

**DATA PROCESSING  
MANAGER'S MANUAL**

# SAVVY

*Sample Vital Registration with Verbal Autopsy*



**MEASURE Evaluation**

**U.S. Census Bureau**



# Data Processing Manager's Manual

## SAVVY

### Sample Vital Registration with Verbal Autopsy



U S C E N S U S B U R E A U  
*Helping You Make Informed Decisions*

This manual was made possible by support from the U.S. Agency for International Development (USAID) under the terms of Cooperative Agreement GPO-A-00-03-00003-00 and from the Office of the Global AIDS Coordinator (OGAC). The opinions expressed are those of the authors and do not necessarily reflect the views of USAID, OGAC, or the United States government.

**August 2007**

**MS-07-26-DPM**



Printed on Recycled Paper



## ACKNOWLEDGMENTS

This manual is one of a series of instructional materials for the Sample Vital Registration with Verbal Autopsy (SAVVY) system prepared by the International Programs Center, Population Division, U.S. Census Bureau, Washington, DC, USA; and the MEASURE Evaluation project, based at the University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

This manual was written by Sarah Osborne and Michael T. Stroot of the U.S. Census Bureau.

Series editors: Philip Setel, Victoria Velkoff, and Loraine West.

The authors wish to express appreciation for review and comment by numerous colleagues, including Dennis Donahue of the U.S. Census Bureau. Copy editing and layout were handled by Nash Herndon and Samhita Brown of MEASURE Evaluation.

### Cover

The cover shows a wood carving by Colombian artist Juan de la Cruz Saavedra from a photograph by Alex Trembl of Durham, NC, USA, and is reproduced with permission from Family Health International of Research Triangle Park, NC, USA.



# Table of Contents

	ACKNOWLEDGMENTS	i
	LIST OF ACRONYMS	v
	SAVVY MANUAL SERIES	vi
<b>1.</b>	<b>What is SAVVY?</b>	<b>1</b>
	DEMOGRAPHIC SURVEILLANCE	2
	MORTALITY SURVEILLANCE WITH VERBAL AUTOPSY	3
	WHAT IS VERBAL AUTOPSY?	3
	ABOUT THIS MANUAL	6
<b>2.</b>	<b>Overview of Data Processing Systems</b>	<b>7</b>
<b>3.</b>	<b>Roles and Responsibilities of the SAVVY Data Processing Manager</b>	<b>9</b>
	RESPONSIBILITIES	9
	TASKS	10
	CONFIDENTIALITY	11
	HONESTY	11
<b>4.</b>	<b>Baseline Census</b>	<b>13</b>
	BASILINE CENSUS QUESTIONNAIRES TRACKING	13
	ENTER AND VERIFY DATA	14
	STRUCTURE EDIT	15
	FILE CONCATENATION	16
	CONSISTENCY EDITS	17
	WEIGHTING DATA	17
	TABULATIONS	17
<b>5.</b>	<b>Verbal Autopsy</b>	<b>19</b>
	VERBAL AUTOPSY CENSUS QUESTIONNAIRES TRACKING	20
	KEYING AND VERIFICATION OF VA FORMS	21
	ASSIGN AND CHECKOUT VA FORMS FOR ICD CODING	22
	TWO-WAY INDEPENDENT ICD CODING OF VA FORMS AND VERIFICATION	22
	ICD CODE RECONCILIATION	23
	CODE ACME/TRANSAX ICD CODE TO SAVVY CAUSE OF DEATH CODE	23
	MERGE WITH VA FILE AND CREATE FINAL ANALYSIS FILE	23

<b>6.</b>	<b>Census Update</b>	<b>25</b>
	PREVIOUS CENSUS DATA FILES	25
	PRINT CENSUS UPDATE MATERIALS BY ASSIGNMENT AREAS	25
	CENSUS UPDATE QUESTIONNAIRES TRACKING	26
	UPDATE EXISTING CENSUS RECORDS AND VERIFICATION	26
	CREATE NEW CENSUS RECORDS	27
	COMBINE NEW AND EXISTING RECORDS INTO A SINGLE FILE	27
	STRUCTURE EDIT	27
	FILE CONCATENATION	28
	CONSISTENCY EDITS	28
	WEIGHTING DATA	28
	TABULATIONS	28
<b>7.</b>	<b>Preparation for Analysis</b>	<b>29</b>
<b>8.</b>	<b>Glossary of Terms</b>	<b>31</b>
<b>9.</b>	<b>Appendices</b>	<b>37</b>
	APPENDIX A: BASELINE CENSUS PROCESSING PROGRAMS	38
	APPENDIX B: PRODUCTION OF CENSUS UPDATE DOCUMENTS	42
	APPENDIX C: CENSUS UPDATE PROCESSING	47
	APPENDIX D: VERBAL AUTOPSY PROCESSING	52
	APPENDIX E: ILLUSTRATIVE TABULATION AND CROSS TABULATION	56

## LIST OF ACRONYMS

AA	assignment area
AIDS	acquired immunodeficiency syndrome
ANC	antenatal care
CSC	census supervisor coordinator
CSPro	Census and Survey Processing System (software)
DSS	demographic surveillance system
HIV	human immunodeficiency virus
ICD, ICD-10	<i>International Statistical Classification of Diseases and Related Health Problems</i> , second edition, 10th revision
ID	identification
KI	key informant
MCH	maternal-child health
MSS	mortality surveillance system
OJT	on-the-job training
QA	quality assurance
SA	supervisory area
SAVVY	Sample Vital Registration with Verbal Autopsy
SBS	SAVVY budget spreadsheet
TB	tuberculosis
TBA	traditional birth attendant
TT	tetanus toxoid
USAID	U.S. Agency for International Development
VA	verbal autopsy
VAI	verbal autopsy interviewer
WHO	World Health Organization

## SAVVY MANUAL SERIES

This series of SAVVY mortality surveillance system manuals, guides, and other documents is available at the MEASURE Evaluation Web site at:

<http://www.cpc.unc.edu/measure/leadership/savvy.html>

*Sample Vital Registration with Verbal Autopsy  
(SAVVY): An Overview*

### **Central office manuals:**

*Data Processing Manager's Manual*, including SAVVY Data System software

*SAVVY Budget Manual*, including SAVVY Budget Template software

*Verbal Autopsy Certifier and Coder's Manual*

### **Field office manuals:**

*Field Office Manager's Manual*

*Census Interviewer's Manual*

*Census Supervisor Coordinator's Manual*

*Census Supervisor's Manual*

*Census Update Interviewer's Manual*

*Key Informant's Manual*

*Verbal Autopsy Interviewer's Manual*

*Verbal Autopsy Supervisor's Manual*

### **Training guides and materials:**

*Census Interviewer Training Guide*

*Census Interviewer's Workbook*

*Census Supervisor Training Guide*

*Census Update Interviewer Training Guide*

*Census Update Interviewer's Workbook*

*Key Informant Training Guide*

*Verbal Autopsy Interviewer Training Guide*

*Verbal Autopsy Supervisor Training Guide*

SAVVY methods for verbal autopsy (including forms, certification, and cause of death assignment and coding) have been developed in collaboration with the World Health Organization (WHO). The WHO publication *Verbal Autopsy Standards: Ascertaining and Attributing Cause of Death* is an essential resource for the application of SAVVY methods.

SAVVY stands for “**S**Ample **V**ital registration with **V**erbal autops**Y**.” SAVVY is a library of best practice methods for improving the quality of vital statistics where high coverage of civil registration and/or good cause of death data are not available. SAVVY is not a substitute for universal civil registration. Its components can, however, fill short- to medium-term needs for critical information on births, deaths, and cause of death at the population level.

SAVVY can be implemented in many ways. It is not always necessary to implement SAVVY as a complete system. There are many circumstances in which various SAVVY methods might be effectively implemented. One option is to adapt specific modules and manuals in order to attain rapid cause-specific mortality estimates as part of a stand-alone data collection exercise, such as a survey. Another potential use of SAVVY methodologies is to augment existing facility-based or administrative data sources. SAVVY methods are integrated into the Health Metrics Network’s vision of “stepping stones” to better vital events monitoring, and are entirely harmonized to the World Health Organization (WHO) International Classification of Diseases (ICD).

The remainder of this chapter will describe the structure of a complete SAVVY system, as part of a long-term strategy to attain universal vital registration and the proper medical certification of deaths. When implemented as a complete system, SAVVY can serve to provide nationally representative information about levels and causes of death, as well as generate many other socio-demographic indicators.

SAVVY includes resources to implement the following:

- ❖ **Demographic surveillance system (DSS)** — DSS is a complete and continuous enumeration of births, deaths, and migration in a geographically defined population.
- ❖ **Mortality surveillance system (MSS)** — MSS consists of the active reporting of deaths in a geographically defined population. Verbal autopsy (VA) interviews are used to determine

the probable causes of death.

- ✦ **Death certification and ICD coding** — Death certification and ICD coding involves application of the tenth revision of ICD and WHO-approved procedures to certify deaths from verbal autopsy interviews and assign a probable cause of death.
- ✦ **Nested surveys** — Nested surveys consist of focused sets of questions and are included in the census update rounds. Examples include surveys on poverty monitoring, reproductive health, health service coverage, and environmental and behavioral risk factors.

In establishing a full SAVVY system, the first step is to select and define representative sample areas. Then a complete baseline census is conducted of all households and residents in those areas. The census information on the residents of each sample area is updated annually.

Following the baseline census and continuously thereafter, a local key informant (KI) notifies a verbal autopsy interviewer (VAI) of all deaths occurring in the KI's assigned area. The VAI then conducts a verbal autopsy interview at the household where that death occurred. This information is used later to determine the most likely cause of death.

The SAVVY system field operational structure will depend on the sample selection, as well as the availability of local resources and logistical support. The diagram on page 5 provides a general outline of the field operations in a prototypical SAVVY system.

Most of the people who work to implement SAVVY are selected with community input and participation. Ensuring the success and sustainability of SAVVY, and the use of the information it generates, depends upon fostering participation and ownership from the community to the national levels.

## DEMOGRAPHIC SURVEILLANCE

Demographic surveillance is designed to collect accurate demographic data for the SAVVY system. The initial step is to conduct a baseline census for the areas within the SAVVY system. During the census, information is collected on age, sex, marital status, and educational attainment for people living within the SAVVY area. Information is also collected about family structures and other socioeconomic characteristics. Typically, interviewers who live within

the community collect the census data and the data are updated annually.

The data collected in the SAVVY censuses are used to calculate many indicators. The census data provide accurate population denominators for calculating rates (e.g., mortality rates). The census data also provide a sampling frame for supplementary surveys that are nested within the system.

Mortality surveillance is designed to produce accurate community-based information on the levels and causes of death. Mortality surveillance involves continuously and actively identifying all deaths that take place in sample areas soon after they happen. Verbal autopsy interviewers visit households where deaths have occurred and interview relatives or caregivers using the SAVVY VA forms. Once completed, these forms are used to determine probable causes of death.

The data collected by SAVVY mortality surveillance techniques are used to calculate many indicators. Among the most important are mortality by age, sex, and specific causes of death. This information can be used at many levels of the health system for planning, reporting, monitoring, evaluating, and priority setting.

The information collected through SAVVY is generally not available from any other source or on an annual basis. Complete coverage and participation by all communities and the individuals residing in sample areas is extremely important.

Verbal autopsy, or VA, is a questionnaire administered to caregivers or family members of deceased persons to elicit signs and symptoms and their durations, and other pertinent information about the decedent in the period before death. SAVVY uses international standards for verbal autopsy forms, death certification and ICD coding procedures developed in collaboration with WHO, the Health Metrics Network, and other stakeholders. Separate verbal autopsy forms are used for the following age groups:

- ❖ perinatal and neonatal mortality (death of a child under 4 weeks);
- ❖ post-neonatal child mortality (age 4 weeks to 14 years); and
- ❖ adult mortality (age 15 years and over).

## MORTALITY SURVEILLANCE WITH VERBAL AUTOPSY

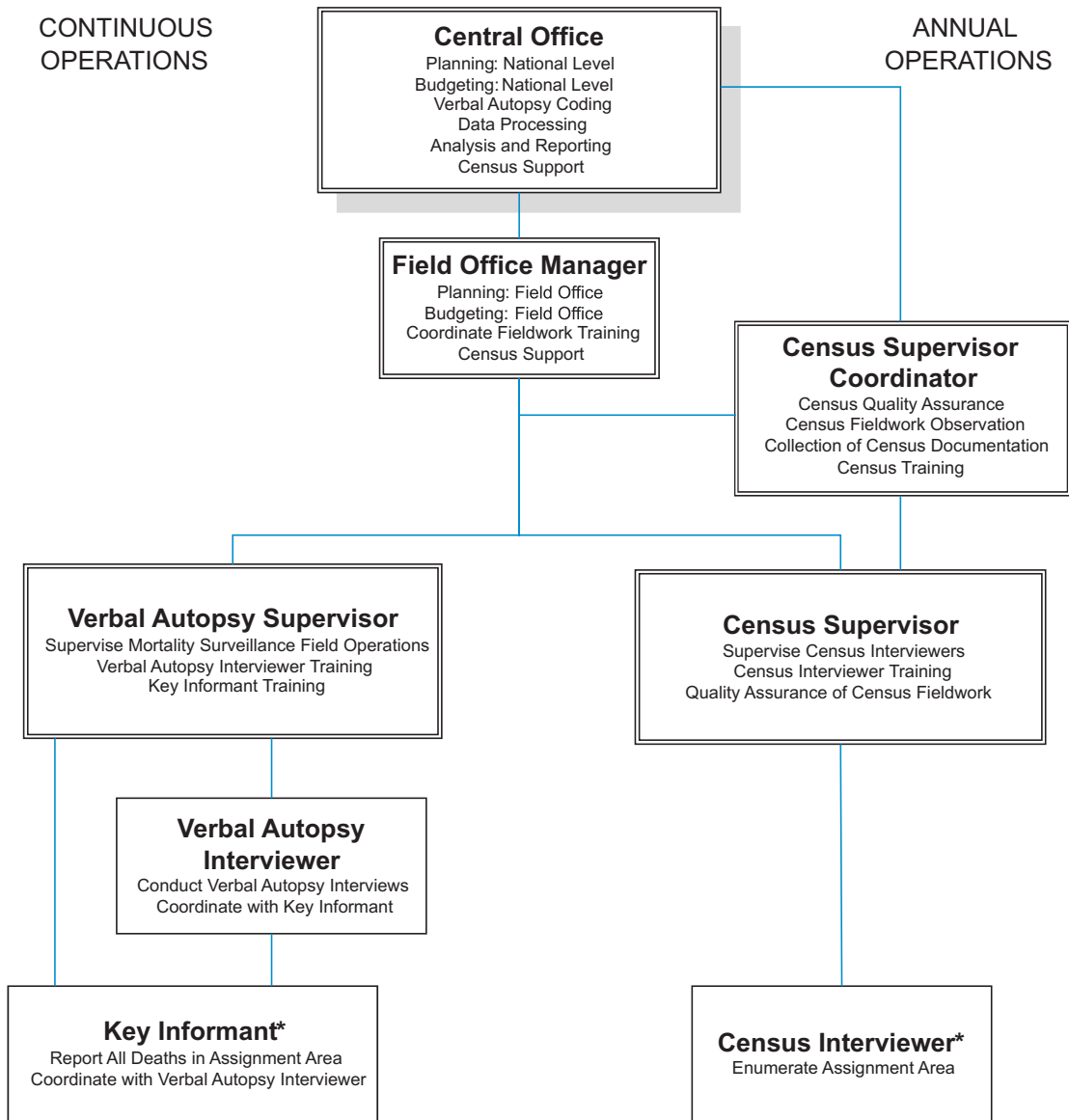
## WHAT IS VERBAL AUTOPSY?

All forms used to conduct these interviews include an open narrative section and a structured symptom duration checklist. There are also questions about the health and status of mothers (in the case of perinatal, neonatal, and child deaths), and questions that specifically relate to all women (15 years of age or older). Additional information is collected about previously diagnosed conditions, medications used, health services used, place of death, and behavioral and environmental risk factors. After the administration of the VA interview, a panel of physicians reviews the forms and assigns a probable cause of death using a method that conforms to international convention.

The purpose of VA is to describe the cause structure of mortality at the community or population level where no better alternative sources exist. VA is not intended to diagnose cause of death at the individual level. While VA has some serious limitations, the shortcomings of the tool are known and quantifiable. These deficiencies, however, should not prevent countries requiring information on causes of death from benefiting from the use of VA when no practical alternative for obtaining these data exist.

The SAVVY Resource Kit has been completely harmonized with WHO recommended tools and procedures. It is understood that there will be a need to carry out a small degree of local adaptation to these materials in each country setting in which they are applied.

Certain applications of verbal autopsy may benefit from a shortened or condensed version of the standard WHO VA forms, for example to reduce the amount of paper required. An alternative layout, containing the same questions and content as the standard WHO VA forms, has been designed. Each of these forms fits onto 4 sides of A4 paper. These forms are available upon request.



\*If possible, these roles should be filled by a single individual.

## ABOUT THIS MANUAL

This manual provides both an overview and a step-by-step discussion of the SAVVY data processing system. There are three main data processing applications needed to support SAVVY. Each of these three focus on one of the three components of the SAVVY system: baseline census, verbal autopsy, and census update.

The manual refers to this set of applications as the SAVVY data processing system. The SAVVY data processing system has been implemented using the Census and Survey Processing System (CSPro) software package, developed and supported by the U.S. Census Bureau.

# Overview of Data Processing Systems

Data processing for the Sample Vital Registration with Verbal Autopsy (SAVVY) system is much more complicated than that of a general population census. In fact, it is an ongoing survey taken periodically for randomly selected areas. After the initial baseline census, there must be matching and updating of the original information based on the census update round. It is important that the SAVVY information be current and accurate in order to produce the most reliable statistics.

If possible, all SAVVY data processing should be at a single central site since the volume of data is low and the need for both operational and quality control is high. Multiple processing centers are usually not recommended.

The processing system has the usual checks for data consistency associated with demographic surveys; e.g., data for males should not include pregnancy information. In addition to this normal processing, however, there is also a critical need for accurate household identification since the SAVVY system is based on matching and updating household information from different sources. In order to allow the original information to be updated throughout the duration of the project, a master file of household information is created following the baseline census. Then, as new household information is collected in census update rounds, this master file is updated to reflect changes. To update the master file with new information, an archive file of the master file is created, then the new and original data are matched based on the household and personal identification. This matching requires that household and personal identification information must be accurate.

The SAVVY implementation partners have used their experience to prepare easily implemented applications for each of the data processing steps. Each of the three main applications has some common tasks. These are as follows:

- ✘ enter data (key)
- ✘ verify data (re-key; this allows for the high degree of accuracy needed)
- ✘ check of essential information and interactive

correction of errors

- ✘ check of less essential information and computer correction of errors
- ✘ update the SAVVY master file using the keyed data.

The step-by-step processing for each application is discussed in the following sections. Each application outputs a single data file and any required publication tabulations. A final module may be added that joins all these data files together by matching on control information. The resulting large file can be used for more thorough analysis. Appendices A through D provide the detailed programs for processing each application.

This manual assumes that the office responsible for SAVVY data processing has at least one person qualified to perform the listed activities. This person can be anyone assigned to the project, regardless of title. The only requirement is that the person be familiar with CPro software. If no one with knowledge of CPro is available, then training should be arranged.

It is also assumed that the processing group includes people with experience in those aspects of the system related to routine production tasks (coding, keying, etc.). Consequently, these activities are not covered in this manual.

# Roles and Responsibilities of the SAVVY Data Processing Manager

The data processing for the Sample Vital Registration with Verbal Autopsy (SAVVY) system will be done at the central office level and the data processing manager is an essential part of the central office team. The duration of the appointment of the SAVVY data processing manager will depend on the operational structure of the SAVVY system in the country of implementation. Levels of authority and hiring will also vary. The central office will hire and train the data processing manager.

The role of the data processing manager is to manage the collection, storage, entry, verification, and editing of SAVVY census and verbal autopsy forms. The data processing manager should be familiar with CPro and the SAVVY data processing system. He or she should understand the concepts behind two-processor verification, electronic check-in of forms, and the use of interactive and batch edits.

The data processing staff will vary throughout the SAVVY system implementation, with greatly increased staffing during peak workloads of the baseline and census update rounds. The data processing manager will keep the central office informed of staffing and hiring needs.

The data processing manager will assure that all procedures established for tracking, keying, and verification are being followed. The data processing manager will also supervise or serve as a liaison for the various coding operations for verbal autopsy (VA) and the International Statistical Classification of Diseases and Related Health Problems (ICD). These procedures will be discussed below.

In ensuring that the data are managed properly and entered in a timely and accurate manner the data processing manager has the following responsibilities:

- ✦ Keep the central office informed of staffing and hiring needs. The number of data processing staff will vary throughout the SAVVY system implementation, with greatly increased staff during peak workloads of the baseline and cen-

## RESPONSIBILITIES

sus update rounds.

- ✘ Assure that all procedures established for tracking, keying and verification are being followed.
- ✘ Supervise or serve as a liaison for the various VA and ICD coding operations.
- ✘ Supervise and assist the data processing team.
- ✘ Assure that data processing is timely and accurate.
- ✘ Assure that proper procedures are followed in keying and verifying the data.
- ✘ Keep the central office apprised of any delays in the data processing or of any known issues in keying accuracy.
- ✘ Manage the production of the census update forms.
- ✘ Maintain regular backups of data according to an agreed backup plan.

## TASKS

The data processing manager will perform the following tasks at the following intervals:

**Before a baseline census** — Plan the data entry human resource needs and plan structural and consistency edits.

**Periodically during census data entry** — Monitor the rate of data entry, make local and off-site backups of the data entered, and receive and store forms from the field sites.

**Once census data entry has been completed** — Archive the completed, unedited census round; and work with survey statistician to create required edits.

**Before a census update round** — Prepare the address listing books, print the census forms, and plan structural and consistency edits.

**Each week** — Monitor the number of VA forms received from each site, monitor the rate of VA data entry, and backup the VA data.

**Each quarter** — Produce an official release of VA data for analysis purposes.

The data processing manager and staff will have access to the SAVVY questionnaires containing confidential information. You must not reveal SAVVY census information to anyone not working with the SAVVY system. Never reveal any information about a particular person or household except to your supervisor. This rule encourages respondents to participate in the SAVVY census because they understand that their personal information will remain confidential.

## CONFIDENTIALITY

You must not accept falsified work under any circumstances. You must make sure that the data processors do not submit falsified work to you and you do not submit falsified work to the central office under any circumstances. The data collected will be of no value if they are not accurate. If you suspect any data submitted have been falsified, you must act upon this assumption immediately. If any data submitted has been falsified, a data processor may be dismissed from his or her job. Similarly, if any of the data that you have submitted is falsified, you may be dismissed from your job.

## HONESTY



## BASELINE CENSUS QUESTIONNAIRES TRACKING

Questionnaires are received by the central office in assignment area bundles. The assignment area (AA) is the unit used for tracking and control of address listing books and forms, both in the data processing unit and in the field. The number of households per AA will vary by country as well as within a country, but typically an AA has an average of about 100 households, corresponding to the workload of one interviewer during the census. Since it is likely that the baseline census will take some time, it is imperative to have a system to ensure that forms are not lost or overlooked.

When an AA bundle is received at the central office, the data processor will use the program **BC0 AA Tracker** in the Sample Vital Registration with Verbal Autopsy (SAVVY) data processing system for check-in and control.

After opening the file, the data processor will key in the cluster number and assignment area from the cover of the address listing book on the top of the AA bundle. The program will provide a drop-down menu from which the data processor selects one of the following actions:

- 1 check in
- 2 sent for coding
- 3 returned from coding
- 4 sent to data entry [and verification]
- 5 returned from data entry
- 6 sent for structure edit [manual correction of essential problems]
- 7 returned from structure edit
- 9 sent to storage [processing completed]
- 0 STOP this assignment area

For initial check-in, the data processor will simply choose number 1. The tracking program will then ask the data processor to ac-

cept the current date as the check-in date. The data processor selects “Yes” and then is asked if he or she has finished the assignment area. If there are no other actions being taken with this assignment area, the coder selects “Yes” and moves on to the next AA bundle.

If the coder desires to enter the date manually rather than accept the current day’s date, this is possible. However, the manually-coded date must be no later than the current date.

If there are further actions to be taken with the AA, the coder selects “No” when prompted. The coder can then select “Sent for Coding” or other actions in the menu.

The initial check-in must be entered in “Add Case” mode, but any of the remaining actions may be taken by simply switching to “Modify Case” mode, and selecting the assignment area of interest. For example, say cluster 01, assignment area 01 will be sent to coding today. Then the data processor simply moves to “Modify Case” mode, picks file 0101 and clicks on the “Sent for Coding” boxes on the screen. Similar prompts as those discussed above guide the data processor in finishing this AA action.

The forms tracking program has built-in logic that does not allow data processors to pick an action out of sequence. A user could not, for example, select “Return from Coding” before selecting “Sent for Coding.” A user can always pick “0 STOP this assignment area” from the drop down menu.

As assignment areas are checked in and modified, the master geographic file (of clusters and assignment areas in the SAVVY census) is modified using the status code. This is the number that corresponds to the action taken; for example, check-in is status code 1, and sent for coding is status code 2. The geographic master file is, among other uses, used in verifying data as it is being entered.

## ENTER AND VERIFY DATA

Typically, an AA bundle of forms (a batch) would be entered by one keyer and verified by another. The task here is to capture the information on the form as accurately as possible. Neither initial entry nor verification should involve any attempt to correct any perceived errors on the form. Cluster and AA values are checked against the master geographic file. If cluster and AA values on the form are not valid, the data processor should notify the supervisor immediately.

To enter the questionnaire data, the data processor will use the program **BC1 Enter Data** in the SAVVY data processing system. The data entry screen looks similar to the form itself. The data

processor moves from field to field as if reading through the form. The data processor must enter a complete ID from section A of the SAVVY census questionnaire to proceed. The complete ID is required. Other values may be skipped, but will trigger a warning flag from the program. All data on the form are entered. There are some checks within the data entry program that occur before verification or the structural edit, such as disallowing fertility information for males.

Once the data are keyed, a second data processor can verify the information. The second processor will select the “Verify Case” mode in the menu. He or she then re-enters the data. The program will stop the data entry if there are any inconsistencies, allowing the processor to confirm that the data entered match the data on the form. The error is resolved when the verifier keys in either (a) the same value as the first data processor or (b) a different value twice in a row. Whatever is accepted or changed by the second data processor (verifier) is what goes into the final file.

Data must be entered and verified (100% or at agreed on rates for items and forms). CSPro allows the user to select items that must be verified by selecting the properties of a given field. Additionally, using the Data Entry Options dialogue box, the user can select the frequency of data verification — every  $n^{th}$  case, for example. Typically, the SAVVY census begins with 100% verification, but this verification is reduced over time as accuracy and common errors begin to be known.

It is also possible to verify the data on a batch basis through file matching.

The structure edit is a basic edit that ensures that an AA file contains the right numbers of the right type of records, that within a given household persons are not missing or duplicated, and that critical data items have been entered correctly.

The structure edit program, run from a desktop icon, is probably the most time-consuming step in processing the baseline census. This could be done after data entry and verification but does require the use of the interactive facility in CSPro.

The data processor or subject matter expert will use the **BC2 Structure** program in the SAVVY data processing system. Selecting the “Modify Case” mode, the user presses F11 to start the interactive edit. The program will stop when it finds a data inconsistency or error and an error message will pop up on the screen. The data

## STRUCTURE EDIT

processor can press F8 to clear the pop-up box, or simply click on the screen.

Now the user has two options. If it is clear how the error should be resolved, the data processor or subject matter expert working on the record should make the correction on the screen and then continue interactively editing the record by pressing F11. If, however, it is unclear as to how to resolve the error, the data processor may move the cursor forward into the next cell and continue interactively editing. These problems will be resolved later in the consistency edit by imputation.

Some values on questionnaires that are tested are essential for accurate reporting. Problems involving these values must be resolved by the data processor or subject area expert using the original questionnaire and local knowledge.

The data processor or subject matter expert will continue interactively editing until the end of the case is reached. The program will automatically bring forward the next case until all cases are completed. If the user wishes to discontinue editing while there are cases remaining in the assignment area, simply pressing the escape [Esc] key will stop the program.

When the structure edit is completed for an entire assignment area, questionnaires will then be returned to the batch tracking center for storage.

## FILE CONCATENATION

Data files from all of the assignment areas must be joined together into a single unedited data file for further processing and raw data must be archived. The single unedited file can be created on a flow basis, that is, if we use a phased-in baseline census then we add the AA's in as we complete them.

A single data processor may use the concatenate data tool in the general CSPro software package on the desktop to create the single unedited file. As successive AA's are complete, the processor need only concatenate the newer files to the single combined file using the same tools and technique. However, since this is not a true append operation, the new combined file requires a new name.

The final output is a structurally correct data file sorted by assignment area. From this data file, two steps may be taken. First, the data processing team can create frequency distributions and tabulations to isolate and magnify any systematic data collection or processing errors that need to be resolved. Resolution of the errors may be written into the consistency edit program, if necessary, or the

consistency edit program can be run while the pre-edit tabulations are being created and analyzed. The second step is the running of the consistency edit program. As mentioned above, this may be done either simultaneously with creating the pre-edit tabulations or as a subsequent step. This process is discussed below.

The final unedited data file is checked for a set of non-critical errors. If found, these are corrected automatically via program logic, using one of several possible imputation methods.

A single data processor should use the **BC3 Final Edit** program to run the consistency edit. When the program is run, the final output will be an edited data file. The consistency edit program will also create a complete list of errors and how they are resolved for each case edited. This output is found in the BC3 Final Edit.lst file.

As noted, part of the error resolution process in the consistency edit may consist of imputing data. A table giving the types and frequencies of data imputations is also generated when the consistency edit is run. This table is called BC3 Final Edit.freq.lst file.

The final error listing generated by the consistency edit program is a listing of missing HU's. The program determines if there are any gaps or duplicates in the sequential listings of buildings, housing units or households and outputs them to the BC3 Missing HU Report.wrt file.

If data need to be weighted it should be done at this step in the process. Adding weights would involve writing a simple batch edit which could weight data on the household or person record. If weights are few, then the weights can be written into the batch edit. If the weighting schema is more complex, weighting data may require a look-up table, also called a parameter list.

The SAVVY implementation partners (U.S. Census Bureau and MEASURE Evaluation) stand ready to provide assistance in this area.

The SAVVY data processing system provides a set of tabulations and cross tabulations. For an illustrative list of the tables available, see Appendix E.

In order to generate the tables, the data processor simply opens up the **BC4 Tables** program in the SAVVY data processing system

## CONSISTENCY EDITS

## WEIGHTING DATA

## TABULATIONS

and selects the table of interest. The data processor selects the green light to run the program. A dialogue box will pop-up to ask the data processor what data file to use to populate the tables. Once the data file is selected, the tables are tabulated and populated automatically.

Additional tabulations and cross tabulations can be created using the CSPro generic software package. As a final step, all the data files and programs used in the baseline census need to be documented and saved for future reference and for possible use in future analysis.

The verbal autopsy (VA) component of the Sample Vital Registration with Verbal Autopsy (SAVVY) system requires accurate data tracking, keying, and verification. VA uses two-way independent coding of VA questionnaires to produce reference death certificates; verification and reconciliation of the death certificates; and ACME/TRANSAX software to ensure that the cause of death is properly coded.

The VA information is probably the most essential to running a successful SAVVY system. The information collected on the verbal autopsy questionnaires are stored in a separate data file. Each form must be processed as accurately as possible so that the information may be linked to final causes of death and International Classification of Diseases (ICD) codes files that are generated later in the process.

Data processing for the verbal autopsy questionnaires includes the following tasks:

- ✘ check in and control each verbal autopsy form
- ✘ enter and verify data from each verbal autopsy form
- ✘ checkout of VA forms to coding/certification
- ✘ check in forms from coding/certification
- ✘ key information from reference death certificates
- ✘ verify data keyed from reference death certificates
- ✘ assign weights to verbal autopsy cases if required
- ✘ tabulate the results
- ✘ archive the data.

## VERBAL AUTOPSY CENSUS QUESTIONNAIRES TRACKING

Verbal autopsy questionnaires are received by the central office on a flow basis. The check-in and control of VA questionnaires is needed to determine if forms are returned from each of the census assignment areas and that the geographic and administrative codes applied to the VA forms are valid. Tracking the forms in the control system will also ensure that the forms are made available to subsequent data processing tasks, and will reduce the likelihood that they will be lost or misplaced. It is recommended that a single person be responsible for check-in and tracking of VA forms.

Check-in and control of VA questionnaires must be done for each form individually. The control system will monitor and track the processing of each VA questionnaire from check-in through data entry, matching, checkout to ICD coders, check-in from ICD coders, and adding of ICD and SAVVY death codes to the VA data file.

When a verbal autopsy questionnaire is received, it is checked in using the **P1 VA Tracker** program in the SAVVY data processing system. The data processor begins the check-in by entering the cluster number, assignment area, building, housing unit, and household number. At this point, the program generates a four-digit serial number for the VA questionnaire — this is the “Person Number.” It is essential that the data processor write this serial number onto the VA questionnaire. A pop-up box will appear on the screen to remind the data processor of this task. Once the serial number is written on the questionnaire, the data processor uses the F8 key to clear the box.

The program will then ask the user to accept the case. If it is accepted by the user, it is considered to be checked in on the current date. The program automatically assigns a status code of 1, i.e. a check-in, and populates the check-in date. The program will then be ready to add the next questionnaire. The data processor can manually stop the program at any time.

Throughout the VA questionnaire processing, the status of the questionnaire may change several times. A brief list of the status codes follows, and will be discussed in detail.

Using DE for “data entry,” the status codes are:

- 1 check in
- 2 sent for DE
- 3 returned from DE
- 4 sent for coding and certification

- 5 returned from coding and certification
- 6 sent to ICD DE
- 7 reconciliation needed
- 8 sent to final ICD DE
- 9 processing completed

Codes are subject to change at implementation, and codes 7 and 8 are optional.

Each VA form must be entered using the SAVVY VA data entry program. Depending on volume of questionnaires, a group of questionnaires can be entered periodically or directly after check-in.

The data processor begins keying the form by opening the **P2 VA Enter Form** program in the SAVVY data processing system. It should be noted that there are three different questionnaires, but only one data entry application. The user will tell the program which form is being used.

The program will default to the “Modify Case” mode upon entry. The user can select any checked-in form to key. Once a form is selected, the program automatically populates the identification information for that questionnaire, including the person number. The first field that the data processor should fill in should be the form type.

The data entry screen looks similar to the questionnaire itself. The data processor moves from field to field as if reading through the questionnaire. All data on the questionnaire are entered. Once the data are keyed, a second data processor can verify the information. The second processor will select the “Verify Case” mode in the menu. He or she then re-enters the data. The program will stop the data entry if there are any inconsistencies, allowing the processor to confirm that the data entered matches the data on the questionnaire or to correct it. Whatever is accepted or changed by the second data processor (verifier) is what goes into the final file.

Data must be entered and verified at the 100% level to ensure the accuracy of the VA questionnaire coding.

The output of this step is a file of SAVVY VA questionnaire data and identification information for each death event recorded.

## KEYING AND VERIFICATION OF VA FORMS

## ASSIGN AND CHECKOUT VA FORMS FOR ICD CODING

The assignment of forms for ICD coding is not part of the data processing per se. Assignments will be made at the supervisory level. The check-in and tracking program, however, can be used to monitor the location and status of the VA questionnaires. The data processors can simply open the **P1 VA Tracker** program in the “Modify Case” mode. Assign status “4 — sent for coding and certification.”

## TWO-WAY INDEPENDENT ICD CODING OF VA FORMS AND VERIFICATION

Each VA questionnaire will be assigned to two ICD coders for independent coding. The ICD coders should be randomly assigned in order to minimize any biases in coding. The codes are assigned independently by each ICD coder. Each ICD coder fills out a SAVVY VA death certificate listing the identification information, three causes of death both in text form and as four-digit ICD codes. The death certificate also contains the ICD coder’s name and the date of coding.

Once both death certificates are complete, the keying and reconciliation operation can begin. It is important to understand that each death will result in the keying of death certificates at least four times. For each death, there will be two independently coded death certificates, and each of these must be coded twice for 100% verification.

To begin the process, the first data processor uses the **P3 ICD Coding** program. The data processor keys in both independent death certificates that are associated with a single verbal autopsy questionnaire, one directly after the other. Once both death certificates have been coded, the program will ask the data processor if this case should be accepted. This is simply a fail-safe mechanism; it allows the data processor to go back and make changes if required. This mechanism can be turned off and cases will simply be saved automatically when the last field is populated. Once the case has been accepted, the verification process can begin.

To verify, a second data processor will enter the program in the “Verify Case” mode and select the case of interest. The data processor will key both the death certificates again. The ICD coding requires 100% verification of keying, so if any of the fields do not match, a dialogue box will appear asking the data processor to resolve the inconsistency. Whatever is accepted or changed by the second data processor (verifier) is what goes into the final file.

Once both death certificates have been verified for keying inconsistencies, i.e., once the second data processor has completed the last line of the second death certificate, the program automatically

checks for consistency between the two death certificates. A dialogue box will pop up with one of the following messages.

1. These two forms do not need resolution. COMPLETED.
2. These two forms do need resolution. PLEASE SET ASIDE.

If all the ICD codes match, then the record is written to a final cause of death (COD) data file. The VA questionnaire and the SAVVY death certificates can then be sent for check-in and archiving.

If the codes assigned independently disagree, the VA forms and death certificates should be sent forward for reconciliation.

The cases that require reconciliation will be sent to a third independent ICD coder for adjudication. For each item the ICD code that is assigned by the majority of coders is assumed to be the correct code, so that each item in the VA form requiring an ICD code will have one and only one ICD code entered in the database. A third SAVVY death certificate will be created with only one set of data. These data must be keyed and verified using the **P4 ICD Resolution** program.

With the final, adjudicated SAVVY death certificate, the data processor will open up the **P4 ICD Resolution** program in the SAVVY data processing system and select “Modify Case” Mode. The data from the adjudicated death certificate is keyed and verified using two independent data processors for entry and verification, as previously discussed in this manual.

Using a parameter file, the output from the ACME/TRANSAX software is coded to a SAVVY cause of death. This is one of about 57 different causes of death at a much less detailed level than the ICD coding. There is no specific tool created to implement this task, but it may be done using the CPro software. The SAVVY implementation partners (U.S. Census Bureau and MEASURE Evaluation) stand ready to provide assistance in this area.

The output from this step is a data file with identification information, the ICD cause of death codes, and the SAVVY cause of death code.

The next step is to merge the output from the ICD and SAVVY cause of death coding with the file of VA questionnaire data that was previously created. Data are merged using all control information to create a final analysis file.

## ICD CODE RECONCILIATION

## CODE ACME/TRANSAX ICD CODE TO SAVVY CAUSE OF DEATH CODE

## MERGE WITH VA FILE AND CREATE FINAL ANALYSIS FILE

Finally, all the data files and programs used for the verbal autopsy processing need to be documented and saved for future reference and for possible use in future analysis. It is important that everything used to process the data (data files, programs, and modules) be well documented, because people using the data for analysis and reporting will probably not be the same people who created the data.

The census update will use the records from the previous census, so the first operation will be to prepare the base listing for each of the assignment areas. For the census update, the assignment area (AA) once again becomes the unit used for tracking and control. The base listing will contain identification codes and names of businesses, institutions, and vacant housing units, and the name of the head of household for each household previously enumerated.

Questionnaires and address listing books for the census update must be printed using a computer program designed for the task so that information collected during the previous census will appear on the forms where needed.

Each country may implement this in a slightly different manner, depending on the capacity of the central office and current standards. CSPPro can be used to output the base listing for assignment areas in any of a number of formats.

This printing operation must be well planned and accurately carried out so that the questionnaires and address listing books are ready when the census update round begins.

The SAVVY system is longitudinal in nature and it is important to identify when individuals entered the SAVVY system. Tracking when information is first collected for an individual is accomplished through the line number on the three census questionnaires (in column 1, "Line #"). In the baseline census, each individual household member is assigned a two-digit line number on the questionnaire, starting with "01" and continuing to "10." If a household has more than 10 members, a second questionnaire is used, with the line number starting with "11" and continuing to "20."

For the census update, a three-digit line number is used on the questionnaires. In section G, the pre-printed census update questionnaire already has the names of individuals first collected in the baseline census. Each of these individuals is identified with a three-digit number that begins with "0" and continues with the same two-digit

PREVIOUS CENSUS  
DATA FILES

PRINT CENSUS  
UPDATE MATERIALS  
BY ASSIGNMENT  
AREAS

number that had been used for that person on the baseline form. In section H of the update questionnaire, where any new household members since the baseline census are entered, the “Line #” column begins each three-digit number with a pre-printed “1.” Information on these individuals is being collected initially in the census update. The census change questionnaire also uses a pre-printed “1” to begin the three-digit numbers in section G since these households and their members are also entering the SAVVY system for the first time, during the first census update round. For a second update, a “0” will begin the three-digit numbers for all individuals who entered the SAVVY system during the baseline year, a “1” will begin the three-digit numbers for anyone initially entered during the first update, and a “2” will begin the three-digit numbers for those who are being added during the second update.

## CENSUS UPDATE QUESTIONNAIRES TRACKING

The census update round uses the **UC0 AA Tracker** program in the Sample Vital Registration with Verbal Autopsy (SAVVY) data processing system for check-in and control of the census update forms. This tracker is similar to the system used by the baseline census.

## UPDATE EXISTING CENSUS RECORDS AND VERIFICATION

To update existing records from the pre-printed SAVVY update questionnaire, the data processor will use the program **UC1 Enter Data** in the SAVVY data processing system. The processor will key in the identification variables as shown on the first page of the census update questionnaire. The program will then pre-populate fields on the screen so that it looks identical to the pre-printed form received in the AA bundle.

The data processor moves from field to field as if reading through the form. All data on the form are entered.

Similar to the baseline census, the data must be verified. Once the data are keyed, a second processor is used to verify the information. As in the baseline census, the second processor will select the “Verify Case” mode in the menu and then enter the data independently. The program will issue a warning flag if there are inconsistencies and the last value entered is used.

Typically, the census update will begin with 100% data verification, with a reduction of the number of fields or frequency of forms verified as accuracy and common errors begin to be known.

Typically, the SAVVY change questionnaires from an AA would be entered by one data processor and verified by a second data processor.

## CREATE NEW CENSUS RECORDS

To create new records from the SAVVY change questionnaire, the data processor will use the program **UC1A Enter New Forms** in the SAVVY data processing system. The data entry screen looks similar to the form itself. The data processor moves from field to field as if reading through the form. The data processor must enter a complete ID from section A of the SAVVY change questionnaire to proceed. The complete ID is required. Other values may be skipped, but will trigger a warning flag from the program. All data on the form are entered. There are some checks within the data entry program that occur before verification or the structural edit, such as determining if the action code entered is valid for a new record and requiring a new household status code.

Once the data are keyed, a second data processor can verify the information. The second processor will select the “Verify Case” mode in the menu. He or she then re-enters the data. The program will stop the data entry if there are any inconsistencies, allowing the processor to verify the data. Whatever is accepted or changed by the second data processor (verifier) is what goes into the final file.

Again, the level of verification is flexible and can be enhanced or reduced over the life of the project.

The SAVVY data processing system has created a batch program to combine the data from the new and existing records.

## COMBINE NEW AND EXISTING RECORDS INTO A SINGLE FILE

A single data processor may use the **UC2 Combine** program in the SAVVY data processing system to create a single unedited file for each AA. The output of this program will be files of combined new and existing census data by assignment areas.

The structure edit is the basic edit that ensures that the AA file contains the right numbers of the right type of records, that within a given household persons are not missing or duplicated and that critical data items have been entered correctly.

## STRUCTURE EDIT

Like in the baseline census, the structure edit for the census update is an interactive application run from the desktop by data processors or subject matter experts. The structure edit for the census update is run using the **UC2 Structure** program in the SAVVY

data processing system. The data processor or subject matter expert will follow the same process as detailed in Chapter 4.

## FILE CONCATENATION

See File Concatenation  
Chapter 4, page 16

## CONSISTENCY EDITS

The final unedited data are checked for a set of noncritical errors using the batch program **UC3 Final Edit** in the SAVVY data processing system. See Consistency Edits in Chapter 4, page 17, for details.

## WEIGHTING DATA

If data need to be weighted, this should be done at this step in the process. Adding weights would involve writing a simple batch edit that could weight data on the household or person record. If weights are few, then the weights can be written into the batch edit. If the weighting schema is more complex, weighting data may require a look-up table, also called a parameter list.

The SAVVY implementation partners (U.S. Census Bureau and MEASURE Evaluation) stand ready to provide assistance in this area.

## TABULATIONS

The SAVVY data processing system provides a set of tabulations and cross tabulations. For an illustrative list of the tables available, see Appendix E.

In order to generate the tables, the data processor simply opens up the **UC4 Tables** program in the SAVVY data processing system and selects the table of interest. The data processor selects the green light to run the program. A dialogue box will pop-up to ask the data processor what data file to use to populate the tables. Once the data file is selected, the tables are tabulated automatically.

Additional tabulations and cross tabulations can be created using the CPro generic software package. As a final step, all the data files and programs used in the census update need to be documented and saved for future reference and for possible use in future analysis.

The information collected at various times in the Sample Vital Registration with Verbal Autopsy (SAVVY) system must be combined to provide input data files, for a more complete analysis of the results. This is the first step. The next step would be producing the base set of predefined tabulations specified by subject matter experts. An optional step after that would depend on the demand for the data in formats suitable for analysis using specialized software, e.g., SPSS. A final step that must be done is to document and store the data for future use.

Data processing in preparation for analysis includes the following tasks:

- ✘ Merge required information from baseline census, census update, and verbal autopsy.
- ✘ Produce tables to describe the combined findings.
- ✘ Produce files for further statistical analysis.
- ✘ Archive the data.

Here is further elaboration on the four tasks:

- ✘ Final analysis requires a combination of data sets collected from the censuses and the verbal autopsies. For example, to calculate the death rate by age and sex for various diseases requires data from the baseline census, census update, and the verbal autopsies. Selected data from these three datasets must be combined into a single dataset for tabulation and analysis. This can be accomplished by using a simple batch application in CPro.
- ✘ Once a combined dataset is created, tables may be produced according to the tabulation plan for the combined data using the CPro Cross Tabulation module.
- ✘ Some of the analysis may require the use of

other statistical analysis packages, such as SPSS, Stata, or SAS. Data from the combined data set can be exported to any of these packages using the CSPro “Export Data” tool.

- ✘ As a final step, it will be important to document the data files, programs, and procedures used in the analysis. This information will be critical to future users of the data who are not likely to know how the data were produced. Post-processing users of the analysis should be able to understand how the results were obtained.

**Address** — The physical location of a building or living quarters.

May be comprised of a building number and street name (e.g., 123 Anywhere Street). See “location description” in this glossary.

**Address listing book** — Used to record the addresses or location descriptions of all known buildings and living quarters in an assignment area. This book is prepared during each census round and is given to the key informant upon completion.

**Assignment area (AA)** — A geographic area established for data collection purposes. An assignment area usually represents the average workload for one census interviewer or one key informant.

**Assignment area map** — Shows the boundaries, features, and landmarks of an assignment area. This map is prepared during each census round and given to the key informant upon completion.

**Assignment area number** — Identifies a specific assignment area on SAVVY maps and forms.

**Callback** — A return visit to an address to complete the required information.

**Canvassing** — Systematically traveling all streets, roads, paths, etc., of each block in an assignment area to identify every place where people live or could live.

**Census** — An official, periodic count.

**Census interviewer** — Responsible for conducting baseline census interviews or update census interviews. He or she is attached to the SAVVY area and is knowledgeable about that assigned area. The census interviewer is also responsible for identifying the boundaries of the assignment area and canvassing the entire assignment area to determine the location of each building, housing unit, and household.

**Confidentiality** — A guarantee that the information respondents provide to a SAVVY employee and SAVVY office will not be revealed to others.

**Duration** — For all signs or symptoms that were not associated with a previously diagnosed condition, nor related to an injury, “duration” is defined as the period starting from the *appearance* of that particular sign or symptom to the cessation of that symptom, regardless of the presence of that sign or symptom at the time of death, and irrespective of whether the sign or symptom appeared intermittently. For example, if a woman began to have fever 10 days before death, but she ceased having fever two days before death, the duration of her fever would be eight days, even if she did not have fever for each and every one of those eight days.

**Head of household** — See “reference person” in this glossary.

**Household** — This is an arrangement in which one or more persons make common provisions for their own food or other essentials for living. These people may have a common budget, be related or unrelated, or a combination. There may be more than one household in a housing unit. In short, a household is defined as a group of people who “eat from the same pot.”

**Household number** — This is a number assigned by a SAVVY census interviewer to each household within a housing unit. The household identification number must be unique within the housing unit.

**Housing Unit** — A housing unit is a separate and independent place of abode intended for habitation by a single household, or one not intended for habitation but occupied as living quarters by a household at the time of the census.

**Key informant (KI)** — A person who lives in the SAVVY assignment area and is responsible for reporting any deaths that occur in her or his assignment area to the verbal autopsy interviewer. The key informant is also responsible for arranging the date and time of the verbal autopsy interview with each bereaved family.

**Location description** — A description of the physical location of a living quarters that tells anyone unfamiliar with the assignment area how to find that living quarters, so the living quarters can be located by another SAVVY employee. This may be an address, if one is available.

**Reference person** — The person who makes decisions for the household on a daily basis and who is a permanent resident of the household (spends the night at least six months out of the year at the house) is the “reference person.” If the household reports someone who spends the majority of his or her time away from

home, then make that person's spouse the reference person. If there is no spouse, then make the eldest family member the reference person, as long as he or she is at least 15 years of age or older. If there is no family member 15 years of age or older, then make the eldest nonfamily member the reference person as long as he or she is 15 years of age or older. If there is no permanent household member who is at least 15 years old, then make the eldest relative or resident the reference person.

**Resident** — See “usual residence” in this glossary.

**Respondent** — There are two types of respondents in the SAVVY system, the **census respondent** and **verbal autopsy respondent**.

*Census respondent* is the person supplying census information about a household and all of its members. The respondent should be a responsible family member of the household. Possible respondents are listed in order of preference:

- ☒ reference person
- ☒ spouse of reference person
- ☒ eldest family member available, at least 15 years of age or older
- ☒ nonfamily member at least 15 years of age or older
- ☒ family member less than 15 years of age
- ☒ nonfamily member less than 15 years of age
- ☒ neighboring reference person or spouse of neighboring reference person
- ☒ local knowledgeable informant

*Verbal autopsy respondent* is the adult being interviewed, who is typically a resident in the household. The respondent must be someone who is able to give reliable and accurate information regarding the members of the household. In the case of a death in the family, the respondent will be able to give information about the circumstances leading to the death. Ideally, the verbal autopsy respondent for the verbal autopsy interview would be the one who cared

for the deceased during the period of illness. Possible verbal autopsy respondents are listed in order of preference:

- ☒ main care-giver of the deceased in the period before death
- ☒ reference person
- ☒ spouse of reference person
- ☒ parents, particularly if the deceased was a child
- ☒ eldest family member available, at least 15 years of age or older
- ☒ nonfamily member at least 15 years of age or older
- ☒ family member less than 15 years of age
- ☒ nonfamily member less than 15 years of age
- ☒ neighboring reference person or spouse of reference person

**Sample area** — The geographic area selected for the SAVVY system.

**Stillbirth** — A baby that shows no signs of life when born. Stillbirths may be due to injuries, illness, infections, or catastrophic events happening to the mother or to the child while in the womb or during birth. In order to distinguish from abortion or miscarriage (for which verbal autopsy is not conducted), the SAVVY verbal autopsy system only includes births that occur after 28 weeks of pregnancy.

**Usual residence** — The place where the person lives and sleeps most of the time. This place is not necessarily the same as a person's legal residence. Also, noncitizens are included if this is their usual residence. Following are some common examples of usual residences:

- ☒ For people temporarily away on vacation or a business trip, their usual residence is the place where they live and sleep most of the time.
- ☒ The usual residence for commuter workers living away part of the week while working is the

residence where they stay most of the week.

- ✦ For children in joint custody, usual residence is where they live most of the time. If time is equally divided, they are counted where they are staying during the interview period.
- ✦ Usual residence for people who own more than one residence is the residence where they live most of the time.
- ✦ Usual residence for college students living away from home while attending college is where they are living at college, if they are in a household. College students living in dormitories are not counted in the SAVVY census.
- ✦ For college students living at their parental home while attending college, usual residence is their parental home.
- ✦ The usual residence for a live-in nanny or other live-in house worker is where that person lives most of the week.
- ✦ For foster children, usual residence is where they are living during the interview period.
- ✦ Usual residence for renters or boarders is where they are living during the interview period.
- ✦ Usual residence for housemates or roommates is where they are living during the interview period.
- ✦ For people in the military who are residing off-base in the country, their usual residence is the place where they live and sleep most of the time. Military people on-base are not counted in the SAVVY census unless they are in residential style housing.
- ✦ Usual residence for staff members living in hospitals, nursing homes, prisons, or other institutions is where they report that they live and sleep most of the time; otherwise it is the living quarters that they inhabit at the institution.
- ✦ For students living in school dormitories but who are not enrolled in college, the usual residence is their parental home.

- ✘ Usual residence for citizens of foreign countries who have established a household or are part of an established household in the country while working or studying, including family members who are with them, is the household where they are residing (spending the majority of their time while in-country).

**Verbal autopsy (VA)** — Verbal autopsy is a process used to collect information (using a specially-designed form) from relatives or caregivers of a deceased person. The process involves interviewing relatives or caregivers of the deceased regarding their knowledge of the symptoms, signs, and circumstances leading to death. The information that is collected is used by medical personnel to assign a probable cause of death for each reported death.

**Verbal autopsy form** — This is a form used to collect information on the history of illness of the deceased and presence of signs and symptoms. The form is to be completed by the interviewer during verbal autopsy interviews. There are three types of verbal autopsy forms used by the SAVVY system:

- ✘ International Verbal Autopsy Questionnaire 1:  
Death of Child under 4 Weeks
- ✘ International Verbal Autopsy Questionnaire 2:  
Death of a Child Aged 4 Weeks to 14 Years
- ✘ International Verbal Autopsy Questionnaire 3:  
Death of a Person Aged 15 Years and Above

**Verbal autopsy interviewer (VAI)** — The verbal autopsy interviewer is the person responsible for conducting VA interviews with the bereaved family members in the household. He or she is attached to the SAVVY area and is knowledgeable about that assigned area. The verbal autopsy interviewer must be accepted by the community in which he or she works. Some requirements of the verbal autopsy interviewer include having attained the highest primary level of the national education system (at minimum) and the ability to speak the dialect of the area to which he or she is assigned.

The following documents are included in the appendices:

- ✘ Appendix A: Baseline Census Processing Programs
- ✘ Appendix B: Production of Census Update Documents
- ✘ Appendix C: Census Update Processing
- ✘ Appendix D: Verbal Autopsy Processing
- ✘ Appendix E: Illustrative Tabulation and Cross Tabulations

## APPENDIX A: BASELINE CENSUS PROCESSING PROGRAMS

### Baseline Census (BC) Processing Programs

This is a summary of basic operations for entering, verifying, editing, and tabulating the information.

#### 1. Forms Tracking (BC0 AA Tracker.ent)

Assumes assignment area (AA) is unit for tracking information. It is likely that BC activities will take some time. It is imperative that forms not be lost or overlooked. This application is to keep track of the AA batches, e.g., groups of forms.

The status codes are (subject to change for an implementation):

- 1 Check in
- 2 Sent for coding
- 3 Returned from coding
- 4 Sent to data entry [and verification]
- 5 Returned from data entry
- 6 Sent for structure edit (manual correction of essential problems)
- 7 Returned from structure edit
- 9 Sent to storage (processing completed)

The ID items are: cluster number, assignment area, status, associated dates given as day, month, year.

This application could be run from desktop icon.

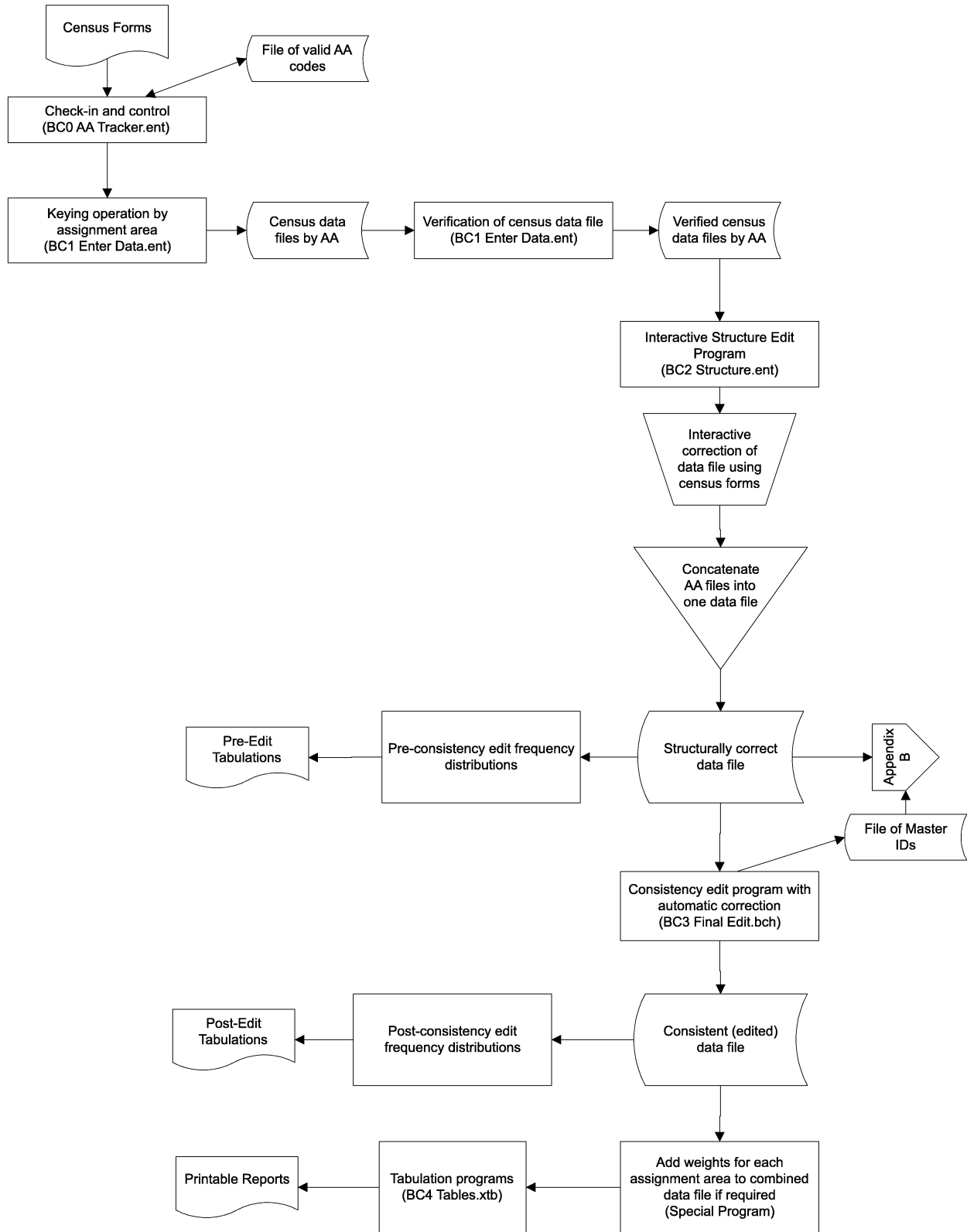
Program Name: **BC0 AA Tracker.ent**

Data File: BC AA tracker.dat (created/modified).

#### 2. Enter and Verify Data (BC1 Enter Data.ent)

A batch (complete enumeration results from one AA) would be entered by one keyer and verified by another. The task here is to capture the information on the form as accurately as possible. Neither entry nor verification should involve any attempt to correct perceived errors on the form. Cluster and AA values are checked against a master file.

Figure 9.1: Baseline Census Processing Flow Chart



A complete ID is required for each questionnaire. Data must be entered and verified (100% or at agreed on rates for items and forms). Forms would then be returned to “Batch Tracking Center” for further processing.

This application could be run from desktop icon.

Program Name: **BC1 Enter data.ent**

Data File: Cluster#\_AA#.dat (created/modified/verified)

ID check File: File of valid IDs at least to the AA level

Note: Tracker data file could be used as ID check file.

### **3. Structure Edit (BC2 Structure.ent)**

This is probably the most time consuming step. This could be done after entry, but requires the use of the interactive facility in CSPro. Some values on questionnaires are tested for consistency since they are essential for accurate reporting. Problems are resolved by a subject matter expert using the original form and local knowledge.

Program Name: **BC2 Structure.ent**

Data File: Cluster#\_AA#.dat (modified)

### **4. Concatenation of “Batch” Files**

Once all AAs have been entered, verified and completed the structure edit they will be concatenated into a single final (unedited) data file for further processing.

Program Tool: Concatenate data (CSPro)

Input Data File: ALL Cluster#\_AA#.dat

Output Data file: Final UNedited BC.dat (created), this file is kept for use with the census update processing — to create address listing and pre-printed questionnaires

### **5. Pre-edit Frequency Distributions**

Perform frequency distributions on all variables using “Actual value” option to determine the quality of the keyed data before the final edit.

Program Tool: Frequency Distribution Tool [CSPro]

Input Data file: Final UNedited BC.dat

Output: Pre-edit Frequency tables for all variables

## **6. Batch Edit (BC3 Final Edit.bch)**

The final unedited data file is checked for a set of not-so-critical problems. If found, these are corrected automatically via program logic.

NOTE: It would be possible to add recodes here that could be used in tables.

Input Data File: Final UNedited BC.dat

Output Data file: Final Edited BC.dat (created)

External File: Master ID file.dat (created from IDs and other information)

External File: BC3 Hot deck.dat (created and used in edit)

## **7. Post-edit Frequency Distributions**

This performs frequency distributions on all variables using the “Actual value” option to determine how the distribution compares to the pre-edit distributions. If distributions are changed significantly, then the edit program should be reviewed.

Program Tool: Frequency Distribution Tool [CSPro]

Input Data file: Final Edited BC.dat

Output: Post-edit Frequency tables for all variables.

## **8. Tabulations (BC4 Tables.xtb)**

This produces a set of tables based on edited data. Tables should be reviewed for accuracy and consistency. Compare distributions to post-edit frequency tables.

Program: **BC 4 Tables.xtb**

Input Data file: Final Edited BC.dat

Outputs: Tables printed and saved as a TBW [CSPro table viewer files]

APPENDIX B:  
PRODUCTION OF  
CENSUS UPDATE  
DOCUMENTS

Production of Census Update Documents

There are two products needed for a census update (UC). Each is produced in two phases. First, an intermediate tab-delimited file must be created as input to a “mail merge” Microsoft Word document that, in turn, will create the final product (i.e., either the AA address listing book or the AA census update form).

**Note:** Make sure the Master ID file is the most current. It needs to be updated after the completion of a census update. (See UC 5 Update Master IDs.bch)

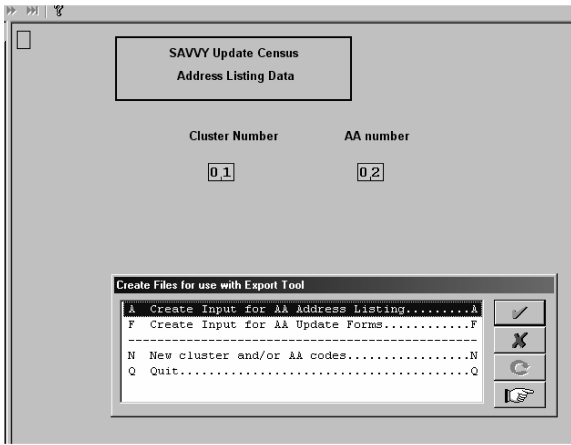


Figure 9.3: Generate UC files menu

1. Main Driver Program (Generate UC Files for Word.ent)

The production of both address listing books and census update forms are based on the cluster and AA codes. Products for each AA must be created separately. The main driver program, “Generate UC files for Word,” is a menu system where the cluster and AA codes are entered. Output files are created and used as input files for the CSPro Export Tool. The output of the Export Tool, a tab-delimited text file, is used to merge with the address listing Microsoft Word document or the census update form document. These produce the AA address listing book and the AA census update form.

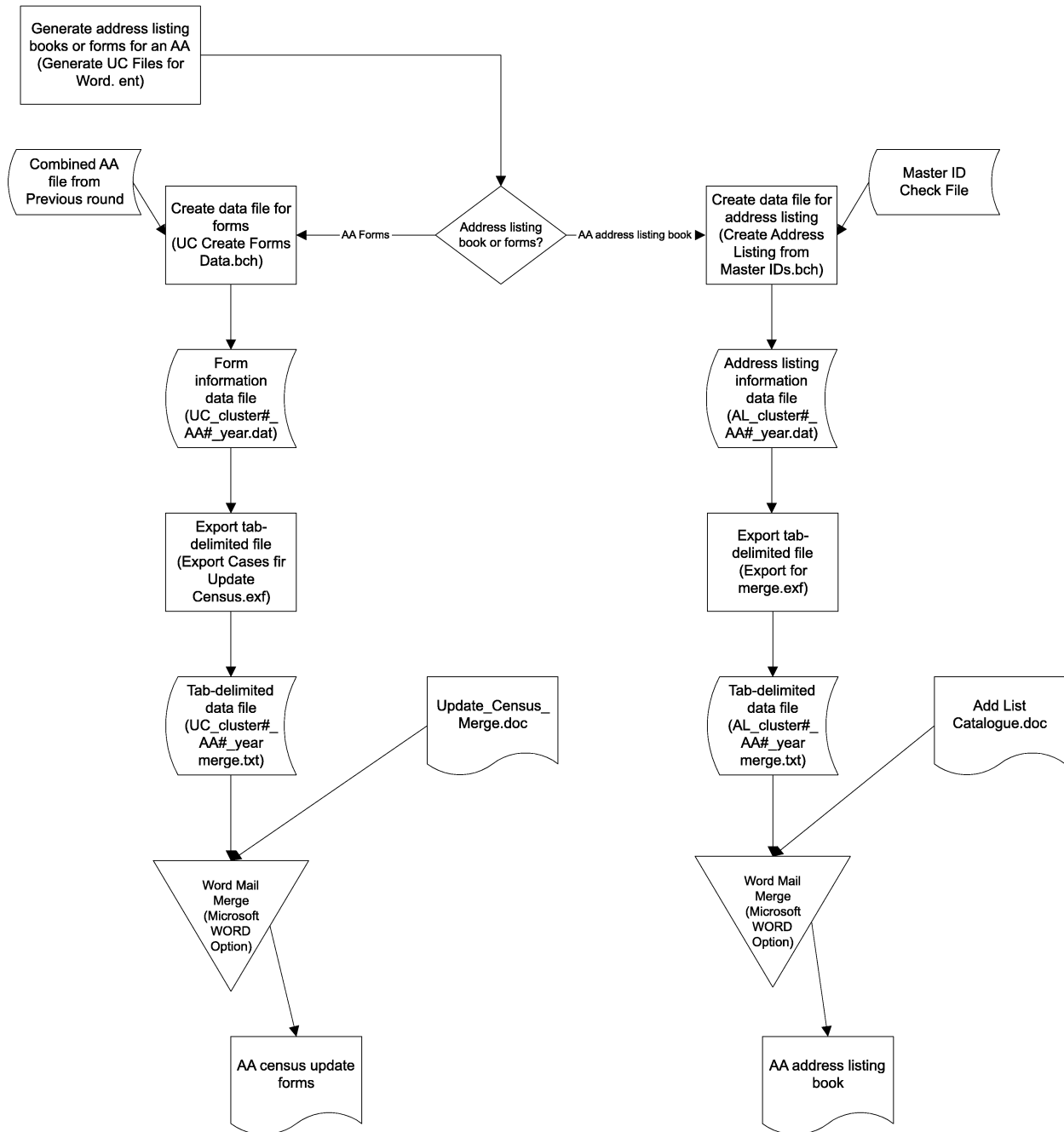
Enter the cluster and AA codes then select either the creation of the input for the export of files for the address listing or the census update forms.

This menu system creates and executes two CSPro Program Information Files (PFF) first, to produce the export data needed for the selected product; and, second, to export it to a tab delimited file that is used in the Microsoft Word mail merge. More information about each is given below.

2. AA Address Listing (Create Address from Master IDs.bch)

This selection (see Figure 9.4) uses the MASTER ID file and creates a tab-delimited output for the individual AA address listing books. The CSPro report generated is named the same as the associated data file. Data and list files are designated as AL\_cluster#\_AA#\_year. List (report) files have a “.lst” extension instead of “.dat,” as used for

Figure 9.2: Production of Census Update Documents Flow Chart



data files. .AL\_01\_02\_2006.lst is a report for cluster #1, AA #2 and year 2006. The sample report is shown at right.

Create text file with address list information for EACH AA.

Input: Master ID file from previous census.

Output: EXTERNAL FILE with address listing information for use with Export Tool (AL\_cluster#\_AA#\_year.dat).

### 3. AA Address Listing (Export for Merge.exf)

CSPro export data tool can be used to produce a tab delimited “merge” file for each AA (see Figure 9.5). These AA files are inputs into the Microsoft Word mail merge.

Input: (UC\_cluster#\_AA#\_year.dat) census update forms data in ASCII file

Output: (UC\_cluster#\_AA#\_year merge.txt) AA tab delimited text file.

Note: There are 10 persons per form. Continuation forms are generated for 11 or more persons in an unit.

### 4. AA Census Update Forms (Update\_CENSUS\_merge.doc)

For each AA, merge tab delimited text file (UC\_cluster#\_AA#\_year merge.txt) with SAVVY Census Update Questionnaire document.

For MAIL MERGE option: Open the “Update\_CENSUS\_merge.doc;” go to “Tools” menu; and select “Mail Merge.”

In “Mail Merge Helper” (see Figure 9.6, click on “Data Source” button (next to “2” in the illustration at left). Select “Open Data Source” option then find a UC\_cluster#\_AA#\_year merge text file created above.

Run merge (next to “3” in the illustration) using given options (see Figure 9.7). For EACH AA, save the document created by the merge (if it appears correct, of course). The cluster and AA codes should be included in the document name. Print the forms, then staple the individual questionnaires together. If there are continuations (more than 10 persons enumerated in the previous census), then continuation forms will be produced. Any continuation should be stapled with the first form.

**Note:** The SAVVY Census Change Questionnaire has NO preprinted information except for the first column in Line #, so it is printed in the normal fashion from Microsoft Word.

```

Application      C:\SAVVY\DP_pgms\AL address listing\Create Address from Master IDs.bch
Type            BATCH
Input Data      C:\SAVVY\DP_pgms\AL address listing\Master ID file.dat
External (UC_ADD_LIST_DICT): C:\SAVVY\DP_pgms\AL data\AL_01_02_2006.dat

Date           Apr 21,2006
Start Time     10:45:19
End Time       10:45:19
                Address Listing Output
                Cluster #1, AA #2

CSPRO Process Summary
-----+-----+-----+-----+-----+
|      8 Records Read ( 100% of input file) |
|      0 Ignored (      unknown,      0 erased) |
|      1 Messages (      1 U,      0 W,      0 E) |
-----+-----+-----+-----+-----+
| Level | Input Case | Bad Struct | Level Post |
-----+-----+-----+-----+-----+
|   1   |      8     |      0     |      8     |
-----+-----+-----+-----+-----+

Process Messages

*** LEVEL 0 PREPROC has 1 messages ( 0 E / 0 W / 1U)
    U -21 *** Processing CLUSTER=01, AA=02 ***

User unnumbered messages:
-----+-----+-----+-----+-----+
| Line | Freq | Pct. | Message text | Denom |
-----+-----+-----+-----+-----+
| 25   | 4    | 50.0 | Entry skipped - does not match Cluster and AA... | 8     |
| 44   | 4    | 0.0  | Address line written... | 8     |
| 47   | 0    | 0.0  | **** Address line problem: %s-%d-%d-%d-%d... | 8     |
-----+-----+-----+-----+-----+

CSPRO Executor Normal End
-----+-----+-----+-----+-----+

```

Figure 9.4: AA address listing book report

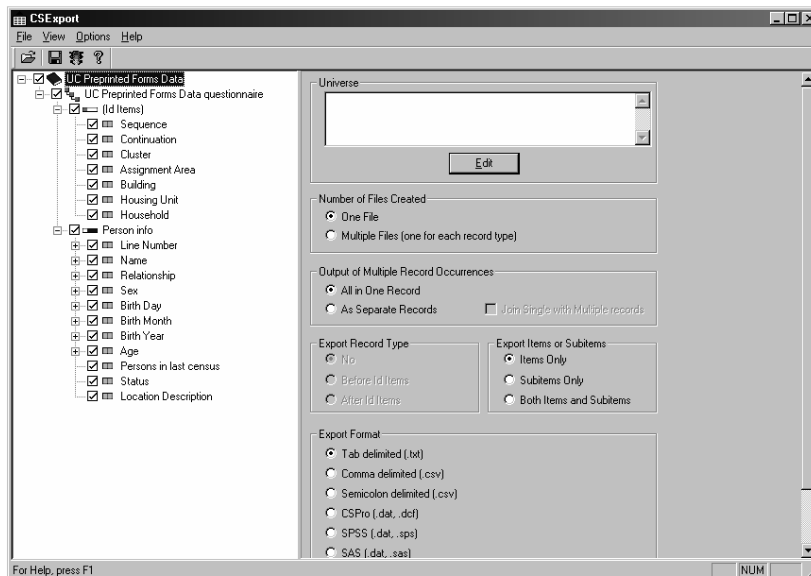


Figure 9.5: AA tab delimited merge file

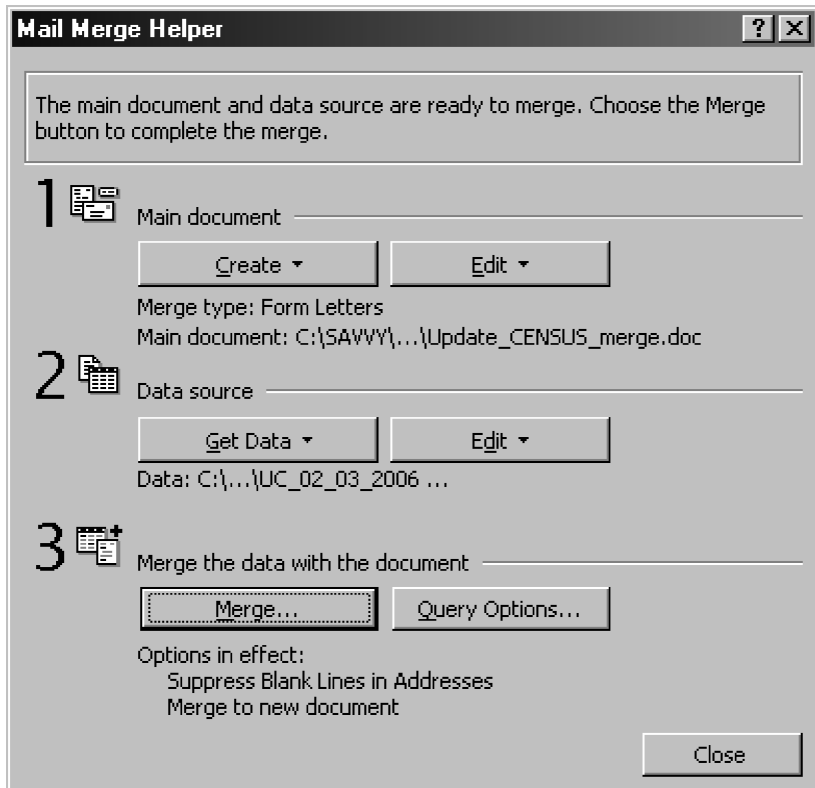


Figure 9.6: Mail merge helper screen



Figure 9.7: Run merge screen

## Census Update Processing

This is a summary of basic operations for entering, verifying, editing, and tabulating the information.

### 1. Forms Tracking (UC0 AA Tracker.ent)

This assumes the assignment area (AA) is the unit for tracking information. It is likely that census update (UC) activities will take time. It is imperative that forms not be lost or overlooked. This application is to keep track of the AA batches, e.g., groups of forms.

The status codes are (subject to change for implementation):

- 1 Check In
- 2 Sent for coding
- 3 Returned from coding
- 4 Sent to data entry (and verification)
- 5 Returned from data entry
- 6 Sent for structure edit (manual correction of essential problems)
- 7 Returned from structure edit
- 9 Sent to storage (processing completed)

The ID items are: cluster number, assignment area, status, associated dates given as day, month, year

This application could be run from desktop icon.

Program Name: **UC0 AA Tracker.ent**

Data File: UC AA tracker.dat (created/modified)

### 2. Enter and Verify Data (UC1 Enter Data.ent, UC1A Enter New Forms.ent)

A batch (complete enumeration results from one AA) might contain preprinted or new household forms. The preprinted forms have IDs that can be verified against the master file of IDs. The new forms have IDs that need to be added to the master file. So we suggest two data capture programs, depending upon whether the forms are preprinted or not. Both applications update the master ID file with the current “status” from the preprinted or new form.

Each group should be entered by one keyer and verified by another. The task here is to capture the information on the form as accurately as possible. Neither entry nor verification should involve any attempt to correct perceived errors on the form.

Complete ID is required. Data must be entered and verified (100% or at agreed upon rates for items and forms). Forms would then be returned to “Batch Tracking Center” for further processing.

This application could be run from desktop icon.

Program Name: **UC1 Enter data.ent** for PREPRINTED forms

Data File: UC1\_Cluster#\_AA#\_year.dat (created/modified/verified)

ID check File: Master IDs File (modified)

Program Name: **UC1A Enter New forms.ent** for NEW forms

Data File: UC1A\_Cluster#\_AA#\_year.dat (created/modified/verified)

ID check File: Master IDs File (modified)

### **3. Combine Preprinted and New Forms (UC2 Combine.bch)**

It might be easier to concatenate the preprinted and new data entry files than to use that as input to this batch program, which moves the data into the standard SAVVY layout.

Program Name: **UC2 Combine.bch**

Data File: UC1\_Cluster#\_AA#\_year.dat + UC1A\_Cluster#\_AA#\_year.dat

External File with correct layout: UC2\_Cluster#\_AA#\_year.dat.

### **4. Structure Edit (UC 2A Structure.ent)**

This is probably the most time consuming step. This could be done after entry, but requires the use of the interactive facility in CSPro. Some values on questionnaires are tested for consistency since they are essential for accurate reporting. Problems are resolved by a subject matter expert using the original form and local knowledge.

Program Name: **UC2A Structure.ent**

Data File: UC2\_Cluster#\_AA#-year.dat (modified interactively)

### **5. Concatenation of “Batch Files”**

Once all AAs have been entered, verified, and the structure edit is completed, the AAs will be concatenated into a single final (unedited) data file for further processing.

Program Tool: Concatenate data (CSPPro)

Input Data File: ALL UC2\_Cluster#\_AA#\_year.dat

Output Data file: Final UNedited UC.dat (created), this file is kept for use with the next census update processing — creation of pre-printed questionnaires

## **6. Pre-edit Frequency Distributions**

Perform frequency distributions on all variables using “Actual value” option to determine the quality of the keyed data before the final edit.

Program Tool: Frequency Distribution Tool [CSPPro]

Input Data file: Final UNedited UC.dat

Output: Pre-edit Frequency tables for all variables

## **7. Batch Edit (UC 3 Final Edit.bch)**

The final unedited data file is checked for a set of not-so-critical problems. If found these are corrected automatically via program logic.

**Note:** It would be possible to add recodes here that could be used in tables.

Input Data File: Final UNedited UC.dat

Output Data file: Final Edited UC.dat (created)

External File: UC 3 Hot deck.dat (created and used in edit)

## **8. Post-edit Frequency Distributions**

This performs frequency distributions on all variables, using the “Actual value” option to determine how the distribution compares to the pre-edit distributions. If distributions are changed significantly, then the edit program should be reviewed.

Program Tool: Frequency Distribution Tool (CSPPro)

Input Data file: Final Edited UC.dat

Output: Post-edit Frequency tables for all variables

### **9. Tabulations (UC 4 Tables.xtb)**

This produces a set of tables based on edited data. Tables should be reviewed for accuracy and consistency. Compare distributions to post-edit frequency tables.

Program: UC 4 Tables.xtb

Input Data file: Final Edited UC.dat

Outputs: tables printed and saved as a TBW (CSPRO table viewer files)

### **10. Create Updated Master ID file (UC 5 Update Master IDs.bch)**

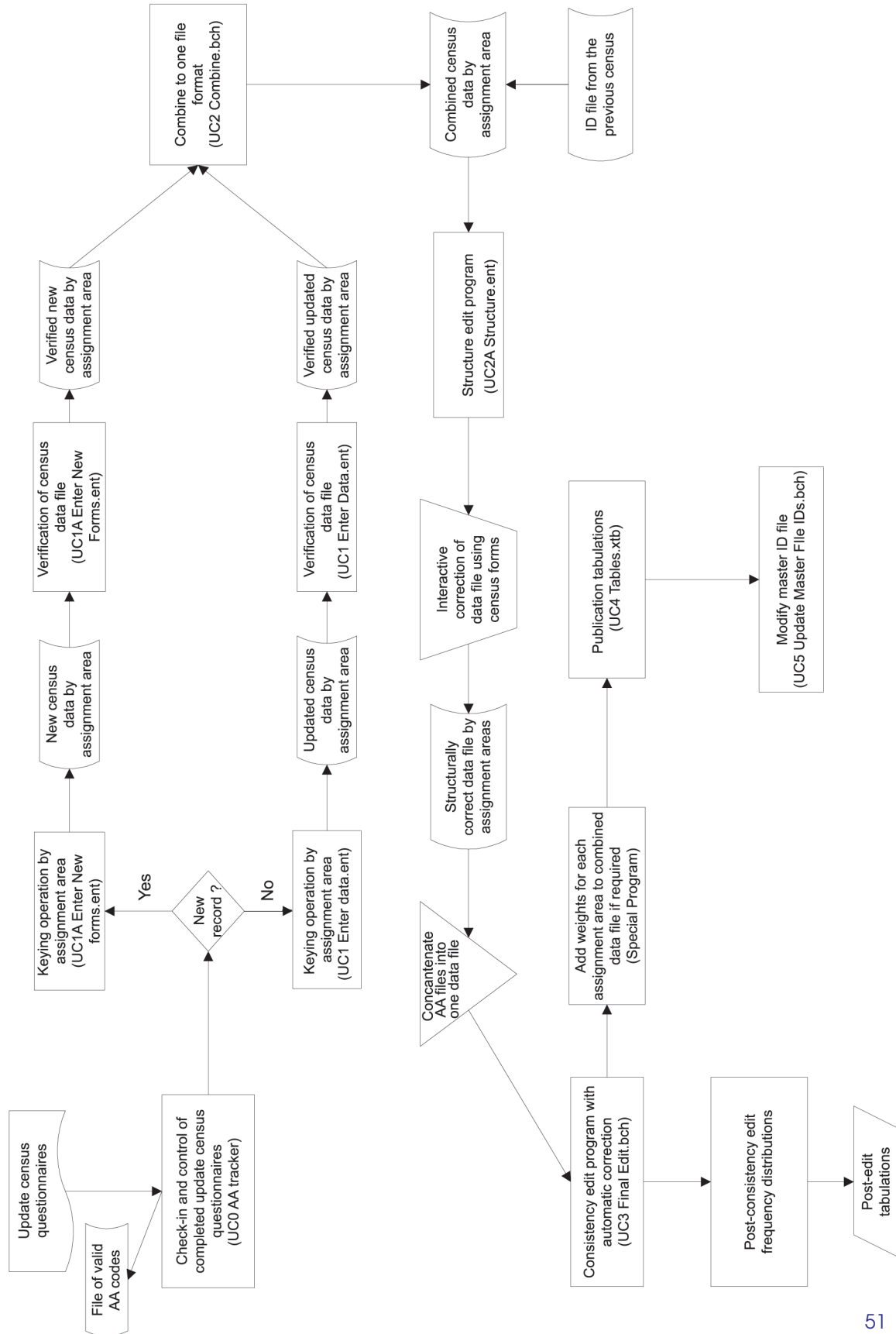
This updates the master ID file, based on “status codes” on both new and preprinted update forms.

Program: UC 5 Update Master IDs.bch

Input Data file: Master IDs file used with two data entry programs.

Output Data file: New updated Master IDs file

Figure 9.8: Census Update Processing Flow Chart



## APPENDIX D: VERBAL AUTOPSY PROCESSING

### 1. Forms Tracking (P1 VA Tracker.ent)

Assumes one person is a focal point for tracking information. It is likely that VA processing will be spread out, occurring throughout the year, so it is imperative to try to keep track of the forms.

The status codes are subject to change for implementation

- 1 Check in
- 2 Sent for data entry (DE)
- 3 Returned from DE
- 4 Sent for ICD coding
- 5 Returned from ICD coding
- 6 Sent for ICD data entry
- 7 Reconciliation needed
- 8 Sent to final ICD data entry
- 9 Processing completed

Codes 7 and 8 are optional.

The ID items are cluster number, assignment area (AA), building number, housing unit number, household number, plus a generated serial number (a four-digit number created and copied on form).

Other items are status codes (listed above) and check-in date, e.g., check-in day, check-in month, and check-in year.

After initial entry, only STATUS should be changed.

This application could be run from desktop icon.

Program Name: **P1 VA Tracker.ent**

Data File: VA ID tracker.dat (created/modified)

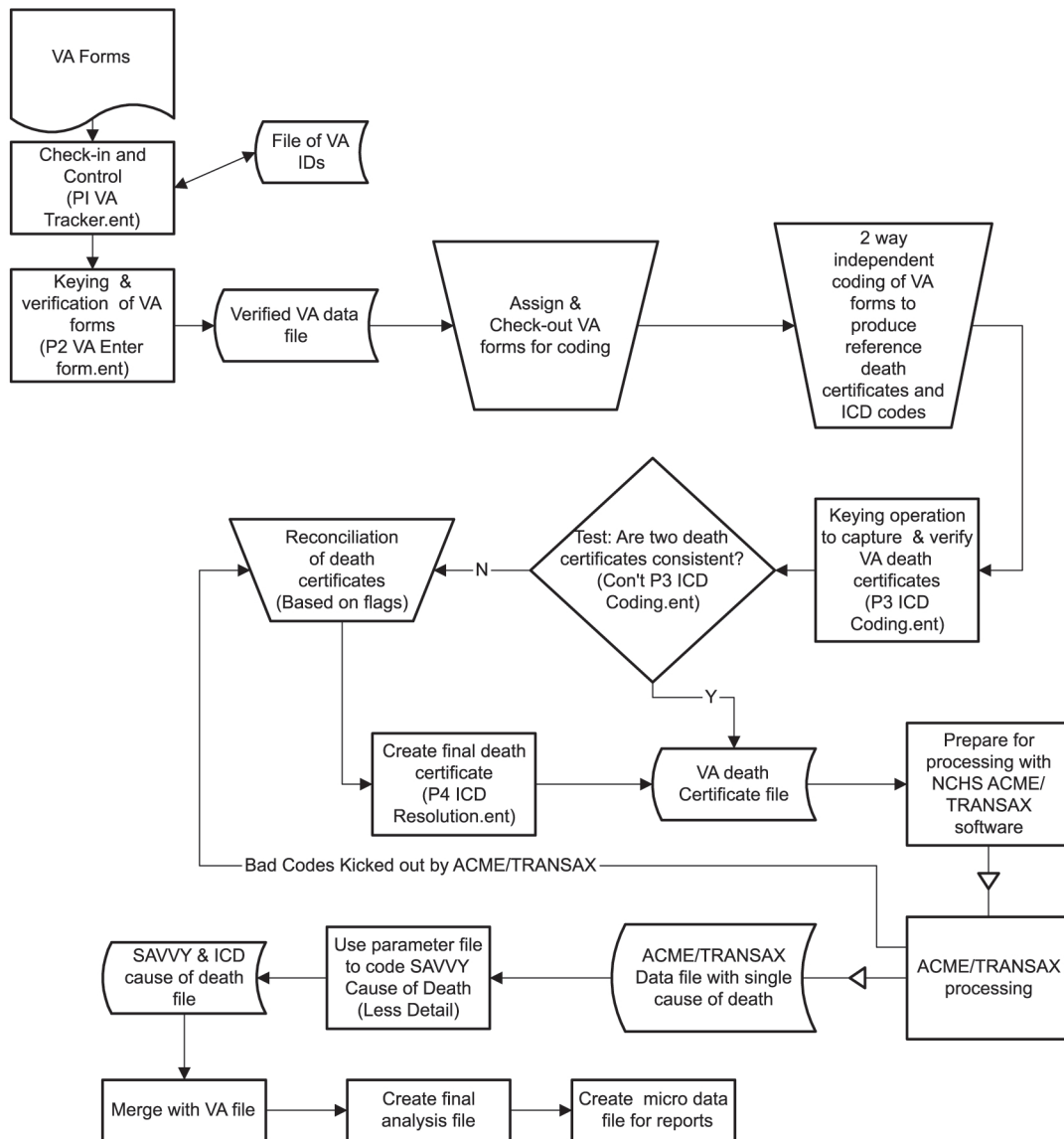
### 2. VA Forms Entry (P2 VA Enter form.ent)

A batch (group) of forms would probably be entered periodically depending on volume. There are three forms, but only one data entry application.

A complete ID is required (including serial number). Data must be entered and verified (100%). Forms would then be returned for ICD coding.

This application could be run from desktop icon.

Figure 9.9: Verbal Autopsy Processing Flow Chart



Program Name: **P2 Enter VA form.ent**

Data File: VA forms.dat (created/modified/verified).

**Note:** Tracking data file could be used as ID check file.

### 3. ICD Coding and Entry (P3 ICD Coding.ent)

This is probably the most time consuming step. This could be done before the verbal autopsy (VA) form is checked in, but both are needed for processing. Once two international classification of diseases (ICD) forms for a VA form are completed, they need to be entered and verified.

After verification, if the two forms match (both sets of three ICD codes are identical), then a record is written to “Final COD” data file, and the VA and ICD forms can be archived. If not, the forms are reviewed, revised, and recoded; and a third reconciliation form is completed (an extra step).

Program Name: **P3 ICD Coding.ent**

Data File: ICD forms.dat (created/modified/verified)

ICD ID file: ICD IDs.dat (created/updated)

Final COD file: Final COD.dat (created/updated).

**Note:** Tracking data file could be used as ICD ID file.

### 4. Reconciliation ICD entry (P4 ICD Resolution.ent)

If necessary (codes on previous forms did not match), a third ICD form will be created with only one set of data. This will be considered final. These data must be keyed and verified also.

Program Name: P4 ICD Resolution.ent

Data File: ICD Resolution forms.dat (created/modified/verified)

ICD ID file: ICD IDs.dat (created/updated)

Final COD file: Final COD.dat (created/updated)

**Note:** Tracking data file could be used as ICD ID file.

### 5. Create ACME Record

At this point, VA forms have been entered and verified, and have a “Final COD” record for each form. This information needs

to be merged to create an ACME input record.

## **6. Add Cause of Death Code (P5 Add Codes.bch)**

The Automated Classification of Medical Entities/Translation of Axis (ACME/TRANSAX) software produces a record (in .TRX file) for each VA with a final cause of death. This is an ICD code (without the decimal point), which must be recoded into one of the special reporting codes for VA (currently 57). Both of these codes must then be copied to the original VA record.

Program Name: **P5 Add Codes.bch**

Data File: VA forms.dat [modified (ICD and VA codes added)]

TRANSAX file: Output from ACME/TRANSAX.trx (read)

ICD to VA code file: Mortality Rows.prn (ICD → VA code table)

Output File: VA file updated with ICD/VA codes

APPENDIX E:  
ILLUSTRATIVE  
TABULATION AND  
CROSS TABULATION

1. Population by single year of age and sex for country
2. Population by single year of age and sex by cluster number
3. Population by single year of age, sex, and parental survival
4. Deaths by single year of age and sex
5. Death rates by single year of age and sex
6. Causes of death by single year of age and sex
7. Causes of death by age groups and sex
8. Population by age, sex, and educational attainment
9. Population by age, sex, and economic activity
10. School attendance by age and sex



MEASURE Evaluation  
Carolina Population Center  
University of North Carolina at Chapel Hill  
CB 8120, 123 W. Franklin St.  
Chapel Hill, NC 27516 USA  
[www.cpc.unc.edu/measure](http://www.cpc.unc.edu/measure)

International Programs Center  
Population Division  
U.S. Census Bureau  
Washington, DC 20233  
[www.census.gov](http://www.census.gov)

MS-07-26-DPM



**USAID**  
FROM THE AMERICAN PEOPLE



**USCENSUSBUREAU**  
*Helping You Make Informed Decisions*