ABSTRACT
At Glaxo Smith Kline we use a proprietary electronic data capture (EDC) tool to database clinical study data. Before authorizing a database, an internal audit is performed to compare the eCRF data printout from Microsoft Access tables with the processed data that is extracted from the EDC output into SAS data sets. The converted data sets are uploaded to the VAX for further processing, data checking, and printing. The data set print utility is written using SAS 6.08 on the VAX and is comprised of three macros that work in tandem to produce the printed output in the desired data sequence for a 10% random list of subjects to be sampled. The eCRF order of variables and data sets repeats in the printed output for each subject. This allows for a direct comparison between the original EDC tables and the SAS data sets found on the VAX.

INTRODUCTION
The extraction process we use involves several SAS programs to get the data from Access tables into SAS data sets. Subsequently, that data is processed using PC SAS by adding keys that are used in our standardized database structures and then normalizing it by transposing its original flat structure. Hence, this necessitates that we compare the data from the CRF in its electronic format (eCRF) to that of the extracted data after processing and placing it in SAS data sets. The EDC tool provides a utility to print the eCRF data in page order for comparison with the printouts of the final SAS data sets. This paper involves an overview of SAS programs written specifically to print the data from the various SAS data sets in page order and subject order comparable to those provided in the printed eCRFs for a formal audit of 10% of the subjects.

THE PROCESS
Three macros were written to do the task at hand: Subj_lst, Varlist, and Prtdat. Subj_lst provides the list of subjects in the order they need to be printed. Varlist provides the data sets and the variables that need to be printed for each data set in the requested order. Prtdat prints the SAS data for the data sets and subjects offered by the two other macros. These three macros work in tandem. The initial calling program that runs the whole process is Subj_lst. Below is the breakdown of each step.

SUBJECT LIST MACRO
The Subj_lst macro has two loops that loop as many times as there are subjects. The first reads the subject number from a macro variable containing the desired subject IDs separated by commas (Figs. 1 and 2). In the second loop each time a subject is taken from that list and provided to a Varlist macro call (Fig. 3).

```
%let subs= (12334, 54321, 33334, 12368, 17890, 18901);
```

Fig. 1 - Macro variable subs containing six subject IDs.

```
data _NULL_; do i=1 to 6; /*<--MODIFY FOR No.OF SUBJECTS*/
call symput('sub'||left(put(i,2.)),scan(symget('subs'),i)); end;
call symput('tot',i-1);
run;
```

Fig. 2 - First loop in Subj_lst.

```
%do i=1 %to &tot;
%varlist(&&sub&i);
%end;
```

Fig. 3 - Second loop in Subj_lst.

VARIABLE LIST MACRO
Macro Varlist defines macro variables containing a list of variables in the order to be printed for each data set (Fig. 4) which reflects the order of variables from the eCRF.

```
%let DEMOG=%str (SESS VSDT TMTNO DOB AGE SEX RACE HT HTU WT WTM))))
```

Fig. 4 - Defining the demography macro variable.

Moreover Varlist also calls the prtdat macro for each of the defined data sets to be printed. It passes the data set name, the subject id and the variable list to the prtdat macro. Figure 5 shows sample calls to prtdat for various data sets (note the order of calls to the prtdat macro provides for the sequence in which the data sets are printed.

```
%prtdat(SCRFLR,&&sub,&SCRFLR);
%prtdat(DEMOGSF,&&sub,&DEMOGSF);
%prtdat(ELIG,&&sub,&ELIG);
%prtdat(DEMOG,&&sub,&DEMOG);
%prtdat(DBHX,&&sub,&DBHX);
%prtdat(DB,&&sub,&DB);
%prtdat(MC,&&sub,&MC);
%prtdat(MCTX,&&sub,&MCTX);
```

Fig. 5 - Calling prtdat.

THE PRINT MACRO
The print macro Prtdat essentially is a print procedure that prints the data sets and provides an appropriate title for each subject and data set whose data is being printed (Fig. 6).
CONCLUSION
Hence the three macros work together to produce the printouts for a given list of subjects, data sets, and variables within those data sets. The figure below shows how one macro calls the other. Follow the subject number as it is given by the subj_lst to the varlist macro, which in turn passes it on to the prtdat macro. The varlist macro also passes the variables for each data set to be printed (Fig 7).

ACKNOWLEDGMENTS
We wish to acknowledge Lori Hall, Shannon Gooding, Pat Borgia and Nellie Hudson who worked with the output of these efforts.

CONTACT INFORMATION
Your comments and questions are valued and encouraged. Contact the authors at:

Michael Acra
Senior Medical Database Programmer
Glaxo Smith Kline
5 Moore Drive
Research Triangle Park, NC 27709
Phone: (919)483-5435
Email: ma39288@gsk.com

Greg Henkle
Senior Medical Database Programmer
Glaxo Smith Kline
5 Moore Drive
Research Triangle Park, NC 27709
Phone: (919)483-5534
Email: glh39233@gsk.com

Web site: www.gsk.com

Note that these macros working together allow for printing the all the CRF data in CRF order for each of the subjects in subject order, thus allowing a direct comparison between original eCRF data and final SAS data (see Fig. 8)

Fig. 6 – The print macro.

Fig. 7 – The three macros.

Fig. 8 – Comparing eCRF data to SAS data.