

Paper P502

Use of SAS/AF V8e to Compare Death Certificate Data with Health Survey Data from the National Center for Health Statistics

Gretchen K. Jones, NOVA Research Company, Hyattsville, MD

Sandra T. Rothwell and Christine S. Cox, National Center for Health Statistics, Hyattsville, MD

ABSTRACT

Determination of Fact of Death is one of the follow-up activities which the National Center for Health Statistics (NCHS) conducts for health surveys. The National Death Index (NDI) is a data base containing information on all U.S. deaths since 1979. To validate a death, the NCHS orders copies of Death Certificates to compare against a more robust set of data than is available from the NDI.

The authors have developed a SAS/AF (V8e) application, using the Form Viewer attached to a SAS Dataset Model, which provides a tool to compare NCHS survey data with actual Death Certificates. Access is provided (through push button control) to frames containing data from three different data files. The user compares and "scores" the extent of match of the data variables on a frame with six pages. The final frame gives a summary of scores from the previous frame, with a box for final evaluation of whether the Death Certificate matches the survey data on the files. The presentation will include a demo of the application.

This paper is appropriate for persons with intermediate SAS skills.

INTRODUCTION

This paper describes a SAS/AF V8e application of the Form Viewer attached to a SAS Dataset Model. The application involves being able to view the National Health and Nutrition Examination Survey (NHANES) data from subjects now believed to be deceased. The purpose is to allow an evaluator to compare variables on the frames with information recorded on a death certificate thought to be a possible match. The evaluator will score and comment using the frames, thus updating the underlying SAS file.

Background is given describing the data and how the previous system worked; the frame-building process is explained (with coding); screen prints of sample frames are presented; and improvements of the application over the previous system are listed.

BACKGROUND ON FOLLOW-UP AND LINKAGE

One of the most useful tools for the epidemiologist is the longitudinal or follow-up study. These studies provide baseline measures that can be compared to subsequent health outcomes to determine risk factors for some of the more common chronic diseases as well as to provide

some information on the natural course of these diseases and conditions.

A major health outcome is the fact of death, the underlying cause of death (or the condition that led to the eventual death), and any other contributing diseases or conditions (such as previous heart attacks, etc.)

THE NATIONAL DEATH INDEX (NDI) AS A PROVIDER OF DATA RECORDED ON DEATH CERTIFICATES

The National Center for Health Statistics (NCHS) provides a data base, the National Death Index (NDI), for use under fairly strict guidelines, that can provide data on fact and cause of death. The NDI user provides a limited set of variables and the NDI software provides all potential matches to the person described by this set of variables. State of death and certificate number are provided on each potential match.

It is then the job of the user to determine which matches could be the death certificate for a subject in an epidemiologic study. The NDI does not provide any more confidential information than is necessary for the user to make some kind of informed decision. For example, if the user provides the subject's Social Security Number (SSN), the NDI provides only information that specifies which digits of the SSN match. The SSN itself is not provided. Another issue is the fact that birth dates can often be misreported in the survey and/or on the death certificate. Often there is a close match on SSN and a birth date and the name is a commonly used one. This causes any match to be tentative. Both probabilistic and deterministic methods are used to winnow out bad matches or to find the very good or perfect matches. There is always a large number which are neither obviously matches nor non-matches. These cases are resolved by ordering copies of the death certificate and reviewing all socio-demographic variables present on the certificate and collected during the survey participant interview. This frame application was developed to facilitate this review process.

REASONS FOR ORDERING DEATH CERTIFICATES

Both the survey and the death certificate itself contain more information than is available from the NDI match. For example, whether the subject ever served in the armed forces and the subject's major occupations or industries in which he or she worked can usually be found on the death certificate. The name of the informant on the certificate is often someone who lived in the house with the subject when they were interviewed either at baseline or at some later follow-up contact. Finding m

matches to this information on certificates can often confirm or deny the match. Ordering the certificates can sometimes discover the death of a twin, where the last names are the same, the first names sound alike, the birth dates match perfectly and the SSNs are very similar. If the twin was not in the study, an incorrect death can be eliminated.

STATEMENT OF THE TASK TO BE PROGRAMMED IN SAS/AF

Data from several SAS source files (a Subject File, a Roster File, and an Address File) are to be displayed on screens so that an evaluator, usually female, (hence feminine pronouns are used in reference to her throughout) may manually compare the information with corresponding data on the death certificate which has been ordered from the State in which the person presumably died.

For every variable on the Subject File looked at, there is a place for the evaluator to enter a "Match Code," which can be an "E" for exact match, a "P" for partial match, or an "N" for no match. Next to the Match Code combo box, there is a text entry field where the evaluator must enter the data from the death certificate if she has selected a "P" or an "N" from the Match Code combo box. And the evaluator is not allowed to enter anything in the text entry field if she has selected an "E" from the Match Code combo box. The frame is programmed so that any box which must have data entered into it has a background color of yellow.

From comparing the address on the death certificate with data from the Address File, the evaluator is able to assign a Code of "N," for no address match, an "S" for State only match, a "C" for City and State matches, or an "E" for exact match with the Household Address, or with the Mailing Address. And lastly, comparing the name of the informant listed on the death certificate with data on the Roster File, the evaluator assigns an "N," a "P," or an "E" again for no match, partial match or exact match on informant in household, or an other contact.

Privacy and security considerations must be programmed into the system. Only someone who takes the privacy oath and has an authorized USERID will be allowed to bring up the system. The evaluator must be able to navigate between screens as much as possible, and when finished with match coding, will go to a decision screen, where she will be able to look at frequencies of match codes she has assigned on previous screens, and to come to a decision as to whether the person described on these files is the one on the death certificate in front of her. She then must have the ability of assigning a "decision" of "N" (not a match), "M" (a match), or "R" (needs to be reviewed by client).

HOW OLD SYSTEM WORKED

The previous system was written in SAS/AF version 6. It was set up to handle one (and only one) survey. It was strictly "fill in the blank", no push buttons, no objects with inheritance. And navigation was difficult. The system

provided no summary statistics such as what proportion of the potential certificates had been reviewed or what proportion of review certificates were considered "true".

THE NEW SYSTEM - HOW IT IS DESIGNED

To implement the new system, it was necessary to use SAS/AF Version 8e, because we needed to make use of the form viewer, which doesn't exist in Version 8. The system consists of five frames or screens, one containing six pages.

The first screen has the Title ("Death Certificate Verification System") and the confidentiality statement. There is a push button for the evaluator to click on if she accepts the statement, and the second screen opens. The evaluator is asked to choose the survey being reviewed from a drop-down box, and to enter her USERID. If the evaluator is found to be authorized, the third screen opens, where she enters the CASEID of the subject for the appropriate survey.

Screen four consists of six pages, which can be navigated using the "page turns" in the upper right or left-hand corners, or the push buttons at the bottom of the frame. Pages 1-3 contain data to be reviewed and updated from the Subject File, Page 4, from the Address File, and Page 5 from the Roster File. Page 6 is the Comment Screen, where the evaluator is given a place to provide additional input if needed.

When the evaluator is ready to make a decision as to whether the death certificate is a match, she navigates to the fifth frame, where statistics as to her previous scoring on individual items are displayed in text control boxes, and she will enter her decision code.

Figure 1 below is a sample of the first page of the Subject Screen. The evaluator will choose E, P, or N from the combo-boxes in the middle column for each variable in the first column. Then if E is not chosen, the evaluator is forced to enter (in the last column) what was written on the death certificate for that variable.

Figure 2 below is a sample Decision Screen. Frequencies of the codes that the evaluator entered on the Subject Screens are shown, as well as the Match codes which she entered on the Roster Screen. After careful consideration, she chooses Match, False, or Review as a Final Decision.

HOW THE APPLICATION IS CODED

Much of the programming of the application could be done automatically by merely dragging and dropping objects onto a frame, and setting attributes through the properties window. However, a considerable amount still had to be hard-coded, using SAS Component Language (SCL). The SCL attached to the frame is called the Frame SCL. When one is using a Viewer attached to a Model, one also uses Model SCL, which can be accessed through right-clicking the mouse on the Viewer. Many of the functions of objects on the Viewer were coded through Model SCL.

Death Certificate Verification System

WESID:	XXXXXX	State of Death:	TX	Death Certificate:	XXXXXX
First Name:	XXXXX				
Middle Name:	XXXXXX				
Last Name:	XXXXX				
Nick Name:					
Other FName:					
Other MName:					
Other LName:					
Date of Death:	01 20 1989				
SSN:	XXXXXX				

Go To MAIN
↩ Subject
↩ Address
↩ Roster
↩ Comment
Go To DECISION

Figure 1. Subject Screen (first page)

NHANES III Death Certificate Verification System

Frequency of Codes from Subject Screen	Match on Household Roster	P
Number of Items coded "E" 10	Match on Contacts	E
Number of Items coded "P" 3	Match on Mailing Address	N
Number of Items coded "N" 5	Match on Household Address	N

Final Decision:

Go To MAIN (New Case)
↩ Subject
↩ Address
↩ Roster
↩ Comment
EXIT

Figure 2. Decision Screen

Following is an example of the Frame SCL for the Subject Screen (see Figure 1).

```
entry SURV: num SURVEY Q1-Q22
IDPASS: char;
dcl num rc;
dcl list msgList=
{'Invalid WESID. Return to MAIN.'};

INIT:

SASDataSet1.table='DCert.Subject5';

SASDataSet1.where='WESID =' ||
    quote(IDPASS);

rc = sysrc();
if rc ne 0 then do;
choice=
MessageBox(msgList,'I','O','Error');
if choice = 'OK' then Call Display
    ('Screen3.Frame',survey,
    surv);
end;
return;
Subroutine:
SASDataSet1._GetColumnText('MA',Q1);
SASDataSet1._GetColumnText('MB',Q2);
SASDataSet1._GetColumnText('MC',Q3);
.
.
.
SASDataSet1._GetColumnText('MU',Q22)
;
return;

PBSubject:

    Formviewer1._gotoPage(1);
    return;

PBAddress:

    Formviewer1._gotoPage(4);
    return;

PBRoster:

    Formviewer1._gotoPage(5);
    return;

PBComment:

    Formviewer1._gotoPage(6);
    return;

PBDecision:

Link Subroutine;
    Call display
('SCREEN8.FRAME',SURV,SURVEY,Q1,Q2,
Q3,Q4,Q5,Q6,Q7,Q8,Q9,
Q10,Q11,Q12,Q13,Q14,Q15,Q16,Q17,Q18,
Q19,Q20,Q21,Q22,IDPASS);
return;
```

```
PBMain:
    Call Display ('SCREEN3.FRAME',
SURVEY, SURV);
return;
```

SURV, SURVEY, and IDPASS are SCL variables that are passed from Screen3. In the INIT section, the Model SAS Data Set is opened, and an observation is searched for where the variable WESID = IDPASS, the ID which the evaluator entered at Screen3. If a match is not found, an error message is displayed and control is returned to Screen3. Otherwise control stays with Screen4. The Subroutine puts the values of the "match codes" that the evaluator has entered into the SCL variables Q1-Q22. The coding for the Subject, Address, Roster, and Comment push buttons merely takes one to the appropriate page on the Form Viewer. The push button for the Decision Screen passes all the previous SCL variables, plus the ones assigned in Subroutine, and transfers control to the Decision Screen (Screen8). The push button for Main merely transfers control to the Main Screen (Screen3).

IMPROVEMENTS UNDER 8E

Under the new system, any number of surveys may have potential death certificate matches loaded. A judicious use of inheritance will ease the programming job for this. We can put more easily understood and used navigation objects on the screens and we can provide better security. We can provide better editing of the data entry fields to be sure all available information is being used in the judgement. We can be sure that there is a decision on each potential match for a given survey before the work is turned back to the linkage staff. We can provide some summary statistics available from a menu or push-button. And, finally, the system will be much more pleasing to the eyes.

CONCLUSION

The National Center for Health Statistics has a need for a tool to enable an evaluator to view and score data from national survey files to compare against Death Certificates which have been ordered from the various states. The purpose is to determine fact of death for particular subjects who have been examined in the NHANES surveys. The Form Viewer with SAS Data Set Model implemented in SAS V8e, which provides such a tool, has been described in this paper.

The application also provides an electronic record of the entire death certificate evaluation process for a given NDI match.

REFERENCES

SAS Institute Inc. *SAS/AF Software: Application Development I Course Notes*. Cary, NC: SAS Institute Inc., 2000.

SAS Institute Inc. *SAS/AF Software: Application Development II Course Notes*. Cary, NC: SAS Institute Inc., 2000.

SAS Institute Inc. *SAS/AF Software: Changes and Enhancements in Version 8 Course Notes*. Cary, NC: SAS Institute Inc., 2000.

ACKNOWLEDGMENTS

The authors wish to thank Cay Loria, formerly of NCHS, who originated the idea of using SAS/AF as a tool in matching survey data with information on the death certificate and Ann Rockett of SAS Institute, whose technical assistance has been invaluable in developing the application.

CONTACT INFORMATION

Gretchen K. Jones
6525 Belcrest Road, Room 730
Hyattsville, MD 20782

Phone: 301/458-4301
Fax: 301/458-4038
Email: gkjones@cdc.gov