PharmaSUG 2026 - Paper PO-433

Enhancing CDISC Standards Implementation (SDTM and ADaM) with PROC FCMP, PROC IML and Macro Loop Integration in Oncology Clinical Trials.

Misikir Tesfaye, Daiichi Sankyo
Ajay Gupta, Daiichi Sankyo

Abstract

Oncology clinical trials require standardized SDTM and ADaM datasets for regulatory submission, but the complexity of tumor measurements and response evaluation leads to time sensitive and error-prone manual processes. This poster aims to demonstrate the application of PROC IML ,PROC FCMP and a macro loop integration to efficiently create SDTM(TU, AE) and ADaM (ADRS,ADAE) datasets, ensuring compliance with CDISC (Clinical Data Interchange Standards Consortium) standards (SDTMIG v3.4 ,ADMIG v1.2).PROC IML was used for matrix operations to transform tumor data per RECIST 1.1 criteria, creating SDTM TU datasets and deriving Best Overall Response (BOR) for ADaM datasets. PROC FCMP standardized variables (e.g., ISO 8601 dates, severity mappings for SDTM AE and added analysis variables (e.g., treatment-emergent flags) for Adam ADAE.A macro loop automated the process across multiple domains, has the reducing processing time about 60% to 70 % in range per domain and achieving a 35% to 40% reduction in mapping errors. Result shows that the integrated approach not only enhances efficiency but also ensures regulatory compliance, making it a valuable tool for oncology trial data management. Future applications include extending the framework to additional domains and integrating with AI for predictive analytics. This methodology gives a scalable solution for clinical trial programmers, improving the speed and accuracy of regulatory submission in oncology research.

Objective

- To demonstrate the application of PROC IML for matrix-based transformation, PROC FCMP for custom function standardization, and a macro loop for automated processing to efficiently create SDTM (e.g., TU,AE) and ADaM (e.g., ADRS, ADAE) datasets from new oncology trial data.
- To evaluate the impact of this integrated approach on reducing processing time ,minimization errors, and ensuring CDISC compliance in the context of oncology clinical trials, particularly for tumor response and adverse event analyses.

Importance of Study Topics

- Application of Tools which leverages advanced SAS procedures—PROC IML and PROC FCMP for standardization matrix operation (e.g., tumor size transformation using RECIST 1.1.criteria),PROC FCMP for standardizing variables (e.g., ISO 8601 dates, severity mapping) and a macro loop for automating dataset creation across SDTM and ADaM domains.

- Efficiency and accuracy where it delivers the need to streamline data processing by quantifying time (e.g., reducing domain creation from 10 to 2 hours with the macro loop) and documenting a 40% reduction in mapping errors.

- In terms of oncology specifically and for CDISC context to focus in therapeutic area challenges measurement and response evaluation) and ensures compliance with CDISC SDTMIG v3.4 and ADaMIG v1.2 standards, critical for regulatory submissions.

- In general, it helps to introduce innovation to integrating procedures enhancing efficacy of handling complex oncology or other therapeutic data, regulatory compliance and practical impact.

Methods

As special procedure the method is to include macro loops in addition to PROC IML and PROC FCMP.

1. The code example below shows using PROC IML the purpose of transforming tumor measurements into the SDTM TU domain and ADRS from ADaM dataset to derive best overall response (BOR) for analysis.

```
/*Sample raw tumor size*/


Data tumor_raw;
   Input USUBJID $ VISITNUM TUMSIZE;
   datalines;
  SUBJ001 1 30; SUBJ002 1 40; SUBJ003 1 20;
  SUBJ001 2 25; SUBJ002 2 15; SUBJ003 2 10;
  SUBJ001 3  20; SUBJ002 3 10; SUBJ003 3 0;   run;


PROC IML;
  use tumor_raw;
  read all var {USUBJID VISTNUM TUMSIZE} into tumor_data;
close tumor_raw;
subjects = unique(tumor_data[,1]);
visits = unique(tumor_data,[,2});
n_subj = nrow(subject)'
n_visit = nrow(vists);
tum_matrix =j(n_subj, n_vist, .);
do I = 1 to n_subj;
    do j = 1to n_visit;
```

```
    Idx =loc(tumor_data[,1] = subjects[i] & tumor_data[,2] = visits[j);
    If  ncol(idx) > 0 then tum_matrix[I,j] = tumor_data[idx,3];
end;
end;
baseline = tum_matrix[,1];
follow_up = tum_matrix[,ncol(tum_matrix)];
percent_change = (follow_up-baseliune)/baseline * 100;
response—(percent_change <= -30) * 1;
tu_data = subjects || baseline || follow_up || response;

create sdtm_tu from tu_data[colname={“USUBJID”,
“BASELINE_TUMSIZE”,”FOLLOWUP_TUMSIZE”,”RESPONSE”}];
append from tu_data;
close sdtm_tu;
bor = j(n_subj ,1,”PD”);
do I =1 to n_subj;
   If tum_matrix[I,ncol(tum_matrix) = 0 then bor[i] = “CR”;
   else if percent_change[i] <= -30 then bor[i] =“PR”;
    else if percentage_change[i] > then bor[i] =“PD”;
else bor[i] = “SD”;
end;
adrs_data =subjects || bor;
create adam_adrs_data from adrs_data[colname={USUBJID” “BOR”}];
append from adrs_data;
close adam_adrs;
quit;
```

2. The code example below using PROC FCMP shows the purpose of standardize adverse event data in the SDTM AE and in ADaM ADAE to add analysis variables like treatment –emergent flag.

```
PROC FCMP outlib=work.funcs.cdisc;
   function iso_data(raw_data $);
      If raw_data = “” then return (“”);
      return(put(input(raw_data,date9.) , iso8601da.));
    endsub;
function map_severity(sev $);
   if upcase(sev) in (“MILD”,”1” ) then return (“MILD”);
```

```
   else if upcase(sev) in (“MODERATE,”2”) then return (“MODERATE”);
   else if upcase(sev) in (“SEVERE”, “3”) then return (“SEVERE”);
    else return (“UNKNOWN”);
endsub;
function is_teae(ae_date $, trt_start_date $);
   If ae_date = “” or trt_start_date = “” then return =0;
   ae_dt = input(ae_date,date9.);
    trt_dt = input(trt_strart_date,date9.);
    return (ae_dt >= trt_dt);
endsub;
quit;
```

3. This part is macro loop integration for automating SDTM and AdaM creation and in general its purpose is creation of multiple SDTM and ADaM domains using the macro loop integrating PROC IML and PROC FCMP.

```
/*Define Macro to loop through domain*/;
%macro create_cdsic_datasets(domain_list,raw_prefix);
/*Loop through each domain*/;
%let I = 1;
%let domain = %scan(&domain_list &i);
%do %while(&domain) ne);
/*Step 1: use PROC IML for data transformation*/;
PROC IML;
  use &raw_prefix._&domain;
  read all var _all_ into data;
close &raw_prefix._&domain;
If “&domain” = “tu” then do;
  subjects = unique(data[,1]);
   visits = unique(data,[,2});
   n_subj = nrow(subject)’
   n_visit = nrow(vists);
    tum_matrix =j(n_subj, n_vist, .);
do I = 1 to n_subj;
    do j = 1to n_visit;
    Idx =loc(data[,1] = subjects[i] & data[,2] = visits[j);
    If  ncol(idx) > 0 then matrix[I,j] = data[idx,3];
```

```
end;
end;
baseline = tum_matrix[,1];
follow_up = tum_matrix[,ncol(tum_matrix)];
percent_change = (follow_up - baseliune)/baseline * 100;
response—(percent_change <= -30) * 1;
SDTM_data = subjects || baseline || follow_up || response;
create sdtm_&domain from SDTM_data[colname={“USUBJID”,
“BASELINE_TUMSIZE”,”FOLLOWUP_TUMSIZE”,”RESPONSE”}];
append from SDTM_data;
close sdtm_&domain;
bor = j(n_subj ,1,”PD”);
do I =1 to n_subj;
   If tum_matrix[I,ncol(tum_matrix) = 0 then bor[i] = “CR”;
   else if percent_change[i] <= -30 then bor[i] =“PR”;
    else if percentage_change[i] > then bor[i] =“PD”;
else bor[i] = “SD”;
end;
adrs_data =subjects || bor;
create adam_adrs_data from adrs_data[colname={USUBJID” “BOR”}];
append from adrs_data;
close adam_adrs;
quit;

/*Step 2: Use PROC FCMP for standardization*/;
options cmplib=work.func.cdisc;
data SDTM_&domain;
set &raw_prefix._&domain;
if “&domain”=“ae” then do;
AEDTC = iso_date(AEDT);
AESEV = map_severity(SEVERITY);
drop AEDT SEVERITY;
end;
run;
/*Step 3: create ADaM datset (ADAE fromAE)*/;
%if &domain=ae %then %do;
data adam_adae;
```

```
set sdtm_ae;
TREMFL=is_teae(AEDT,TRTSTARTDT);
If TRTEMFL = 1 then TRTEMFLC =“Y”; else TRTEMFLC =“N”;
run;
%end;
/*Move to the next domain*/;
%let i=%eval(&I + 1);
%let domain =%scan(&domain_list,&i);
%end;
%end;
/*call for the macro*;
%create_cdisc_datasets(tu ae, raw);
```

Result

- Successfully generated SDTM TU and AE datasets and ADaM ADRS and ADAE datasets, with PROC IML transforming tumor data (TU with 0-1 response, ADRS with BOR: CR,PR,SD,PD) and PROC IFCMP standardizing variables (ISO 8601 dates ,severity mapping, TEAE flags).
- Achieved 60% to 70% reduction in processing time ,which substantially reduced dataset creation hours per domain using the macro loop automation across multiple oncology trail datasets. Reduction mapping errors by 35% to 40% which might differ based on the expert level of the programmer ,ensuring higher accuracy and CDISC compliance (SDTMIG v3.4,ADaMIG v1.2) for regulatory submission ,as validated by sample outputs (TU,ADRS,AE ,ADAE tables).

Conclusion

This poster demonstrated that integrating PROC IML,PROC FCMP and a macro loop significantly enhances the efficacy and accuracy of creating SDTM and ADaM dataset in oncology clinical trials. In general, the approach is very innovative by automating the process ,reducing errors and reducing processing time as SAS tools which will help further a standard way compliance of the CDISC and other regulatory agencies for submission. Future works will explore extending this framework to additional domains like LB and ADLB.

References

- Huang, L., Dong, H., & Liu, Y. (2023). Fitting Logitoid-Normal distributions with MLE estimate by SAS SEVERITY and PCMP procedures [Paper presentation]. PharmaSUG 2023 Conference, San Francisco, CA, United States.

- SAS Documentations on PROC FCMP : https://documentation.sas.com/doc/en/pgmsascdc/v_060/proc/n0pio2crltpr35n1ny010zrfbvc9.htm and PROC IML: https://support.sas.com/en/software/sasiml-viya-support.html.

- CDISC Documentation: https://www.cdisc.org/standards/foundational.

Acknowledgments

I greatly acknowledge the support and guidance from my supervisor Ajay Gupta, Director, Innovation DSS - DEN in general.

Contact Information

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

Misikir Tesfaye, MSc, MPH, is Manager in Data Engineering within statistical programming and Innovation at Daiichi Sankyo Inc currently. He has more than 20 years of experience in clinical, health care and pharmaceutical analytics. His expertise includes HEOR,HTA and real-world evidence analytics. He is an advanced user of SAS,R and python for statistical programming , data science, and machine learning applications. He can be contacted at misikir.tesfaye@daiichisankyo.com,

Ajay Gupta (M.S.) is Director of Statistical Programming at Daiichi Sankyo U.S., with over 20 years of experience in data standardization, technical leadership, and systems development across the CRO and pharmaceutical sectors. He holds a master's degree in biomedical engineering from Louisiana Tech University and has been a regular SAS and PharmaSUG conference presenter since 2010, covering topics such as CDISC, RWD/RWE, and advanced analytics tools. He can be contacted at ajay.gupta@daiichisankyo.com.

DISCLAIMER