

Paper 094-31

Archival Documentation of a SAS® Database via the Web

Joel Achtenberg and Elizabeth M Hornbeck
Washington University School of Medicine
St. Louis, Missouri

ABSTRACT

The Collaborative Longitudinal Evaluation of Keratoconus (CLEK) Study was a multi-center observational study funded by the National Eye Institute to characterize vision, corneal changes, and quality of life in keratoconus patients, and to determine the progression of the disease over time. A total of 1,209 keratoconus patients at 16 participating clinics were examined annually for 8 years.

Because CLEK is the first systematic, longitudinal study of keratoconus, an eye condition about which very little is known, study data will be an especially valuable resource to investigators outside the study.

To provide access to the data following the study's completion, we have produced a web-based archive containing extensive documentation of the study and its database. The archive includes traditional text and standard SAS documentation. We supplemented the traditional documentation with PDF images of all study data collection forms, annotated with SAS dataset variable names for the associated form fields and with links from each variable name to details about that variable and a summary of its contents.

We will describe the methods and tools used to produce this archive, and illustrate with both code fragments and sample screens.

BACKGROUND

The database for the CLEK Study is composed of 21 SAS datasets: one for each of the four major data collection forms used in the study, several administrative datasets, and datasets built from raw data sent from our two reading centers. More than 1000 variables are represented.

Since the study began enrollment in 1994, numerous revisions have been made to the study protocol and forms and, consequently, numerous additions and deletions of variables from the associated SAS datasets. A common format library is shared by all study datasets, and formats in the library have also undergone changes over the years.

Data collection for the study ended in 2004, with the last patient follow-up visit completed in August of that year. Analysis, preparation of manuscripts, and the closeout and archiving of study materials have continued since then.

To archive CLEK study materials and data, we have produced an archive website containing historical information and material useful for those interested in understanding the study and interpreting its results and publications.

DESIGN CONCEPT

Three objectives went into the design and selection of content for this Web Archive:

- To provide a history of the study and overview of its organization and operation.
- To provide sufficient details of the protocol, procedures and data collection forms to aid in interpreting the study's published results

- To provide an introduction to the compiled database, its organization and contents, for those outside investigators who may wish to use it for further analysis.

The archive includes standard text documents, including the study's Manual of Procedures, a historical overview and timeline, bibliography of published manuscripts and publications, and more. For this SUGI presentation, however, we focus on the documentation of the study's SAS database, including both traditional SAS documentation and the extended documentation developed specifically for this archive.

TRADITIONAL DOCUMENTATION OF THE SAS DATABASE

PROC CONTENTS and PROC FORMAT

Traditional documentation of a SAS database begins with printing of the output from PROC CONTENTS for all datasets in the library. For our archive, we used the Output Delivery System (ODS) to format the output as a Portable Document Format (PDF) file, and include both a directory of all the datasets (not shown) and contents for each individual dataset (see Figure 1: PROC CONTENTS).

The PDF output file for this and all of our archive documents described below, is routed to a directory on our web server, where it can be linked from the archive menu pages.

Collaborative Longitudinal Evaluation of Keratoconus--CLEK DataSet Contents for MASTER datasets, as of 09/13/2005			
The CONTENTS Procedure			
Data Set Name	MASTER.EL	Observations	1300
Member Type	DATA	Variables	16
Engine	V8	Indexes	0
Created	Tuesday, July 26, 2005 02:27:12 PM	Observation Length	56
Last Modified	Tuesday, July 26, 2005 02:27:12 PM	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	YES
Label			
Data Representation	WINDOWS_32		
Encoding	wlatin1 Western (Windows)		

Alphabetic List of Variables and Attributes						
#	Variable	Type	Len	Format	Informat	Label
7	AGEGT12	Char	1			Is patient at least 12
12	CONDITN	Char	1			Cataract, etc. both eyes
14	CONSDATE	Num	5	MMDDYY10.	MMDDYY6.	Date consent form was signed
13	CONSENT	Char	1	SYN.		Consent form signed
9	DEFECT	Char	1			Vogts, Fleischers, etc
1	EDITDATE	Num	8	MMDDYY10.	MMDDYY6.	Edit Date
15	ELIGIBLE	Char	1			Does Patient appear Eligible?
2	EXAMDATE	Num	8	MMDDYY10.	MMDDYY6.	Date of CLEK Eligibility

Figure 1: PROC CONTENTS

We also used ODS to produce PDF output from PROC FORMAT, documenting the many custom formats shared by variables in our database (see Figure 2: PROC FORMAT).

Collaborative Longitudinal Evaluation of Keratoconus--CLEK CLEK Custom Format Library Used for Value Formatting Current to 07/12/2005				
Format Type	Format Name	Width	Value(s)	Label
Format	\$ABSCYL	18		MISSING OR PK/EPIK
			1	(<1.5 D)
			2	(1.5 D and up)
Format	\$AGEDEC	8	1	10 - <20
			2	20 - <30
			3	30 - <40
			4	40 - <50
			5	50 - <60
			6	60 - <70
			7	70 - <80
			8	80 - <90
Format	\$ALLNONE	22	1	All of the Time
			2	Most of the Time
			3	A Good Bit of the Time
			4	Some of the Time
			5	A Little of the Time
			6	None of the Time
			OTHER	*
Format	\$AXCH	19		MISSING OR PK/EPIK

Figure 2: PROC FORMAT

CODEBOOK

Throughout our study we have used a Codebook, combining in one concise document the information from PROC CONTENTS and PROC FORMAT for all variables in all datasets used by the study. The codebook provides a more comprehensive view of the database than CONTENTS or FORMAT viewed separately, and is more intuitive and accessible to non-programmers.

The codebook is produced, entirely in SAS, by merging the output from CONTENTS and FORMAT within a dataset that formats the information for easy, convenient access by both SAS users and those unfamiliar with SAS. (Note: this program was written in 1995 and has been in use continuously since then, without modification to use features since added in more recent versions of SAS.)

At the study's conclusion, we produced a PDF version of the codebook for inclusion in the archive web site (illustrated in Figure 3).

Collaborative Longitudinal Evaluation of Keratoconus--CLEK					
CODEBOOK: Documentation of variables (with value formatting) in CLEK Datasets (as of 09/13/2005)					
VARIABLE	DATASET	DESCRIPTION	L	FORMAT	INFORMAT
ARFFDRT	MASTER.F010605E	FDACL, CPIC Read Date	8	MMDDYY10.	MMDDYY10.
ARHF	MASTER.F010605E	Habitual Fix, Apical Relation, CPIC	1	\$P_TOUCH.	\$
		1 = Definite Touch			
		2 = Touch			
		3 = Clearance			
		4 = Def. Clearance			
ARHFFDT	MASTER.F010605E	Habitual, CPIC Photo Date	8	MMDDYY10.	MMDDYY10.
ARHFFDR	MASTER.F010605E	Habitual, CPIC Reader	3	\$3.	\$
ARHFFDT	MASTER.F010605E	Habitual, CPIC Read Date	8	MMDDYY10.	MMDDYY10.
AS	MASTER.K072105	Sim. Keratometry Index-Steep (deg:0-180)	8	5.1	
AS	MASTER.K100504	Sim. Keratometry Index-Steep (deg:0-180)	8	5.1	
AS	MASTER.K100504	Sim. Keratometry Index-Steep (deg:0-180)	8	5.1	
ATHOME	MASTER.VQ	Diff. entertaining friends in own home	1	\$DIFF1.	\$
		1 = No difficulty			
		2 = A little difficulty			
		3 = Moderate difficulty			
		4 = Extreme difficulty			
		5 = Stopped b/c eyeght			
		6 = Stopped b/c other			
AUTOACC1	MASTER.EX	Auto Accidents?	1	\$Y/N1.	\$
		1 = Yes			
		2 = No			
		(other) = *			
AUTOACC2	MASTER.EX	Auto Accidents: how many cars involved?	3	2.	
AUTOACC3	MASTER.EX	Auto Accidents: how many cars repaired?	3	2.	
AUTOACC4	MASTER.EX	Auto Accidents: cost to fix your car?	4	5.	
AUTOACC5	MASTER.EX	Auto Accidents: did you go to hospital?	1	\$Y/N1.	\$
		1 = Yes			
		2 = No			
		(other) = *			
AUTOACC6	MASTER.EX	Auto Accidents: anyone else to hospital?	1	\$Y/N1.	\$
		1 = Yes			
		2 = No			
		(other) = *			

Produced by "PROGS.PROGRAMS_CODEBK2.SCL"

Figure 3: Codebook

BEYOND TRADITIONAL: ANNOTATED FORMS

In order to facilitate an in-depth understanding of the relationship between the data collection forms used in the study and the elements of the database produced from those forms, we built an annotated copy of each form with "hot links" providing details of the variables associated with each element on the form. Building each annotated form required two steps: generating a PDF file containing the detailed information about each variable, and creating an image of the form using the associated variable name to label the field and provide a clickable link to the details.

USE SAS MACROS TO PRODUCE PDF FILES WITH DETAILS ABOUT EACH VARIABLE

We designed a simple set of SAS macros to combine standard variable descriptors (variable name, label, type, length, format, informat) for each field with a statistical summary of the contents of that field, using PROC MEANS for numeric fields, and PROC FREQ for character fields.

(Note: The complete macro code and a typical invocation are included as an appendix to this paper; in the discussion which follows, only fragments are included.)

The first macro, %ANNOTATE, is invoked once for each dataset to be processed. It receives five parameters: the library and dataset names (LIB, DSN), an optional dataset alias (DSNREF), the name of a variable in the dataset to be used for statistical sub-grouping (STATSBY), and a list of variables for which a statistical summary is not to be produced (EXCL). For each variable in the dataset, %ANNNOTATE first gathers dictionary information about the variable using SAS File I/O functions VARNAME, VARLABEL, VARLEN, VARTYPE, VARFMT, VARINFMT.

```
%let dsid=%sysfunc(open(&dset));
%if &dsid %then %do;
  %let nfmts=0; %let ninfmts=0; *--initialize counters--;
  %let nvars=%sysfunc(attrn(&dsid,nvars));
  %do i=1 %to &nvars;
```

```

%let existfmt=0; %let existinf=0;
%let vname=%sysfunc(varname (&dsid,&i));
%let vlabel=%sysfunc(varlabel(&dsid,&i));
%let vlabel=%sysfunc(compress(&vlabel,%str('')));
%let vlen=%sysfunc(varlen (&dsid,&i));
%let vtype=%sysfunc(vartype (&dsid,&i));
%let vfmt=%sysfunc(varfmt (&dsid,&i));
  %if x&vfmt=x %then %let vfmt=Undefined;
  %else %do;
    *-- first remove $ and periods from the format--;
    %if %substr(&vfmt,1,1)=$ %then %let vfmt=%substr(&vfmt,2);
    %let vfmt=%substr(&vfmt,1,%eval(%length(&vfmt)-1));
    *--determine if format/informat is in the library--;
    %if &vtype=C %then
      %let existfmt=%sysfunc(exist(library.formats.&vfmt..formatc));
    %if &vtype=N %then
      %let existfmt=%sysfunc(exist(library.formats.&vfmt..format));
    *--run MAKEFMT macro to build PDF file for this format--;
    %if &existfmt %then %makefmt(&vfmt,&vtype,format);
  %end;

```

%ANNOTATE then formats and includes that information in a set of centered TITLE statements.

```

*--Use TITLE statements to add dataset & var details to freq/MEANS output--;
title1 j=c "*****";
title2 j=c "Dataset is &DSET";
title3 j=c "Variable Name is &VNAME" ;
title4 j=c "Variable Label is '&VLABEL'";
title5 j=c "Variable Type is &VTYPE";
*--TITLE statements for formats/informats include imbedded URLs--;
%let click=Click here for details;
%if (&existfmt) %then
  title6 j=c "^S={URL='format_&VFMT..PDF'} Variable Format is &vfmt - &Click" ;;
%if (&existfmt=0)%then
  title6 j=c "Variable Format is &vfmt";;
%if (&existinf) %then
  title7 j=c "^S={URL='infmt_&VINFMT..PDF'} Variable Informat is &vinfmt--&Click";;
%if (&existinf=0) %then
  title7 j=c "Variable Informat is &vinfmt";;
title8 j=c "*****";

```

Finally, %ANNOTATE runs either the MEANS procedure (for numeric variables) or the FREQ procedure (for character variables). PROC FREQ is run as a two-way crosstabulation and MEANS is run with a CLASS statement, using the specified STATSBY variable to group the data into classification levels. The output of FREQ or MEANS, with the added information in the TITLE statements, is formatted as PDF and routed to a file in the web archive directory.

```

*--Summarize by value of specified STATSBY variable, usually by year--;
*--Omit the summary if var is on excluded list*/
%if %index(x&excl,&vname)=0 %then %do;
  %if &vtype=C %then %do;
    proc freq data=&dset;
      tables &STATSBY*&vname;
      %if &statsby=EXAMDATE %then FORMAT &STATSBY YEAR4.;
      ;run;
    %end;
  %if &vtype=N %then %do;
    proc means data=&dset; var &vname;
      %if &statsby=EXAMDATE %then FORMAT &STATSBY YEAR4.;
      %if &vfmt=:MMDDYY %then FORMAT &vname YEAR4.;
      ;CLASS &STATSBY;
      ;run;
    %end;
  %end;
%else %do;

```

```

data _null_ ; file print footnotes ;
put "No Statistical Summary for this variable" ; run ;
%end ;

```

%ANNOTATE also invokes the second macro, %MAKEFMT, to gather details of the custom format and/or informat, if any, for that variable, and produces a PDF page containing that information. The link to the format/informat detail page is imbedded in the associated TITLE statement produced by %ANNOTATE (above).

```

%MACRO MAKEFMT(vfmt,vtype,ftype);
title; footnote;
%let file=&basefile.\&ftype._&vfmt..PDF ;
%let existfile=%sysfunc(fileexist(&file));
%if &existfmt and not &existfile %then %do;
ODS PDF file="&file" style=jfastyle bookmarklist=hide;
proc format fmlib lib=library;
%if &vtype=C %then select $&vfmt.; %else select &vfmt.;; run;
%end;
ODS PDF close;
%MEND MAKEFMT;

```

The result after running %ANNOTATE for each dataset is a PDF file for each variable incorporating in one screen per variable detailed dictionary information for the SAS variable, links to the associated format and/or informat, and a brief bivariate summary of the data in the CLEK database. The next step in building our archive was to provide access to these detail pages from copies of the original CLEK data collection forms.

USE ADOBE ACROBAT TO PREPARE ANNOTATED FORMS

We used Adobe Acrobat Professional 7.0 to convert each of the forms used in the study from the original Microsoft Word document to a PDF file. We then added text form field annotation to display the associated SAS variable name on or adjacent to the entry field or checkbox for each variable.

A sample from one of our annotated forms is illustrated in Figure 4, at right. The grey boxes are the annotations containing variable names.

Figure 4 shows a sample of an annotated form titled "CLEK Eligibility Worksheet: EL". The form includes several sections with questions and checkboxes, where variable names are highlighted in grey boxes:

- Page 1 of 1**, 5/24/95, CLEL01.01, FORMVER
- EDITDATE & MOVEDATE are system supplied variables**
- Coordinating Center use only** (box): Rec'd, Log'd, Chk'd, Entr'd
- Clinic:** SITE (grey box)
- Date:** EXAMDATE (grey box), mm dd yy
- Use this form to screen patients** (checkbox)
- Send to Coordinating Center** (checkbox)
- 1. Inclusion Criteria:**
 - AGEGT12** (grey box): Is patient at least 12 years old? (checkboxes 1, 2, 9)
 - IRREGEYE** (grey box): At least **one eye** has irregular corneal surface such as: (checkboxes 1, 2, 9)
 - Keratometric mires
 - Retinoscopic reflex
 - Red reflex
 - DEFECT** (grey box): At least **one eye** has one of the following: (checkboxes 1, 2, 9)
 - Vogt's striae in the deep stroma
 - Fleischer's ring of at least 2mm arc
 - Corneal scarring characteristic of keratoconus
 - FOLLOW** (grey box): Able to complete 3 years of follow-up? (checkboxes 1, 2, 9)
- 2. Exclusion Criteria:**
 - TRANSPL** (grey box): Bilateral Corneal Transplants? (checkboxes 1, 2, 9)
 - CONDITN** (grey box): Does patient have any of the following ocular conditions in **both eyes**? (checkboxes 1, 2, 9)
 - Cataract
 - Intraocular lens implant
 - Macular disease
 - Optic nerve disease*, eg. optic neuritis, optic atrophy
 - *Eligible patients can have glaucoma
- 3. Consent Form:**
 - CONSENT** (grey box): Has patient signed Informed Consent Form? (checkboxes 1, 2)
 - CONSDATE** (grey box): If yes, give date consent form was signed: (checkboxes mm, dd, yy)
- 4. Does this patient appear to be eligible?** (checkboxes 1: Yes, 2: No, 9: Uncertain)
- 5. Give Pt. ID number and name code assigned to this patient:** IDNUM (grey box) -- NAMECODE (grey box)
- 6. Person completing form:** (first, last) OR Slit Lamp Examination Certification (checkboxes), Telephone (), Doctor responsible for patient's eye care

Figure 4: Annotated Form

We then assigned a mouse up function to each variable name, linking the variable name to the PDF file that displays the variable detail page (PDF) created by %ANNOTATE. The mouse up function opens the PDF file whenever the user clicks and releases the mouse button over the variable name. Figure 5 illustrates how, by clicking on the variable name CONSENT in the annotated form, the user can then view the variable detail page for that variable. The second arrow in the figure illustrates how the user can then click on the format/informat name (if any) to get details of the formatting of the variable's values.

3. Consent Form:

CONSENT Has patient signed Informed Consent Form? 1 2 30

CONSDATE If yes, give date consent form was signed : / / 31
mm dd yy

If ✓'s appear in any shaded box, the patient is **not** eligible.

4. Does this patient appear to be eligible? **ELIGIBLE** 1 **Yes** 2 **No** 3 **Uncertain** 37

Dataset is master.EL
Variable Name is CONSENT
Variable Label is 'Consent form signed'
Variable Type is C
Variable Format is YN - Click here for details
Variable Informat is Undefined

EXAMDATE(Date of CLEK Eligibility)	CONSENT(Consent form signed)		
	Yes	No	Total
1995	700	47	747
	53.85	3.62	57.46
	93.71	6.29	
	56.45	78.33	
1996	540	13	553
	41.54	1.00	42.54
	97.65	2.35	
	43.55	21.67	
Total	1240	60	1300
	95.38	4.62	100.00

Frequency Percent Row Pct Col Pct

```

-----
|          FORMAT NAME: SYN          LENGTH: 3  NUMBER OF VALUES: 3
| MIN LENGTH: 1  MAX LENGTH: 40  DEFAULT LENGTH 3  FUZZ: 0
|-----|-----|-----|-----|
| START          | END          | LABEL (VER. V7|V8 08SEP2003:13:51:38) |
|-----|-----|-----|-----|
| 1              | 1            | Yes                                     |
| 2              | 2            | No                                      |
| **OTHER**     | **OTHER**   | *                                       |
|-----|-----|-----|-----|

```

Figure 5: Linking to Variable Details

CONCLUSION

We believe the CLEK archive, and the tools used to build it, can serve as a model for documentation of any research study or business application, not only as an archive at the conclusion of the project, but as a readily accessible, easily useable reference while the project is on-going or under development. The tools we developed are adaptable to any SAS database and easily customized to meet any special circumstances.

We invite you to visit our web site at <https://vrcc.wustl.edu/clekarchive> to see the full scope of documentation included in our archive and to try out the interactive annotated forms for yourself.

HIPAA COMPLIANCE NOTE

The CLEK Study carefully guards the confidentiality of its participants and is in full compliance with the Health Insurance Portability and Accountability Act (HIPAA). The web archive does not include access to any raw study data, and no patient identifiers are included in the summary statistics displayed on the annotated form detail pages. All data made available to researchers, independently of the web site, is carefully cleansed of patient identifying information in full compliance with HIPAA prior to release.

ACKNOWLEDGMENTS

The Collaborative Longitudinal Evaluation of Keratoconus (CLEK) Study is supported by the National Eye Institute/National Institutes of Health, grants EY10419, EY10069, EY10077, EY12656, and EY02687. It also was supported by Conforma Contact Lenses, Paragon Vision Sciences, CIBA Vision Corporation, the Ohio Lions Eye Research Foundation, and the Research to Prevent Blindness Foundation.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Joel Achtenberg
Washington University School of Medicine
660 South Euclid, Campus Box 8203
St Louis, Missouri 63110
Phone: 314-362-6562
Fax: 314-362-0231
E-mail: joel@vrcc.wustl.edu

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.
Other brand and product names are trademarks of their respective companies.

APPENDIX: FULL TEXT OF ANNOTATE AND MAKEFMT MACROS, AND TYPICAL INVOCATION

```

*-----;
options nodate nonumber formchar='|----|+|---+=|-/\<>*' ;
%get_time; *--formats current date for use in titles/footnotes--;
footnote "Program: &progname   Prepared: &curdate";
libname MASTER "S:\MASTER";
%let basefile=           W:\CLEKARCHIVE\Website\AnnotatedForms;
%let basehref=file:///W:/CLEKARCHIVE/Website/AnnotatedForms;
ods listing close;
*-----;
*--CREATE MACRO ANNOTATE                               --;
*-----;
%MACRO ANNOTATE(LIB,DSN,DSNREF,STATSBY,EXCL);
%global nfmts ninfmts; %global existfmt existinf;
%let dset=&lib.&dsn;
%let nchar=0; %let nnum=0;
%let excl=%upcase(&excl);
%let statsby=%sysfunc(upcase(&statsby));
%let excl   =%sysfunc(upcase(&excl));
ods results off; ods trace off;
ods pdf file="&basefile.\&dsnREF._contents.pdf"
        style=styles.jfastyle bookmarklist=hide;
ods pdf title=" ";
ODS pdf exclude contents.dataset.enginehost;
ods noproctitle;
ods proclabel='';
title2;
proc contents data=&dset; run;
ODS PDF close;
%let dsid=%sysfunc(open(&dset));
%if &dsid %then %do;
    %let nfmts=0; %let ninfmts=0; *--initialize counters--;
    %let nvars=%sysfunc(attrn(&dsid,nvars));
    %do i=1 %to &nvars;
        %let existfmt=0; %let existinf=0;
        %let vname =%sysfunc(varname (&dsid,&i));
        %let vlabel=%sysfunc(varlabel(&dsid,&i));
        %let vlabel=%sysfunc(compress(&vlabel,%str(%')));
        %let vlen  =%sysfunc(varlen  (&dsid,&i));
        %let vtype =%sysfunc(vartype (&dsid,&i));
        %let vfmt  =%sysfunc(varfmt  (&dsid,&i));
        %if x&vfmt=x %then %let vfmt=Undefined;
        %else %do;
            *-- first remove $ and periods from the format--;
            %if %substr(&vfmt,1,1)=$ %then %let vfmt=%substr(&vfmt,2);
            %let vfmt=%substr(&vfmt,1,%eval(%length(&vfmt)-1));
            *--determine if format/informat is in the library--;
            %if &vtype=C %then
                %let existfmt=%sysfunc(exist(library.formats.&vfmt..formatc));
            %if &vtype=N %then
                %let existfmt=%sysfunc(exist(library.formats.&vfmt..format));
            *--run MAKEFMT macro to build PDF file for this format--;
            %if &existfmt %then %makefmt(&vfmt,&vtype,format);
        %end;
    %let vinfmt=%sysfunc(varinfmt(&dsid,&i));
    %if x&vinfmt=x %then %let vinfmt=Undefined;
    %else %do;
        %if %substr(&vinfmt,1,1)=$ %then %let vinfmt=%substr(&vinfmt,2);
        %let vinfmt=%substr(&vinfmt,1,%eval(%length(&vinfmt)-1));
        *--determine if format/informat is in the library--;
        %if &vtype=C %then
            %let existinf=%sysfunc(exist(library.formats.&vinfmt..infmtc));
        %if &vtype=N %then
            %let existinf=%sysfunc(exist(library.formats.&vinfmt..infmt));
        *--run MAKEFMT macro to build PDF file for this input format--;
        %if &existinf %then %makefmt(&vinfmt,&vtype,infmt);
    %end;
%end;

```



```

%end;
ODS PDF file="&basefile.\&dsnREF._&vname..PDF"
      style=jfastyle bookmarklist=hide startpage=never;
ods escapechar="^";
*--Use TITLES to add dataset and var details to freq/MEANS output--;
title1 j=c "*****";
title2 j=c "Dataset is &DSET";
title3 j=c "Variable Name is &VNAME" ;
title4 j=c "Variable Label is '&VLABEL'";
title5 j=c "Variable Type is &VTYPE";
*--TITLE statements for formats/informats include imbedded URLs--;
%let click=Click here for details;
%if (&existfmt) %then
  title6 j=c "^S={URL='format_&VFMT..PDF'} Variable Format is &vfmt-&Click" ;;
%if (&existinf=0) %then title6 j=c "Variable Format is &vfmt";;
%if (&existinf) %then title7 j=c
  "^S={URL='infmt_&VINfmt..PDF'} Variable Informat is &vinfmt - &Click" ;;
%if (&existinf=0) %then title7 j=c "Variable Informat is &vinfmt";;
title8 j=c "*****";
*--Summarize by value of specified STATS BY variable, usually by year--;
*--only do Stats summary if not excluded and if not a certification field*/
%if %index(x&excl,&vname)=0 and %sysfunc(index(&vname,CERT))=0 %then %do;
  %if &vtype=C %then %do;
    proc freq data=&dset;
      tables &STATSBY*&vname;
      %if &statsby=EXAMDATE or &statsby=DOD %then FORMAT &STATSBY YEAR4.;
    ;run;
  %end;
  %if &vtype=N %then %do;
    proc means data=&dset; var &vname;
      %if &statsby=EXAMDATE or &statsby=DOD
        %then FORMAT &STATSBY YEAR4.;
      %if &vfmt=:MMDDYY %then FORMAT &vname YEAR4.;
      ;CLASS &STATSBY;
    ;run;
  %end;
  %end;
  %else %do;
    data _null_; file print footnotes;
    put "No Statistical Summary for this variable"; run;
  %end;
  ODS PDF close; *--stop ODS output--;
%end;
%let rc=%sysfunc(close(&dsid));
%end;
%ELSE %PUT Attempt to open &dset failed - %sysfunc(sysmsg());
%MEND ANNOTATE;

*-----;
*--CREATE MACRO MAKEFMT(VFMT,VTYPE,FTYPE) - FTYPE is either FORMAT or INFMT--;
*-----;
%MACRO MAKEFMT(vfmt,vtype,ftype);
title; footnote;
%let file=&basefile.\&ftype._&vfmt..PDF ;
%let existfile=%sysfunc(fileexist(&file));
%if &existfmt and not &existfile %then %do;
  ODS PDF file="&file" style=jfastyle bookmarklist=hide;
  proc format fmtlib lib=library;
    %if &vtype=C %then select $&vfmt.; %else select &vfmt.;
  ; run;
%end;
ODS PDF close;
%MEND MAKEFMT;

```

```
*-----;
*-- TO RUN THE ANNOTATE MACRO FOR EACH SELECTED DATASET      --;
*-- PARAMETERS ARE LIBRARY, DATASET, STATS BY, and list of EXCLUDED Variables -;
*-- EXCLUDED VARS DO NOT GET A PROC FREQ OR MEANS           --;
*-----;
%annotate(master,DT,DT,dod      ,idnum namecode formdate editdate movedate dod);
%annotate(master,EL,EL,examdate,idnum namecode examdate editdate movedate consdate);
*--etc, for each dataset--;
```