

Challenges in Converting the National Crime Victimization Survey to Blaise

Charlie Carter, Roberto Picha, Michael Mangiapane, Alexis Arrington, Daniel Moshinsky, U.S. Census Bureau

1. Introduction

The National Crime Victimization Survey (NCVS) is a major national longitudinal survey conducted by the U. S. Census Bureau for the Bureau of Justice Statistics (BJS) with a sample size of about 56,000 households annually. This survey provides personal victimization and property crime data on types and incidences of crime, monetary losses and physical injuries due to crime, along with characteristics of the victims and offenders. Although the NCVS is a household-based survey, it collects crime data for every household member who is age 12 and older. The sample is divided into six panels, with the first interview occurring in a given month and again at 6-month intervals. Each household in sample is interviewed by telephone or personal visit for a total of seven times during a three-year period.

In this paper we describe the challenges we encountered converting NCVS from a Paper and Pencil Interviewing (PAPI)/Computer-Assisted Telephone Interviewing (CATI) CASES instrument to a Computer Assisted Personal Interviewing (CAPI)/CATI Blaise instrument.

1.1 Survey Challenges

- Our first challenge was to create an approach for collecting multiple crime reports without negatively impacting instrument performance, i.e., the time delay between questions, for selecting in any order the next available household member to interview, and for navigating effortlessly between the various sections of the survey instrument.
- The second challenge involved developing an approach for retrieving details on the reported crimes and comparing them between members of the household so that the interviewer can determine if the crime is a duplicate of one that has already been reported by another household member during the current interview period as well as crimes reported six months ago during the prior enumeration.
- The third challenge with the NCVS instrument was integrating an “Editing and Coding” (E&C) instrument into the regular collection instrument and only allowing a certain set of questions for review, editing, or coding for each eligible household member, while maintaining the integrity of all of the other survey questions that were not E&C eligible.

1.2 NCVS Technical Description

The NCVS data model was developed with Blaise 4.6 software. It is a large and complex data model, containing 55 blocks and 4741 block instances. The NCVS data model produces over 60 ASCII relational output files. There are a total of 162,656 data fields with a total length of 2,289,961 bytes defined in the model. There are many different

data types and relationships in the data model. The technical description for the NCVS data model is shown in Table 1.

Table 1, Technical Description

1. Overall counts	Value
Number of uniquely defined fields	971
Number of elementary fields	924
Number of defined data fields	162656
Number of defined block fields	47
Number of defined blocks	55
Number of embedded blocks	1
Number of block instances	4741
Number of key fields	1
Number of defined answer categories	295
Total length of string fields	2131051
Total length of open fields	0
Total length of field texts	69159
Total length of value texts	39263
Number of stored signals and checks	21424
Total number of signals and checks	21424

2. Data fields	Number	Length
fields in data file		
Integer	11107	45493
Real	3	21
Enumerated	83049	87616
Set	14868	25676
Classification	0	0
Datatype	5	40
Timetype	8	64
String	53616	2131051
Open	0	-
Total in data model	162656	2289961

2. Multiple Crime Reports

The NCVS is unique among other surveys administered at the U.S. Census Bureau because of the total number of household members that are eligible to be interviewed, i.e., up to 30 members age 12 and older. A maximum of 30 crime reports can be collected for each member. If an individual respondent is not available or refuses to be interviewed, that person is either interviewed at a later date or flagged as a non-interview. Once the interviewer is successful at interviewing an eligible household member during the first enumeration, usually age 18 and older, that person is flagged as the household respondent. All of the other eligible household members are individual respondents who must also be interviewed before the survey close out date or be flagged as a non-interview.

The distinction between the household respondent and the individual respondent is important because there are multiple questions within the survey that are slated only for the household respondent. They include questions about the household composition and their characteristics, household income, housing characteristics, vandalism, and use of telephone.

Additionally, there are questions that are common to both the household respondent and the individual respondents. Those common questions include, but are not limited to, the screener questions, which include a series of probes about different categories of crimes. The screener questions are asked to report the number of incidents in the reference period and the number of times they have occurred. The household respondent and the individual respondent also share in common the crime report questions that are asked to collect details about all the crimes reported in the screener questions as well as questions about current employment and mobility [see Figure 2.1 and Figure 2.2].

For the most part, the number of crimes reported in the screener questions determine how many crime reports will need to be collected. The large number of questions and responses in both the screener and crime report blocks produce an even larger number of skip patterns because the questions are interdependent.

Our initial approach to collecting multiple crime reports for a respondent posed several challenges, as we will describe below.

2.1 Instrument Design for Multiple Crime Reports

The data collection process starts with the household respondent; the instrument collects their screening questions and crime reports, if any. Once that interview is completed the instrument takes the interviewer to the *SetupRoster* question where all members eligible for interviewing are displayed; the interviewer will then select a household member available at the time of interview. Once selected the interviewer will start interviewing until that respondent has completed their interview. The instrument will go back to *SetupRoster* allowing the interviewer to select another member or conclude the interview if all eligible members were interviewed.

In an ideal world, all of the eligible household members are at home and available to complete their interview. However, there will be situations when a respondent has to terminate their interview early, or an eligible household member may not be available for interview. The interviewer must be able to resume a partial interview with a respondent

from the previous termination point, or start an interview with another available household member.

Figure 2.1, A Question from the Screening Section.

I'm going to read some examples that will give you an idea of the kinds of crimes this study covers. As I go through them, tell me if any of these happened to you in the last 6 months, that is, since January 05, 2006. Was something belonging to YOU stolen, such as --

- ♦ Read each category

- Things that you carry, like luggage, a wallet, purse, briefcase, book -
- Clothing, jewelry, or cellphone -
- Bicycle or sports equipment -
- Things in your home - like a TV, stereo, or tools -
- Things outside your home such as a garden hose or lawn furniture -
- Things belonging to children in the household -
- Things from a vehicle, such as a package, groceries, camera, or CDs - OR
- Did anyone ATTEMPT to steal anything belonging to you?

- ♦ Ask only if necessary:

Did any incidents of this type happen to you?

Figure 2.2, A Question from the Crime Report Section.

Did this incident happen ...

- ♦ Read each category until respondent says 'yes', then enter appropriate precode.

11. In your home or lodging?

12. Near your home or lodging?

13. At, in, or near a friends's/relative's/neighbor's home?

14. At a commercial place?

15. In a parking lot or garage?

16. At school?

17. In open areas, on the street, or on public transportation?

36. Somewhere else?

2.2 Challenges of Collecting Multiple Crime Reports

In order to facilitate testing early in the development, sections of the survey had been programmed as stand-alone modules. When the instrument was first integrated, screener questions were arranged in arrayed blocks, each block corresponding to a potential eligible household member. Crime report questions were also organized into an array sized according to the maximum potential number of crimes each household member may report (30), for each of the 10 categories of eligible crimes. The resultant data model consisted of some 9000 ($30 * 10 * 30$) nested blocks, exceeded the 16,300 Blaise page limit, and failed to compile. After we reduced the size of one of the arrays associated with the maximum number of crimes allowed for each screener question, we avoided the compilation error; however, the performance remained undesirable.

The instrument load time was about 45 seconds, and there was a re-execution delay of about 2 seconds when proceeding from question to question. The poor performance was, in part, due to the large number of arrayed blocks being on route at the same time. Clearly, a conceptually different approach was called for.

2.3 Solution for Collecting Multiple Crime Reports

The survey architecture was re-designed to separate the data collection activities from data storage. Data would be collected in one non-arrayed block, and then copied into a separate arrayed storage block. This approach takes the large arrayed blocks that had slowed performance off the re-execution path, and also allows for a more flexible retrieval of partial interviews and navigation between household members.

The re-designed instrument works as follows. First, the screening questions are collected for the current respondent in the *Screening* block (see Figure 2.3). After all the screening questions are answered, the data is copied into a separate arrayed block, *NCVSScreenData* – the data storage structure for the screening questions. Then, a count of reported incidents for the respondent is calculated. If incidents have been reported, crime reports are collected in the *CrimeReport* block. After the completion of each crime report, the data is again copied to a separate arrayed storage block, called *NCVSCrimeReportData*. The interviewer can then pick another household member to interview, and the process repeats. Data collection blocks are re-used over and over again for every eligible household respondent, and for each new crime report. The data collection blocks are set to EMPTY each time their data is sent to the storage blocks, ensuring that when rules are re-executed they will be sent on route again. This iterative process continues until all crime reports have been collected and stored for each crime reported by the respondent in the Screening section – and until all available household members have been interviewed. It must also be noted that each element of the storage block arrays has the same exact field structure as the corresponding data collection block, which makes copying data from one to the other a matter of simple assignment.

Data is also copied into the storage arrays for partial interviews -- whenever the interviewer needs to schedule a follow-up interview with the respondent. Another advantage of using the storage arrays is that they allow the retrieval and continuation of a partial interview at follow-up time. Counters and flags within the instrument are employed to keep track of the household members with partial interviews, and of how many and which crime reports remain to be completed for them.

Figure 2.3, Critical Fields Used in the Data Storage Approach

Field Names	Field Definitions
NCVSScreenData	ARRAY[1..30] OF BNCVS1
Screening	BNCVS1
NCVSCrimeReportData	ARRAY[1..30] OF BNCVS2
CrimeReport	BNCVS2
SetupRoster	bSetupRoster

If the selected member cannot continue with the interview, the interviewer can setup a callback with an appointment for a later date. The interviewer can continue by selecting another member in the household from the *SetupRoster* block.

The instrument controls part of that process with the use of special person-level flags that keep track of the completion status of each household member, and, if there has been a termination, the location of the last saved crime report within the *NCVSCrimeReportData* array, and the number of crime reports that remain to be collected. In addition, each block contains identifying information (e.g. person number and incident number). These flags make possible the retrieval of any incomplete screening or crime report section. The appropriate data is retrieved from the storage block using the flags, and it is then copied into the data collection block. This functionality is critical to the implementation of un-duplication and editing and coding capabilities, as we will discuss in more details in Sections 3 and 4 of this paper.

A Brief summary of the pros and cons of this approach are listed in Figure 2.4 below.

Figure 2.4, A Summary of the Pros and Cons of the Developed Solution.

<p>Pros</p> <ul style="list-style-type: none"> • Allows the re-use of the same block as many times as needed to collect all the incident and crime reports. • Greatly improves the performance of the instrument when navigating from question to question. • Reduces the number of variables within the instrument. <p>Cons</p> <ul style="list-style-type: none"> • Prohibits returning to an incident already collected. • Complex coding for setting control flags deters changes to the rules for setting these flags.
--

3. Un-Duplication

In the course of an interview, it is likely that multiple respondents in a household will mention the same crime that happened in the last year, either in the current interview reference period of the last six months or the prior interview reference period. As part of the system requirements for the NCVS, the instrument displays certain details of the crimes to the interviewer so that they may compare the reported crimes for similarities and make decisions about whether the compared crimes are duplicates. This process is known as un-duplication and it is used to review the reported crimes to determine if they are suspected duplicates of one another. If any duplicates are found, one of the crimes

reported in the current period will be marked for deletion from the data that will be processed into the final tabulated criminal victimization data.

3.1 Un-duplication Process and Structure

The process of un-duplication of the reported crime incidents takes place after all of the incidents have been collected from the respondent. All incidents that were reported in the current interview are compared against one another with details about both crimes being displayed on the screen for comparison. If an interviewer suspects that the two displayed crime reports are duplicates, one of the crime reports is flagged by the instrument to be deleted during post-processing; then the instrument moves to the next crime report comparison. This process is repeated by comparing all of the current incidents with up to four incidents that were reported in the previous interview six months ago, if any were reported.

The main structure of the un-duplication table is a cross-reference table that only puts on-route the reported crimes from the current interview that had not been compared to one another. Each row and column represents reported crimes in the current interview. When comparing the current respondent's crime reports to all crime reports completed this interview period, the displayed rows would represent all crimes reported by all respondents, and the displayed columns would be the crimes that were reported by the respondent just interviewed. If any crime is reported as a duplicate, the crime represented in the row of the table is the one that is marked as a duplicate, and any columns past the duplicated crime are taken off-path since there is no reason to continue the comparisons. The cursor moves on to the next crime to be compared until all crimes from the current interview have been compared.

A second table is used to compare all crimes reported by the current respondent to any crimes that were reported in a previous interview. Each row represents the crimes reported by the current respondent, and each column represents a crime that was reported in the previous interview. Just like the previous table, if a crime is marked as a duplicate, any columns past the duplicated crime are taken off-path. Also, if any crime was marked as a duplicate in the previous table, it will not come up for comparison in this table. The cursor moves on to the next crime until all comparisons have been made.

3.2 Challenges of Un-duplication

The biggest challenge to un-duplication was being able to retrieve details on the reported crimes and display them in such a way that the interviewer could quickly compare the two crimes against one another to determine if one is suspected to be a duplicate of the other. To simplify the instrument for collecting the details of the crime, one block is used to collect the current crime report. After the details are collected, the report is moved into an array for storage since the interviewer does not retrieve data from storage when they're collecting the crime report. While every single detail in the crime report did not need to be displayed to the interviewer, they did need a certain information from the crime report, which included the date, location, which respondent reported it, the type of crime (theft, assault, burglary, etc.), who was there, if weapons were used, what was taken if it was a theft, what injuries there were, and any additional information that the interviewer reported in a summary question at the end of the crime report.

The immediate problem during implementation was that while we were able to retrieve the correct information from the storage array and piece it together for un-duplication, the amount of data being brought in was too large for Blaise to handle inside a single block. We were trying to use defined parameters to bring each piece of information out of the storage array into un-duplication and we exceeded Blaise's limits on the size of one block (Length of data for one block exceeds 65519 positions). The block in question was not the un-duplication block; it was the core instrument block that drives the main interview of the NCVS, including screening questions and the collection of details on the crime. There wasn't much we could do to change the core block or the other interviewing blocks, so adjustments had to be made to un-duplication.

Another challenge of un-duplication was making the screen readable to the interviewer. We needed to display some type of summary of the crime report details that the interviewer could read, it also needed to be formatted in such a way that the two crimes could be compared without forcing the interviewer to look all over the screen to compare each detail. Ideally we wanted to put the summaries on the screen next to one another vertically so that the two could be compared line by line, saving the interviewer time since they could look across the screen to make the comparison.

3.3 Solutions for Un-duplication

The initial approach of using parameters to bring in the crime information was abandoned for another method of displaying the crime details by moving the un-duplication block to a higher-level block in the instrument. A bonus of creating the un-duplication display at this level was that the un-duplication display was able to re-use some text fills that were already defined, saving the instrument the overhead of creating a number of new text fills. Tests of the maximum size of the string that would be used for displaying the crime data in un-duplication were done, and the maximum was 635 characters, which turned out to be more than enough for our purposes.

To make the un-duplication display readable to the interviewer it was calculated that the instrument could display 90 characters per line on the screen since we wanted to display the two crime reports side-by-side. A substring of 45 characters for each displayed line was created. Each line contained the data item and a buffer string of up to 45 blank characters concatenated. Blaise would automatically cut off the string at 45 characters, whether there was part of the data or this string buffer there. This was done to keep everything aligned horizontally and vertically, especially if there were blank lines because one crime description was shorter than the other. The instrument would generate one string with all of the summarized information to be brought into the un-duplication block. Inside the un-duplication block we used a direct reference to the un-duplication string that we had created in this higher-level block. Each line of the un-duplication string from each incident was brought over and concatenated together so that they were aligned. A separate block was used to create the same un-duplication display for any crimes reported in the previous enumeration that were sent on the input file, and the same process was used to put the un-duplication display strings together.

Below is an example of un-duplication screens for each un-duplication table. Un-duplication is being run for the second respondent, who reported two incidents during their interview (incidents three and four). In the example from Figure 3.1, the comparisons are being made with two incidents that were reported by the first respondent (incidents one and two). Since the interviewer answered on the first row that incident one

is a duplicate of incident three, the comparison of incident one and four is taken off-route in the table. The interviewer would proceed to the second table of un-duplication if applicable. Figure 3.2 illustrates un-duplication of crime reports three and four in the current interview being compared with three crime reports made in the NCVS interview during the previous enumeration.

Figure 3.1, Comparing Incidents Reported in the Current Enumeration

- Compare the incident data below. Do not discuss incidents with the respondent.
- Is the incident on the left a duplicate of the incident on the right?

<p>INCIDENT REPORTED DURING CURRENT ENUMERATION Enum: 02 Ln No: 1 Inc. No: 2</p> <p>MONTH: 2 WHERE: In own dwelling/att garage. Respondent and other HHLD member(s) present WEAPON: Blunt object, THREATEN: Weapon present, NOTHING TAKEN AND NO ATTEMPT TO TAKE ANYTHI SUMMARY: offender tried to break into house , was warned off, police called, offender fou nd 2 days later, replaced door</p>	<p>INCIDENT REPORTED DURING CURRENT ENUMERATION Enum: 02 Ln No: 2 Inc. No: 4</p> <p>MONTH: 2 WHERE: On street. Respondent was present. WEAPON: Knife, ATTACK: Stabbed/cut w/ knife, THEFT: Cash, Wallet, Credit cards, SUMMARY: mugged in parking lot of mall, cut o n arm defending self, stitched at hospital, w allet found hours later</p>
--	--

1. Yes
 2. No

	Undup Cur Enum								
Incident[1]			1						
Incident[2]			2						
Incident[3]									

Figure 3.2, Comparing Incidents Reported in the Previous Enumeration

- Compare the incident data below. Do not discuss incidents with the respondent.
- Is the incident on the left a duplicate of the incident on the right?

<p>INCIDENT REPORTED DURING CURRENT ENUMERATION Enum: 02 Ln No: 2 Inc. No: 4</p> <p>MONTH: 2 WHERE: On street. Respondent was present. WEAPON: Knife, ATTACK: Stabbed/cut w/ knife, THEFT: Cash, Wallet, Credit cards, SUMMARY: mugged in parking lot of mall, cut on arm defending self, stitched at hospital, wallet found hours later</p>	<p>INCIDENT REPORTED DURING PREVIOUS ENUMERATION Enum: 1 Ln No: 01 Inc. No: 2</p> <p>MONTH: 09 WHERE: Commercial parking lot. WEAPON: Other personal/portable objects, SUMMARY: ART PENS STOLEN FROM PURSE IN CLASR ROOM/VALUE \$12/REPORTED TO SCHOOL OFFICIAL/NO DAMAGE DONE/OFFENDER WAS A COWORKER</p>
--	--

1. Yes
 2. No

	Undup Prior Enum	Undup Prior Enum	Undup Prior Enum	Undup Prior Enum
Incident[2]				
Incident[3]	2	2	2	
Incident[4]	2			
[5]				

4. Editing and Coding

Editing and coding (E&C) is the process of reviewing data that was collected during interviewing for quality assurance. Criteria which determined whether the E&C instrument would be called depended on any race category set to “other” for the first time the case was interviewed [see Figure 4.1], there were one or more crime incidents reported in the household at time of interviewing [see Figure 4.2] – either telephone [CATI] and/or personal [CAPI], screener information at the person level - person called police, or person didn’t call police regarding a potential crime [see Figure 4.3].

The NCVS E&C Operation in CASES was conducted after the NCVS CATI production interviewing closeout using a separate CASES instrument and clerical editing of the PAPI forms served a comparable purpose. Based on conversations between the stakeholders and the authors, it was decided that one instrument was to be used for data collection whether CATI and/or CAPI interviewing, then the same instrument would be used for E&C of CATI and CAPI cases once regular interviewing was complete.

Figure 4.1 Race = 6 (Other)

The screenshot shows the 'National Crime Victimization Survey -- NCVS Questions ver 14.02' window. The interface includes a menu bar with 'Forms', 'Answer', 'Navigate', 'Options', and 'Help'. Below the menu is a navigation pane with 'Main', 'HH Roster', 'FAQs', and 'Incident Review'. The main content area displays a table of household members:

LNO	NAME	REL	ORG	RACE	RACE SPECIFY
1	Ray Ray	Ref Person	No	1	
2	Teresa Ray	Wife	No	1	
3	Junior Ray	Son	No	1 6	Greek

Below the table is a list of race categories with checkboxes:

- 1. White
- 2. Black or African American
- 3. American Indian or Alaska Native
- 4. Asian
- 5. Native Hawaiian or Other Pacific Islander

At the bottom, there is a data entry table with columns for 'Race', 'Review Catchall1', and 'Review Catchall2'. The rows correspond to the household members in the table above.

	Race	Review Catchall1	Review Catchall2
2	Teresa Ray		
3	Junior Ray		

The status bar at the bottom of the window displays: 00000001 | RACEROSTER_EC | 8:40:22 AM | 6-12-2007 | Talking To: | About: Ray Ray | 289/1492

Figure 4.2, Crime Incident Reports for the Household

National Crime Victimization Survey -- NCVS Questions ver 14.02

Forms Answer Navigate Options Help Show Watch Window

Main | HH Roster | FAQs | Incident Review

◆ Enter the precode to select the next incident to edit or enter 31 if finished editing all incidents

L	I	S	REF	WHAT
N	N	T	CODE	HAPPENED
O	C	A		
1	1	1	P	flower pot stolen

Review incident roster

00000001 | INCROSTER_EC | 8:42:03 AM | 6-12-2007 | Talking To: | About: Ray Ray | 291/1492

Figure 4.3, Catchall Questions.

National Crime Victimization Survey -- NCVS Questions ver 14.02

Forms Answer Navigate Options Help Show Watch Window

Main | HH Roster | FAQs | Incident Review

? [F1]

CATCHALL QUESTIONS

◆ Incident Description: **bike stolen from neighbor's yard**

◆ Select up to 3 precodes, separate with commas

<input type="checkbox"/> 11. Rape	<input type="checkbox"/> 21. Motor Vehicle Theft
<input type="checkbox"/> 12. Attempted Rape	<input type="checkbox"/> 22. Attempted Motor Vehicle Theft
<input type="checkbox"/> 13. Robbery	<input type="checkbox"/> 23. Motor Vehicle Accident
<input type="checkbox"/> 14. Attempted Robbery/Threatened Robbery	<input type="checkbox"/> 24. Vandalism (Against Household or Household Member's Property)
<input type="checkbox"/> 15. Assault	<input type="checkbox"/> 25. Prowlers/Peeping Toms
<input type="checkbox"/> 16. Attempted Assault/Threatened Assault	<input type="checkbox"/> 26. Crime Against Household, Other than above
<input type="checkbox"/> 17. Burglary	<input type="checkbox"/> 27. Not a Crime
<input type="checkbox"/> 18. Attempted Burglary	<input type="checkbox"/> 28. Crime Against Someone Else or Society
<input type="checkbox"/> 19. Larceny (Household or Personal)	<input type="checkbox"/> 29. Unable to Classify
<input type="checkbox"/> 20. Attempted Larceny (Household or Personal)	

	Race	Review Catchall1	Review Catchall2
2	Teresa Ray		
3	Junior Ray	1,2,3,4,5	

00000001 | CATCHALL2_EC | 8:41:38 AM | 6-12-2007 | Talking To: | About: Ray Ray | 289/1492

4.1 Editing & Coding Process and Structure

The NCVS E&C Blaise Instrument was the EXACT same instrument modules/blocks as the production NCVS instrument; however, it was to display only parts of the Screener section and Crime Report section, not the entire sections of each. Only the leading questions of crimes in the Screener, total number of times the crime occurred, and description of each crime were to be displayed, such as something was stolen, someone broke in, motor vehicle theft, attacked where, attacked how, sexual assault, called police crime and not called police crime. Specific data from the Crime Report section, which consisted of particular details of each crime, were to be displayed, not all of the Screener and Crime Report sections.

4.2 Challenges of Editing & Coding

A major performance issue occurred while attempting to run the instrument *and* populate the E&C block with data collected from interviewing with maximum 30 household members and maximum 30 incidents each; the instrument run time was greater than 60 seconds [one minute], which far exceeded the 20 seconds or less instrument run time anticipated for a survey.

Another performance issue occurred during the *initial* design of the E&C block whereby the block was designed as a flat file of maximum 30 household members with each household member record with maximum 30 incidents worth of data, which exceeded the block length size allowed of 32767 for any one record.

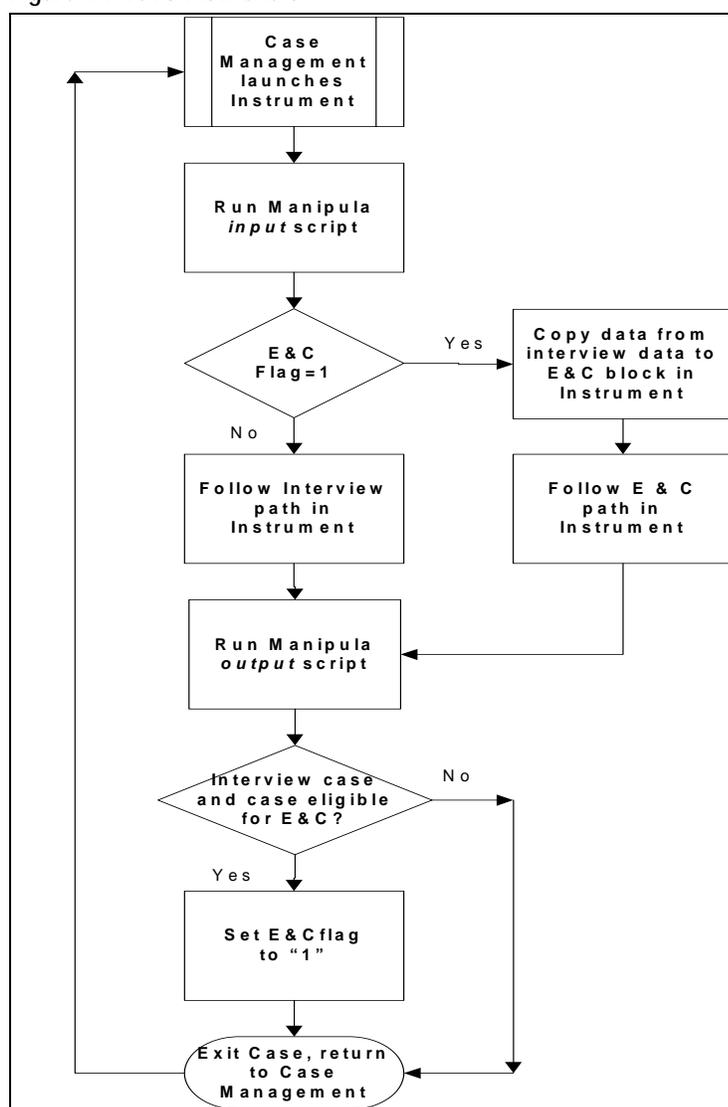
The E&C instrument was based on the premise that it would follow a different path than the production instrument. This instrument will read in a parameter that is passed in by the E&C control scripts that tells it to follow the E&C path instead of the interviewing path. In dealing with this path, although the instrument *started* with all of the production data from the original production case, some of that data was being “lost” due to the routing of the E&C instrument and off-path data for E&C questions may not appear in the Blaise databases. This challenge was how to run the E&C component of the instrument without losing the data that was collected during interviewing.

As specifications changed in the interviewing component, sometimes the same changes that were made to the interviewing instrument were not reflected in the editing and coding component which created discrepancies in screen transparencies, the flow of data and skip pattern changes just to name a few.

4.3 Solution for Editing and Coding

To resolve the performance issues – instrument run time, the E&C block would be populated with the interviewing data collected; via the Manipula *input* script *after* the E&C flag was set to “1”. This flag would be set using an Manipula *output* script *after* interviewing data collected/case complete. This was the quickest and most efficient mode to update the E&C blocks without tying up resources before calling the instrument [see Figure 4.4].

Figure 4.4 NCVS Flow Chart



To resolve the exceeding block size, a separate E&C component block was created to be a separate block within the instrument and to include only those variables that were needed from the interviewing blocks. The E&C block consisted of three tables – one table consisting of E&C household roster [see Figure 4.5]; the second table which arrayed the maximum 30 household members with race, called police crime, not called police crime, and two E&C catchall questions [see Figures 4.1 and 4.3]; and a third table that arrayed the maximum 30 crime report incidents that matched incidents to the person number and line number [see Figure 4.2]. The E&C household roster displayed pertinent data about the household member, i.e., name, age, gender, relationship, member status; the second table would be displayed *only* when at least one member of the household had a race = other *or* at least one member of the household had responded “Yes” to the called police crime *or* not called police crime questions [i.e. Catchall Questions]; the third table would be displayed when at least one crime report was recorded, then the household member data would be displayed, selecting one incident at a time along with the contents of each incident to keep the buffer size at a minimum [see Figure 4.6].

Figure 4.5 E&C Household Roster

National Crime Victimization Survey -- NCVS Questions ver 14.03

Forms Answer Navigate Options Help

Main | HH Roster | FAQs | Incident Review

NCVS Editing/Coding Roster

Control Number : 999999999999999999 Telephone Center/RO: 28

Phone Number: (800) 555-1212

Address:
100 NEW STREET ADDRESS
NEW CITY NAME, AZ 22222 1111

L		S	A	R	S	M	DATE OF	H
N		SLF/	T	G	E	E	LAST	H
0	NAME	PRXY	A	E	L	X	INTERVIEW	R
1	Ray Ray	Self DONE-Int	30	Ref Person	M	yes	7-18-2007	X
2	Wife Ray	Self DONE-Int	30	Wife	F	yes	7-18-2007	
3	Junior Ray	Self DONE-Int	13	Son	M	yes	7-18-2007	

1. Enter 1 to continue

Household

00000031 | SCREEN1 | 11:15:55 AM | 7-18-2007 | Talking To: | About: Ray Ray | 288/1492

Figure 4.6, Crime Report Summary – Beginning of E&C Crime Report

National Crime Victimization Survey -- NCVS Questions ver 14.02

Forms Answer Navigate Options Help Show Watch Window

Main | HH Roster | FAQs | Incident Review

◆ YOU ARE EDITING LINE: 1 Ray Ray /INCIDENT NUMBER : 1

◆ SUMMARY:
Summary #1 - Ray Ray, SQTheft in January, flower pot stolen from own yard

1. Enter 1 to continue

Why deleted in prod.

Review deleted incidents

Incident Summary

00000001 | SUMMARY_EC | 8:42:25 AM | 6-12-2007 | Talking To: | About: Ray Ray | 292/1492

The situation of data being “lost” due to the routing of the E&C instrument was dealt with by strategically placing keep statements at block levels within the instrument so that

no data would be lost when running the E&C component, and all other components [CAPI and CATI interviewing (data collection) blocks] were off-line/disabled. To ensure the keep statements worked properly and that data collected was actually being kept, a sample of test data were reviewed using the Watch Window debug feature, during both interviewing and E&C stages to see at what point the data was being lost, and at what point the keep statement would be most effective; block level or field level.

The problem of disabling the E&C component of the instrument without losing data collected during interviewing was accomplished via an E&C flag initialized to “blank” or “empty” within the instrument according to the transaction code in the Manipula transaction script that calls the instrument’s E&C component. The E&C component of the instrument was disabled while interviewing, but *after* the interviewing data collection was done and the case was determined to be eligible for E&C, the E&C flag was set to “1” via the Manipula output script.

With specification changes in the interviewing component, it was critical to ensure that the same changes that were made to the interviewing instrument were reflected in the editing and coding component for consistency and data review accuracy. This included, but was not limited to, screen text changes, variable name changes, and skip pattern changes. The instrument development team had to work closely and notify team members when changes were made to both components.

5. Conclusion

The NCVS is a multi-functional and complex survey that has operated continuously since 1972. The BJS uses the crime data for the dissemination of public files that are released on a semi-annual basis. When we first started converting NCVS from a PAPI/CATI CASES interview to a CAPI/CATI Blaise instrument, we encountered multiple challenges as discussed in this paper and would suggest the following recommendations for future survey efforts with a similar survey architecture.

- In order to improve performance, separate the data collection activities from the data storage when there are multiple respondents that are being asked the exact same series of questions
- Develop the E&C instrument as a second stage instrument instead of integrating it into the regular data collection instrument.

We were able to surmount all of the challenges discussed in this paper through good communication with our internal sponsor, which started by educating the sponsor about how their design decisions may adversely impact development issues such as instrument performance. That was our key to a successful deployment of the NCVS instrument into the field for data collection in July of 2006.

6. Disclaimer

The views expressed on (statistical, methodological, technical, or operational) issues are those of the author(s) and not necessarily those of the U.S. Census Bureau.