



GCE SOLUTIONS

Derive Value From Excellence ...



Convert the Clinical SAS
datasets from RTF
files (CT08)

Murali Neela, GCE
solutions, India



Introduction :

- RTF(Rich Text Format) table file format in Pharmaceutical Industry
- Easy insertion of Tables in clinical study reports
- Special formatting of Table contents eg; bolding of significant p-values
- Validation of Table/listings programmatically
- Comparing multiple version of the same RTF
- Availability of additional fine format for external raw data



Uses of after converting the RTF file to DATASETS

- Removes all of the embedded RTF formatting tags
- Creates a SAS dataset that has a separate observation for every line of text in each RTF file
- Eliminating all unrelated rows and transforming the file to create the primary final dataset with variables col1 to coln and titles



Creating the RTF(Rich text format) output using Proc Report

```
proc report data=testdata.final nowd spacing=1 split='*' headline;  
  columns ('--' subjid bthdt age sex newrace ethnic height weight);  
  
  define subjid / order width=10;  
  define bthdt / display 'Birth Date' width=10;  
  define age / display width=7 'Age*(years)';  
  define sex / display width=6 'Sex';  
  define newrace / display width=9 flow 'Race';  
  define ethnic / display width=10 flow 'Ethnicity';  
  define height / width=6 'Height*(cm)';  
  define weight / width=6 'Weight*(kg)';  
  
  title2 'Listing of demographics and baseline characteristics';  
  title3 'Full Analysis Set';  
  
run;
```



Output file (RTF)

The SAS System
Listing of demographics and baseline characteristics
Full Analysis Set

Subject ID	Birth Date	Age (years)	Sex	Race	Ethnicity	Height (cm)	Weight (kg)
200-002	1983-02-02	29	FEMALE	ASIAN	NOT HISPANIC OR LATINO	193.3	93.8
300-003	1964-03-03	48	MALE	WHITE	NOT HISPANIC OR LATINO	190.0	90.7
400-004	1975-04-04	37	FEMALE	BLACK OR AFRICAN AMERICAN	NOT HISPANIC OR LATINO	186.7	87.6
500-005	1985-05-05	27	MALE	WHITE	NOT HISPANIC OR LATINO	183.3	84.4
600-006	1980-06-06	32	FEMALE	WHITE	NOT HISPANIC OR LATINO	180.0	81.3
700-007	1954-07-07	58	MALE	BLACK OR AFRICAN AMERICAN	NOT HISPANIC OR LATINO	176.7	78.2
800-008	1988-08-08	23	FEMALE	WHITE	NOT HISPANIC OR LATINO	173.3	75.1
900-009	1990-09-09	22	MALE	WHITE	NOT HISPANIC OR LATINO	170.0	72.0



Sample code

```
filename output "&output";
data output;
    retain page 1;
    infile output missover length = 1 lrecl = 2000;
    input string $varying2000. 1;
    if index(string, '\par {\f0\fs18}') > 0;
    /** remove rtf header information **;
    length text $200.;
    text = substr(string, 16);
    if substr(text, length(text)) = '}' then substr(text, length(text)) = '';
    /** find last footnote **;
    _n + 1;
    if find(text, '/report/') = 1 or find(text, '/-` _ _ _ ') = 1 or
find(text, '/` _ _ _ ') = 1 then do;
        _foot = 1;
        page + 1;
    end;
    if lag(_foot) = 1 then _n = 1;
    if _foot = 1 then call symputx('ps', _n);
    /** mark page category discription **;
    %if %length(&rowkp) > 0 %then %do; if _n + 1 in (&rowkp) then _label = 1; %end;
    %if %length(&page) > 0 %then %do; if mod(page, &pg_s) = &pg_t; %end;
run;
```



After using macro created SAS dataset

	col1	col2	col3	col4	col5	col6	col7	col8
1	200-002	1983-02-02	29	FEMALE	ASIAN	NOT HISPANIC OR LATINO	193.3	93.8
2	300-003	1964-03-03	48	MALE	WHITE	NOT HISPANIC OR LATINO	190.0	90.7
3	400-004	1975-04-04	37	FEMALE	BLACK OR AFRI...	NOT HISPANIC OR LATINO	186.7	87.6
4	500-005	1985-05-05	27	MALE	WHITE	NOT HISPANIC OR LATINO	183.3	84.4
5	600-006	1980-06-06	32	FEMALE	WHITE	NOT HISPANIC OR LATINO	180.0	81.3
6	700-007	1954-07-07	58	MALE	BLACK OR AFRI...	NOT HISPANIC OR LATINO	176.7	78.2
7	800-008	1988-08-08	23	FEMALE	WHITE	NOT HISPANIC OR LATINO	173.3	75.1
8	900-009	1990-09-09	22	MALE	WHITE	NOT HISPANIC OR LATINO	170.0	72.0

	text	group
1	Listing of demogr...	1004
2	Full Analysis Set	1008
3	--	2003
4	Subject ID	2014
5	Birth Date	2015
6	Age (years)	2016
7	Sex	2017
8	Race	2018
9	Ethnicity	2019
10	Height (cm)	2020
11	Weight (kg)	2021
12	--	3004



PROC COMPARE AND RESULTS

```
proc compare base=dev compare=val listall;  
run;
```

Observation Summary

Observation	Base	Compare	ID
First Obs	1	1	STUDYID=CRAD001H2307E1 USUBJID=CRAD001H2307E1_4020002 EXTRT=EVEROLIMUS EXSTDTC=2016-04-21 EPOCH=EXTENSION TREATMENT 1 EXDOSTOT=2 EXENDTC=
Last Obs	77	77	STUDYID=CRAD001H2307E1 USUBJID=CRAD001H2307E1_4028003 EXTRT=TACROLIMUS EXSTDTC=2017-09-28 EPOCH=EXTENSION TREATMENT 1 EXDOSTOT=2 EXENDTC=

Number of Observations in Common: 77.

Total Number of Observations Read from WORK.DEX_D: 77.

Total Number of Observations Read from WORK.DEX_Q: 77.

Number of Observations with Some Compared Variables Unequal: 0.

Number of Observations with All Compared Variables Equal: 77.

NOTE: No unequal values were found. All values compared are exactly equal.

