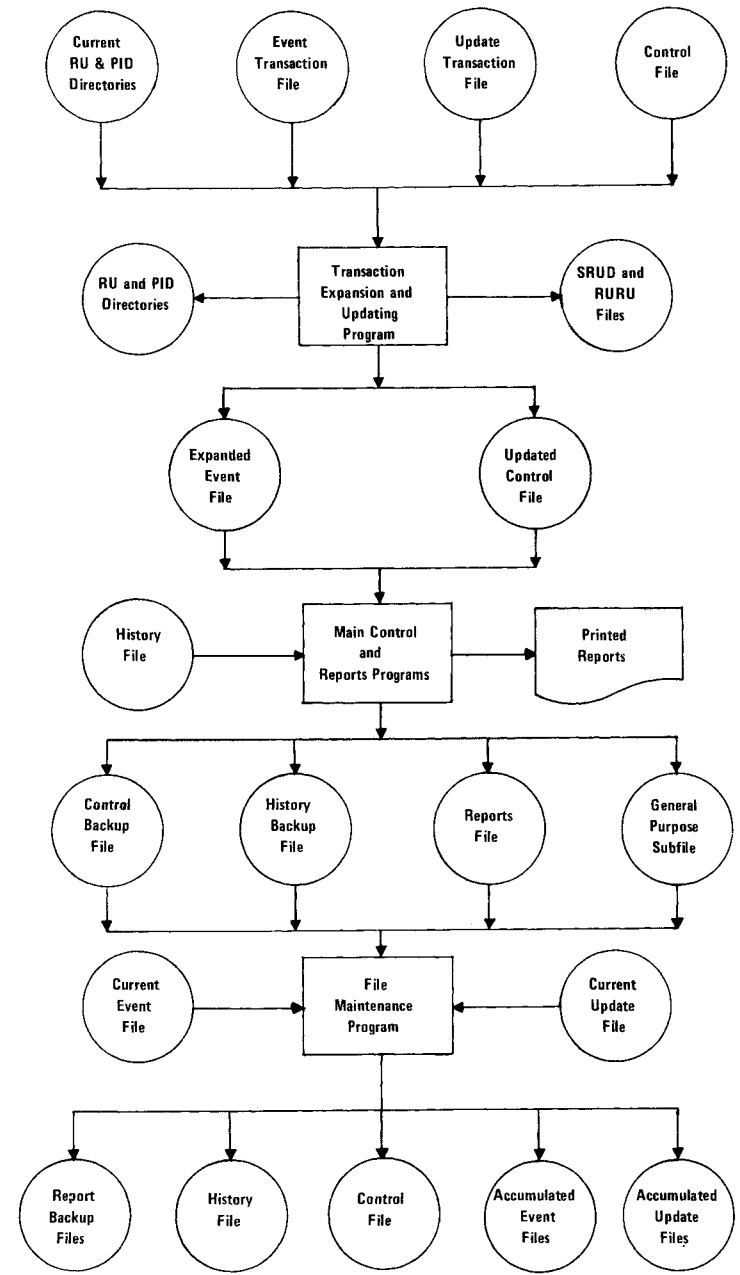


Figure 2. CONTROL SYSTEM FLOW CHART

Figure 3. COUNT OF PARTICIPANTS BY EVENTS

<u>EVENT ID</u>	<u>COUNT</u>	
121	6	In Field - Round 2
139	1	Comp. pkg. rec'd for split RU - prob in Phase I,II,III edit
145	1053	Comp. non-interview package received - Round 2
146	12	Complete package - passed DR edit - Round 2
147	122	RU was unable to be scheduled for Round 3
148	155	Field reported comp. interview never rec'd in-house - Rnd 2
151	17	Ready for scheduler - edited data, Round 2 (auto. generated)
152	1	Scheduler ready to request Summary Rep. & Control Card
161	480	In Field - Round 3
165	1	Summary Report received - Round 3
167	3	Control Card received - Round 3
177	1	Comp. pkg. rec'd for assigned RU w/prob. Phase I,II,III edit
179	3	Comp. pkg. rec'd for split RU w/prob. in Phase I,II,III edit
185	728	Comp. non-interview pkg. received - Round 3
186	178	Completed package - passed DR edit - Round 3
191	1421	Ready for scheduler - edited data (auto. generated) - Rnd. 3
193	478	Req. for Summary Rep. & Control Card generated - Round 3
195	5290	Control Card generated - Round 3
197	18773	Summary Report generated - Round 3
199	11028	Round 3 Holdover
TOTAL:	39752	

Figure 4. EXAMPLE OF TABLE HEADINGS

THE FOLL COMP PKG-SPLIT RU W/PROB IN PH IV EDIT 7 DAYS PENDING
 THE FOLL COMP PKG-SPLIT RU W/PROB IN PH IV EDIT 14 DAYS PENDING
 REQUEST FOR SR AND CC HAS NOT BEEN GENERATED AFTER 14 DAYS
 CONTROL CARD HAS NOT BEEN GENERATED AFTER REQUEST 14 DAYS AGO
 QUES. REJECTED BY TRANS. BUT NOT REBATCHED AFTER 4 DAYS
 QUES. NOT KEYED 14 DAYS AFTER BATCHING
 QUES NOT TRANSMITTED 7 DAYS AFTER KEYED EVENT
 QUES. TRANSMITTED BUT NOT EDITED AFTER 7 DAYS
 CC REJECTED BY TRANSMISSION BUT NOT REBATCHED AFTER 4 DAYS
 CC NOT KEYED 14 DAYS AFTER BATCHING
 CC NOT TRANSMITTED 7 DAYS AFTER KEYED EVENT
 CC TRANSMITTED BUT NOT EDITED AFTER 7 DAYS
 THE FOLL QUES. KEYED, TRANSMITTED, OR EDITED WITHOUT BATCHING EVENT
 THE FOLL SR KEYED, TRANSMITTED, OR EDITED WITHOUT BATCHING EVENT
 THE FOLL R6 CC KEYED, TRANSMITTED, OR EDITED WITHOUT BATCHING EVENT
 THE FOLL HSS KEYED, TRANSMITTED, OR EDITED WITHOUT BATCHING EVENT
 THE FOLL R5 CC KEYED, TRANS, OR EDITED WITHOUT BATCHING EVENT
 THE FOLL R5 RUFs KEYED, TRANS, OR EDITED WITHOUT BATCHING EVENT
 R6 EVENT RECEIVED BUT NEVER AT EVENT 281
 THE FOLL EMPL. SUPP KEYED, TRANS, OR EDITED WITHOUT BATCHING EVENT
 THE FOLL R6 ACCESS TO CARE SUPP KEYED, TRANS, OR EDITED WITHOUT BATCH E
 THE FOLL HS ARE BATCHED BUT NOT EDITED,(SR PROB)
 THE FOLL RUFs ARE BATCHED BUT NOT EDITED,(SR PROB)
 THE FOLL R5 SUPP KEYED, TRANS, OR EDITED WITHOUT BATCHING EVENT
 THE FOLL INS SUPP KEYED, TRANS, OR EDITED WITHOUT BATCHING EVENT
 THE FOLL ACCESS TO CARE SUPP BATCHED BUT NOT EDITED, (SR PROB)