Paper 30-25

Generating Case Report Form Tabulation

David Wade, SmithKline Beecham Pharmaceuticals,Collegeville, PA
Shi-Tao Yeh, EDP Contract Services, Bala Cynwyd, PA

ABSTRACT

This paper describes a novel and efficient algorithm for producing a Case Report Form Tabulation (CRT) listing. Where for the purposes of this paper, the authors define a CRT listing as a clinical report that interleaves a series of individual data listings (e.g., demographic characteristics, laboratory data, and vitals) by patient.

The approach defined in this paper applies an efficient data integration process, and an important page consolidating post-processing algorithm.

The SAS® product used in this paper is SAS BASE® on a UNIX platform.

INTRODUCTION

SAS programmers in the pharmaceutical industry support the clinical data analysis and reporting process. One of the more frequently requested reports is the Case Report Form Tabulation (CRT). It is a clinical report that interleaves individual data listings (e.g. demographic characteristics, laboratory data, and vitals) together for each patient.

There are a number of methods to produce a CRT. A typical CRT listing program creates intermediate SAS data sets, and stores them in a temporary SAS Library. Then the reporting routine reads in the temporary data to interleave each patient's information in the listings.

This paper presents an alternative approach that is simpler, more efficient, and produces a smaller number of pages of output. The approach skips the previously described first step of creating intermediate SAS data sets, and replaces it with an easier process for CRT reporting. The new approach consists of five steps, and it is discussed in the following sections using sample SAS code.

DATA FILE

For illustrative purposes, five data files from a hypothetical clinical trial are used in this paper. The variables in each data file are:

Data File Name	Variable List
su_dem	pid, dem_race, dem_racet, dem_sex, dem_age, dem_wt, dem_ht, dem_bsa
su_mhist	pid, mh_diag, mh_pref, mh_disc1, mh_pcb
su_cvhis	pid, cv_vis, cv_dat, cv_disc, cv_dis, cv_tpp, cv_ansyn
su_trx	pid, trx
su_chfcs	pid, cc_iscm, cc_cauc, cc_cau, cc_diagd, cc_ncp

STEP 1: DEFINE HIT LIST

A hit list data set identifies the subset of patients that will be included in the CRT listings.

```
%macro hitlist;

data hitlist ;
  set mart.trx;
  if trx ^= ' ';
  label trxl = 'TREATMENT GROUP';
  keep pid  trx;
proc sort;by pid;

run ;

%mend hitlist;
```

STEP 2: CREATE WORK DATA SET

The data are read from a standard Data Mart. The Data Mart is a collection of SAS data sets that is defined by a standard data structure for clinical reporting needs.

```
%macro crt_dsug;

%macro setdata(dsin=, dsout=);
data &dsout;
 set mart.&dsin;

proc sort;by pid;

data &dsout;
 merge &dsout hitlist;
 by pid;
run;
%mend setdata;

%setdata(dsin=su_dem, dsout=dem1);
%setdata(dsin=su_mhis, dsout=mhis1);
%setdata(dsin=su_cvhis, dsout=cvhis1);
%setdata(dsin=su_chfcs, dsout=chfcs1);

%mend crt_dsug;
```

STEP 3: SET UP REPORTING PROGRAM

The primary reporting SAS macro, **crt_psug**, consists of two SAS macros. The first SAS macro defines all of the customized report display items needed. If the user does not create this macro, then PROC REPORT will apply default characteristics. The second macro, **crt_psug**, is used to read the data and to display it with a DO loop for selecting one patient at a time.

```
%macro crt_psug;
%macro def;
define dm_race / 'Race' width=10 ;
define dm_racet / 'Race(Specify)' width=15 ;
define dm_sex  / 'Sex' width=8  ;
define dm_age  / format=3.0 width=9 'Age(Yrs)' ;

define dm_wt   / format=6.0 width=11
'Weight(kg)' ;
define dm_ht   / format=6.0 width=11
'Height(cm)' ;
define dm_bsa   / format=6.3 width=8 'BSA(M2)' ;
   .
   .
 /* more define statements */
   .
   .
%mend def;

options missing=' ';
%do jj = 1 %to &pnum;
proc report data=dem1 headline  split='^';
     by &sortby pid;
     column ( "==" dm_race dm_racet dm_sex
dm_age dm_ht dm_wt dm_bsa );
     where pid= "&&p&jj";
     %def;
        run;
   .
   .
  /* more PROC REPORT statements  */
   .
   .

%end;
run;
%mend  crt_psug;
```

STEP 4: EXECUTE MAIN PROGRAM

The main program controls execution of the individual macros, and populates the macro variables needed for each patient (pid).

```
%macro crt00(hitlist=, crtmacro=, sortby=) ;

proc sort data=&hitlist out=_crt(keep=pid)
nodupkey ;
  by &sortby pid; run ;

*****************************************;
* Create the macro variables for each pid ;
*****************************************;
%let pnum=0;
data _null_ ;
  set _crt nobs=_nobs ;
  call symput("p" || compress(put(_n_, 10.)),
pid) ;
  call symput("t" || compress(put(_n_, 10.)),
&sortby ) ;
  call symput("pnum", put(_nobs, 4.));
run ;


%&crtmacro ;


%mend crt00;
```

PROGRAM INVOCATION AND OUTPUT

The program is invoked by defining the SAS library where the input data are stored, the hit list of patients required for the CRT, and the main program macro calls.

These individual components are simplified in this example for illustrative purposes. However, the authors developed the approach to be implemented within a typical complex SAS reporting environment.

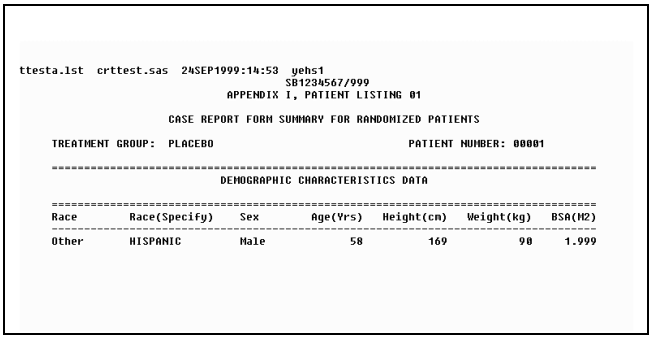
```
libname mart '/home/yehs1/mart';


%hitlist ;

%crt_dsug;


%crt00(crtmacro=crt_psug, sortby= trx ) ;
run;
```

Samples of Three pages of the raw output from Step 4 are presented in Figures 1 through 3. Note that only a small portion of each page is filled at this point.



ttesta.lst crttest.sas 24SEP1999:14:53 yehs1
SB1234567/999
APPENDIX I, PATIENT LISTING 01

CASE REPORT FORM SUMMARY FOR RANDOMIZED PATIENTS

TREATMENT GROUP: PLACEBO PATIENT NUMBER: 00001

DEMOGRAPHIC CHARACTERISTICS DATA

Race	Race(Specify)	Sex	Age(Yrs)	Height(cm)	Weight(kg)	BSA(M2)
Other	HISPANIC	Male	58	169	90	1.999

Figure 1 Demography Output.

ehs1/code/crttesta.lst crttest.sas 24SEP1999:14:53 yehs1
SB1234567/999
APPENDIX I, PATIENT LISTING 01

CASE REPORT FORM SUMMARY FOR RANDOMIZED PATIENTS

TREATMENT GROUP: PLACEBO PATIENT NUMBER: 00001

PRESENTING CONDITIONS DATA

Verbatim Term	Preferred Term	Disease Classification (Level 1)
ABDOMINAL AORTIC ANEURYSM	ANEURYSM, AORTIC	CIRCULATORY SYST
ABDOMINAL AORTIC ANEURYSM GRAFT REPAIR	OPERATION, OTHER VESSELS	OPERATIONS
ARTHRITIS	ARTHROPATHY	MUSCULOSKEL/CONNECT TISSUE DIS
BOWEL OBSTRUCTION	INTEST OBSTRUCTION	DIGESTIVE SYST
HERNIA	HERNIA, ABDOMINAL	DIGESTIVE SYST
HYPERCHOLESTEROLEMIA	CHOLEST/TRIGLYCERIDE, ELEVATED	ENDOCR/METAB/IMMUNITY DISORD

Figure 2 Presenting Conditions Output.

ttesta.lst crttest.sas 24SEP1999:14:53 yehs1
SB1234567/999
APPENDIX I, PATIENT LISTING 01

CASE REPORT FORM SUMMARY FOR RANDOMIZED PATIENTS

TREATMENT GROUP: PLACEBO PATIENT NUMBER: 00001

CV HISTORY DATA

OBSERVATION	DATE		CV SYMPTOM/HISTORY	STATUS	ANSWER
SCREENING	16MAR1999	1	Angina	Treated Presently	Yes
		2	Hypertension	Treated Presently	Yes
		3	Diabetes,(insulin dependent)	Treated Presently	No
		4	Diabetes,(non-insulin dependent)	Treated Presently	No
		6	Stroke	Had In Past	No
		7	TIA	Had In Past	No
		8	Peripheral vascular disease	Had In Past	No
		9	CABG	Had In Past	Yes
		10	PTCA	Had In Past	No
		11	Pacemaker fitted	Had In Past	No
		12	Implantable cardiac defibrillator	Had In Past	No
		13	Peripheral vascular surgery	Had In Past	No
		14	Carotid endarterectomy	Had In Past	No

Figure 3 Cardiovascular History Output.

STEP 5: PERFORM POST PROCESSING

The final step of this approach uses a post-processing method to reduce the number of pages of output. Step 5 consolidates the output across pages wherever possible, while maintaining clarity of presentation.

As noted above, the raw output from Step 4 is a series of individual PROC REPORT outputs. This type of presentation begins a new page for each PROC REPORT statement regardless of the amount of remaining space on the page. Thus, one may generate many thousands of pages of output having only a portion of each page filled with text.

The post-processing objective is to read the raw list file output from Step 4, and convert it into the final product. It consolidates more than one PROC REPORT table onto single pages where they will fit, and splits PROC REPORT tables across pages only where the end of a page of output is encountered.

Our post-processing algorithm encompasses two relatively simple SAS macros which, using the following approach can be readily created for a specific user's needs. We found our two specific macros to be capable of processing a 15,000 page CRT listing in less than 1 minute on our particular SAS system.

The first of the two SAS macros uses a single SAS DATA STEP to read through the text file from Step 4 and gather a small amount of information. The macro call has a single parameter that specifies the name of the text file to be processed (shown as step4.lst in the following example:

```
%igcrt( inname = ./step4.lst);
```

The first macro gathers the following information from the text file and places the results into three macro variables using CALL SYMPUT:

- The line number that contains the page number is identified. This is found as simply the first non-blank line on each page.
- The line number of the last line of the page header is identified. This is found as the last line at the top of the pages that is constant across all pages.
- The maximum number of lines for a page in the input file is identified. The SAS GOPTION PAGESIZE value could also be used for the same purpose.

The second of the two SAS macros uses another relatively simple DATA STEP to read through the text file from Step 4 and consolidate it. This second macro uses the three pieces of information gathered and stored as macro variable values. The second macro call has two parameters that specify only the name of the text file to be processed and the name of the text file to be output:

```
%ppcrt( inname = ./step4.lst,
        outname = ./final.lst);
```

- The second macro reads the constant header information for all pages from the first page of the output. The line number of the last line of the page header was identified by the first macro. Following this, the header is skipped over as the macro reads the rest of the pages.
- Starting with the first page and repeating whenever a new page is printed to the output

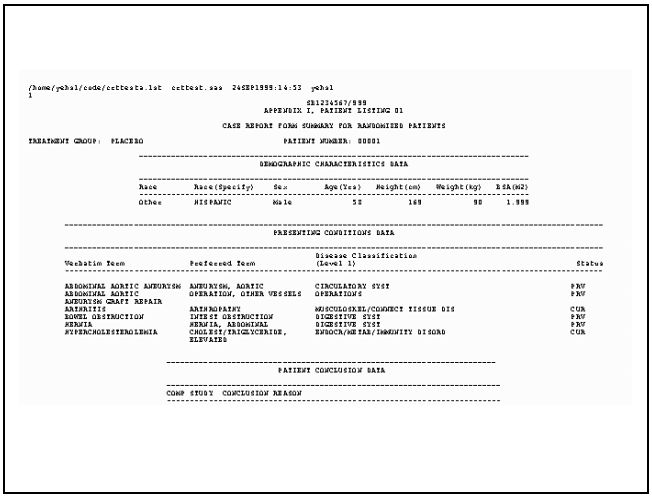
file, the macro prints the page header information on the top of each page, and pages are sequentially renumbered as each new page is output. The line number of the page number line was identified by the first macro.

- Once a header is skipped over on an input page, the macro then reads each line from the input file and prints it to the output file. The macro keeps track of the number of lines of output written to the current page of output.

- If the number of lines output written to a page equals the maximum number of lines per page identified by the first macro, then a new page of output is started.

- New pages are also started whenever a new BY line value is encountered (e.g., patient). The line number of the BY line was identified by the first macro.

Figures 4 through 6 present 3 pages of sample output from Step 5 for two patients. Users of this approach would likely wish to customize the format of the final outputs to meet their specific needs. For example, in Figures 4 and 6,we have chosen to present missing patient conclusion data as a set of headers with no data rows below them.



/home/yehal/code/cctteats.lst cctteat.sas 24SEP1999:14:53 yehal
1

SR1234567/999
APPENDIX I, PATIENT LISTING 01

CASE REPORT FORM SUMMARY FOR RANDOMISED PATIENTS

TREATMENT GROUP: PLACEBO PATIENT NUMBER: 88881

DEMOGRAPHIC CHARACTERISTICS DATA

Race	Race(Specify)	Sex	Age(Yrs)	Height(cm)	Weight(kg)	BSA(M2)
Other	HISPANIC	Male	52	168	88	1.998

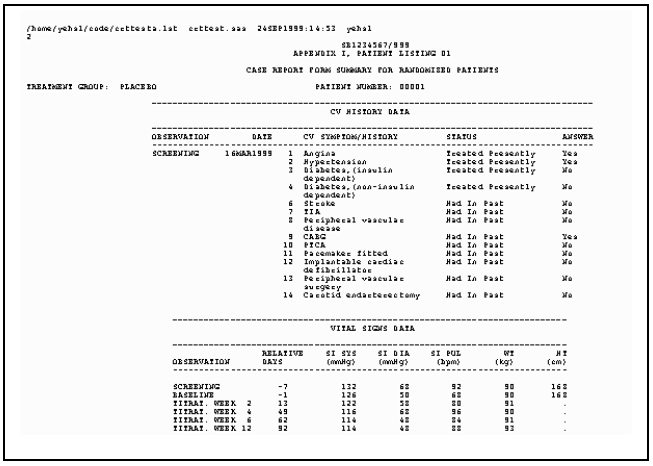
PRESENTING CONDITIONS DATA

Verbatim Term	Preferred Term	Disease Classification (Level 1)	Status
ABDOMINAL AORTIC ANEURISM	ANEURYSM, AORTIC	CIRCULATORY SYST	PAV
ABDOMINAL AORTIC ANEURISM GRAFT REPAIR	OPERATION, OTHER VESSELS	OPERATIONS	PAV
ARTHRITIS	ARTHROPATHY	MUSCULOSKEL/CONNECT TISSUE DIS	CUR
BOWEL OBSTRUCTION	INTEST OBSTRUCTION	DIGESTIVE SYST	PAV
HERPIA	HERNIA, ABDOMINAL	DIGESTIVE SYST	PAV
HYPERCHOLESTEROLEMIA	CHOLEST/TRIGLYCERIDE, ELEVATED	ENDOCR/METAB/IMMUNITY DISORD	CUR

PATIENT CONCLUSION DATA

COMP STUDY CONCLUSION REASON

Figure 4 Sample output page 1.



/home/yehal/code/cctteats.lst cctteat.sas 24SEP1999:14:53 yehal
2

SR1234567/999
APPENDIX I, PATIENT LISTING 01

CASE REPORT FORM SUMMARY FOR RANDOMISED PATIENTS

TREATMENT GROUP: PLACEBO PATIENT NUMBER: 88881

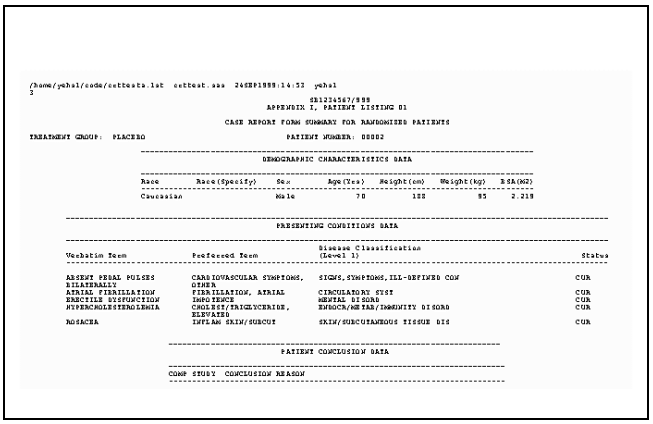
CV HISTORY DATA

OBSERVATION	DATE		CV SYMPTOM/HISTORY	STATUS	ANSWER
SCREENING	16MAR1999	1	Angina	Treated Presently	Yes
		2	Hypertension	Treated Presently	Yes
		3	Diabetes,(insulin dependent)	Treated Presently	No
		4	Diabetes,(non-insulin dependent)	Treated Presently	No
		5	Stroke	Had In Past	No
		7	TIA	Had In Past	No
		8	Peripheral vascular disease	Had In Past	No
		9	CABG	Had In Past	Yes
		10	PTCA	Had In Past	No
		11	Pacemaker fitted	Had In Past	No
		12	Implantable cardiac defibrillator	Had In Past	No
		13	Peripheral vascular surgery	Had In Past	No
		14	Carotid endarterectomy	Had In Past	No

VITAL SIGNS DATA

OBSERVATION	RELATIVE DAYS	SI SYS (mmHg)	SI DIA (mmHg)	SI PUL (bpm)	WT (kg)	HT (cm)
SCREENING	-7	132	82	92	98	182
BASELINE	-1	126	78	88	98	182
TIRAT. WEEK 2	13	122	72	88	91	.
TIRAT. WEEK 4	45	116	68	86	90	.
TIRAT. WEEK 8	62	114	68	84	91	.
TIRAT. WEEK 12	92	114	68	82	93	.

Figure 5 Sample Output Page 2.



/home/yehal/code/cctteats.lst cctteat.sas 24SEP1999:14:53 yehal
3

SR1234567/999
APPENDIX I, PATIENT LISTING 01

CASE REPORT FORM SUMMARY FOR RANDOMISED PATIENTS

TREATMENT GROUP: PLACEBO PATIENT NUMBER: 88882

DEMOGRAPHIC CHARACTERISTICS DATA

Race	Race(Specify)	Sex	Age(Yrs)	Height(cm)	Weight(kg)	BSA(M2)
Caucasian		Male	70	182	95	2.218

PRESENTING CONDITIONS DATA

Verbatim Term	Preferred Term	Disease Classification (Level 1)	Status
ABSENT PEDAL PULSES BILATERALLY	CARDIOVASCULAR SYMPTOMS, OTHER	SIGNS,SYMPTOMS,ILL-DEFINED CON	CUR
ATRIAL FIBRILLATION	FIBRILLATION, ATRIAL	CIRCULATORY SYST	CUR
ERECTILE DYSFUNCTION	IMPOTENCE	MENTAL DISORD	CUR
HYPERCHOLESTEROLEMIA	CHOLEST/TRIGLYCERIDE, ELEVATED	ENDOCR/METAB/IMMUNITY DISORD	CUR
ROSACEA	INFLAM SKIN/SUBCUT	SKIN/SUBCUTANEOUS TISSUE DIS	CUR

PATIENT CONCLUSION DATA

COMP STUDY CONCLUSION REASON

Figure 6 Sample Output Page 3.

For our tested sample CRT documents, the post-processing algorithm was able to reduce the number of pages of output by 50%. Thus, saving many thousands of pages of paper. The resulting consolidated document was also considered to be more easily interpretable.

CONCLUSIONS

Using this new approach, we were able to produce CRT listings quickly and efficiently. The post-processing algorithm has wide-ranging potential application for SAS reporting.

This approach provides the following features:

* Avoids storage of the intermediate data files, by using work files directly.
* Reduces the size and complexity of the program for better understanding of the approach and program maintenance.
* Introduces post processing to rearrange the SAS output, and achieve a page saving rate of approximately 50%.
* Provides a useful framework for the rapid development and modification of a CRT program.

TRADEMARKS

AUTHORS

David Wade, M.S.
(610)917-6923(W)
E-mail:
david_wade@prodigy.net

SmithKline Beecham Pharmaceuticals

Shi-Tao Yeh, Ph. D
(610)917-5883(W)
E-mail:
shitaoyeh@us.sina.com

EDP Contract Services™