Paper 213-2011

A SAS Macro to Validate Business Rules for Phasing Derivations in the Clinical Data Repository

Ellen Xiao, Merck Sharp & Dohme Corp., Rahway, NJ

ABSTRACT

This paper describes the business rules for phasing derivations and the related concepts in Merck's Clinical Data Repository (CDR). It introduces the %CHKPHASERULE macro, a tool to validate the phasing derivations for all related domains in the CDR. This macro was created utilizing SAS/MACRO with version 9.1.3. The business rules for phasing derivations includes two parts (ruleXXA and ruleXXB). By using the Metadata Setup specifications, the programmers can easily identify which domains relate to which part of this business rule to ensure the macro is being called correctly. This tool can test to see if the phasing derivations are following the tie-breaker rules defined in Metadata setup for different domains. This reduces many tedious jobs to check epochs derived from domain to domain manually in the CDR. It offers benefits and process improvement to the business rules validation programmers.

INTRODUCTION

A Clinical Data Repository (CDR) is a real-time database that consolidates data from a variety of clinical sources to present a unified view of a single patient. Merck's CDR features FDA's Janus database with extensions, CDISC SDTM based repository, SDTM, and SDTM+ views. It is a system that supports collected and derived data, standards-based data exchange, frozen files, vocabulary updates, study pooling and phasing. Phasing derivations for clinical trial data is one of the important parts for database derivation which need to follow certain business rules. This paper will provide some basic concepts related to phasing derivation, will describe the related business rules and will introduce a tool to validate the phasing derivations in the CDR.

SOME BASIC CONCEPTS

An Epoch -- describes a segment or period of time within a trial (e.g. Screening, Placebo run-in, Treatment, Follow-up). When a study is blinded, and treatment assignments are unknown, collected data are phased by epoch start and end rules.

An Element -- also describes a segment or period of time within a trial based on start and end rules but includes unblinded treatment and non-treatment details (e.g. MK-XXXX 5 mg, Follow-up).

An Element Code (ETCD) -- is a short name of an element (e.g. SCR, PBO, MKXX, PS) which can be used for programming and sorting.

An Epoch Sequence (TAETORD) -- describes the order (e.g. -10, 10, 20, 30, 40) that elements occur within an Arm / Treatment Group.

Study Day Relative to Epoch -- identifies how many days into an Epoch an event or intervention started (such as Adverse Event or Concomitant Therapies). This derivation should be calculated on all domains that contain a date field which describes when an event or intervention started. It is an integer number of days, and should not be 0.

THE BUSINESS RULE TO DERIVE EPOCH

One of the key tasks for the CDR to apply business rules to correctly derive the phase is to complete the Metadata setup appropriately. Metadata setup is study-specific. Below is an example of Metadata setup for epoch and start event function.

A	B	D	G	H	I	J	K
EPOCH	EPOCH Sequence	Treatment Flag	Off Set	Off set Unit	EPOCH Start Event (General Description)	EPOCH Start Event	Start Event Function Name
Screening	10	N			Birth Date	SUPPDM.BRTHDTI	PH_DM_BIRTH_DATE
Metformin Run In	20	N			Visit 2 Date	SUPPSV.SVSTDTI where	PH_SV_DT_V2
Placebo Run In	30	N			Date of first non-zero dose from label A or B	MIN SUPPEX.EXSTDTI where EX.EXSPID IN (A,B) AND	PH_EX_ST_DT_MIN_NZ_A_B
Treatment I	40	Y			Date of first non-zero dose bottles C,D,E,F	MIN SUPPEX.EXSTDTI where EX.EXSPID IN (C,D,E,F) AND	PH_EX_ST_DT_MIN_NZ_CDEF
Post-Treatment I	50	N	5	day	For subjects with a status of discontinued or completed use the maximum Date of the last non-zero dose (Label	MAX SUPPEX.EXENDTI where (EX.EXSPID IN (C,D,E,F) AND SUPPEX.EXNUMDOS <> 0) AND DS.DSSCAT = ('DISCONTINUED') or	PH_EX_EN_DT_MAX_NZ_C2F_DSC_CMP

The derivation phases data based on defined epoch and element start and end event rules. When a study is blinded, epoch values will be derived and element, element code and epoch sequence will be blinded. When a study is unblinded, all four variables will be populated with study-specific values based on start and end event rules.

This business rule exists in two parts. One part phases data using start dates (-- STDTI) and the other part phases data by collection date (-- DTI). When metadata are defined for the CDR, a study's list of domains will be separated into two groups; those to be phased by Start Dates (-- STDTI) will have XXA applied and those to be phased by Collection Dates (--DTI) will have XXB applied.

Example of Metadata setup related to Business Rule applying:

	B	E	F	G
1	SDTM Domain	Derivation	Comments	Enable
2	AE	BR001_AGE_AT_AE_ONSET		Y
3	AE	BR005_START_DAY_RELT_SUB_RF_DT		Y
4	AE	BR006_STOP_DAY_RELTO_SUB_RF_DT		Y
5	AE	BR013_TIME_SINCE_LAST_DOSE	Should not be enabled for vaccine studies	Y
6	AE	BR022_AE_DURATION		Y
7	AE	BR037_MEDDRA_PT_AND_PSOC	Should only be used for CRO studies	N
8	AE	BR050A_EPOCH		Y
9	AE	BR052_INTENSITY_CODE		Y
10	AE	BR053_ACTION_CODE		Y
11	AE	BR054_CAUSALITY_CODE		Y
12	AE	BR072_CONTINU_INDICTR		N
13	AE	BR076_STOP_DAY_REL_TO_EPOCH		Y
14	CM	BR005_START_DAY_RELT_SUB_RF_DT		Y
15	CM	BR006_STOP_DAY_RELTO_SUB_RF_DT		Y
16	CM	BR010_DRUG_NAME	Should only be used for CRO studies	N
17	CM	BR021_DUR_BETWEEN_TWO_DT		Y
18	CM	BR050A_EPOCH		Y
19	CM	BR072_CONTINU_INDICTR		N
20	CM	BR076_STOP_DAY_REL_TO_EPOCH		Y
21	DM	BR002_AGE_AT_SUBREF_STRT_DT_TM		Y
22	DM	BR003_SUBJECT_REF_START_DATE		Y
23	DM	BR004_DAY_RELTO_SUBJECT_REF_DT		Y
24	DM	BR009_DAYS_ON_TREATMENT	Should not be enabled for vaccine studies	Y
25	DM	BR014_PLANNED_ARM_CODE		Y
26	DM	BR034_INVESTIGATE_SITE_COUNTRY		Y
27	DM	BR046_DATE_OF_DISCONTINUANCE	Should not be enabled for vaccine studies	Y
28	DM	BR047_RELA_DAY_OF_DISCONTINUCE	Should not be enabled for vaccine studies	Y
29	DM	BR063_ACTUAL_ARM_CODE		Y
30	DM	BR082_DM_SITENUM		Y
31	DM	BR084_SUB_REF_END		Y
32	DS	BR004_DAY_RELTO_SUBJECT_REF_DT		Y
33	DS	BR005_START_DAY_RELT_SUB_RF_DT		Y
34	DS	BR050A_EPOCH		Y
35	EG	BR004_DAY_RELTO_SUBJECT_REF_DT	Should only be used if ECG data is collected	Y
36	EG	BR050B_EPOCH	Should only be used if ECG data is collected	Y
37	EG	BR091_BASELINE_FLAG	Should only be used if ECG data is collected	Y
38	EG	BR065_ECG_CATEGORY		N

Epochs will start based upon a predetermined event and will end when the next epoch starts. In the case where data can be assigned to more than one epoch (also referred to as a tie-breaker situation) specific predetermined rules will be used to determine the correct epoch.

Example of Metadata setup related to tie-breaker rules.

A	B	C	D	E	F
SDTM Domain Name	Early EPOCH Name	Late EPOCH Name	Assigned To EPOCH Name	Tie-Breaker Qualifier	Tie-Breaker Value
AE	Pre-Study	Treatment	Treatment	None	None
CM	Pre-Study	Treatment	Treatment	STRF	DURING
EX	Pre-Study	Treatment	Treatment	None	None
DS	Pre-Study	Treatment	Treatment	None	None

Tie-breaker rules can be defined for certain variables. The business rule also lists them in priority order. If multiple variables are coordinated to break ties for specific data, and if conflicting values exist, the data will be phased based upon the defined priority order list.

A MACRO TO VALIDATE THIS BUSINESS RULE

A macro %CHKPHASERULE can be used to validate this Business Rule for phasing derivation. Major steps of this macro are listed below.

(1) Import the CDR Metadata Setup Specifications by reading in the ***Epoch_and_Start_Events*** sheet.

```
PROC IMPORT OUT=WORK.MSS
            DATAFILE= "&uatdirxls\&mssfile"
            DBMS=EXCEL REPLACE;
     SHEET="Epoch_and_Start_Events$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
Run;
```

(2) Count the total number of distinct epochs listed in the Metadata Setup Specification. Create a macro variable called ***totepoch*** to indicate the total number of epochs defined in the study. Use this macro variable as a loop indicator in the next step.

```
data _null_;
      set mss end=eof;
      where epoch ne ' ';
      call symput('epoch'||left(_n_),left(epoch));
       call symput('seq'||left(_n_), left(epoch_sequence));
      if eof then call symput('totepoch',left(_n_));
run;
```

(3)Subject Element(SE) domain includes one record per Actual Element per Subject. SE contains a set of treatment intervals for each subject covering both active medication as well as off-treatment periods (like screening, washout and post-treatment).

The macro reads in the data from the SE domain and transposes it to make one Subject, one record to include all the Actual Elements for each Subject including.Pre-Study phase start date time, end date time, Treatment phase start date time, end date time, Post-study phase start date time and end date time, etc. .

Subject	Trial Epoch	sestdat	sesttime1	sestdti1	seendate1	seentime1	seendti1	Trial Epoch	sestdate2	sesttime2	sestdti2	seendate2	seentime
)00200005	Pre-Study	871	.	75254400	17776	47100	1535893500	Treatment	17776	47100	1535893500	17782	72300
)00200006	Pre-Study	-3501	.	-302486400	17730	48660	1531920660	Treatment	17730	48660	1531920660	.	.
)00200007	Pre-Study	700	.	60480000	17666	70500	1526412900	Treatment	17666	70500	1526412900	17667	70500
)00200008	Pre-Study	349	.	30153600	.	.	.		.	.	.	.	.
)00200018	Pre-Study	3349	.	289353600	17785	54300	1536678300	Treatment	17785	54300	1536678300	.	.

```
%do j=1 %to &totepoch;
    data se&j;
       set uatsddub.se;
       if epoch="&&epoch&j";
           if sestdti ne '' then do;
                 sestdate&j=input(sestdti,is8601da.);
                 sesttime&j=input(scan(sestdti,2,'T'), time8.);
               sestdti&j=sum(input(sestdti,is8601da.)*24*60*60,
                           input(scan(sestdti,2,'T'), time8.));
           end;
           if seendti ne '' then do;
                seendate&j=input(seendti,is8601da.);
               seentime&j=input(scan(seendti,2,'T'), time8.);
               seendti&j=sum(input(seendti,is8601da.)*24*60*60,
```

```
                        input(scan(seendti,2,'T'), time8.));
                end;
            rename epoch=epoch&j;
        keep usubjid epoch sestdti&j sestdate&j sesttime&j seendti&j
              seendate&j seentime&j;
        run;
    %end;

 data new_se;
    merge
    %do j=1 %to &totepoch;
      se&j
    %end;;
     by usubjid ;
 run;
```

(4) Test to see if there are any Subjects existing in other domains, but without any records in the SE domain. If so, there's a mismatched epoch in the tested domain.

```
data nose_frm_domain;
   merge domain1(in=inae) se1(in=inse);
   by usubjid;
   if inae and not inse;
   myepoch=' ';
   if epoch ne myepoch then flag1="mis-match EPOCH in &domain.domain";
run;
```

(5) Test to see if there's a tie between epochs which need to follow the metadata setup specification.

For example, per the Metadata Setup Specification provided above, if there's a tie between Pre-Study and Treatment phases, check variable ***STRF.*** If this variable has a value of ***"DURING"***, the epoch will be assigned as treatment, otherwise it will be Pre-Study. If either element start date time or event start date time doesn't include a time portion, only the date parts from both will be compared to identify if there's a tie between epochs. If both of them include date time, then compare the date and time from both to identify if a tie exists.

```
data check_dom;
   merge domain1 new_se1;
   by usubjid;
   length flag $20;
   %if %upcase(&domain)=CM %then %do;
       if upcase(CMSTRF)='DURING' then do;
          if startdt='' then flag='';
           else do;
               if starttime=. or seentime1=. then do;
                   if startdate=seendate1 then flag='tie at epoch1';
             end;
              else do;
                 if startdt=seendti1 then flag='tie at epoch1';
                  else flag='no tie';
              end;
          end;
       end;
        else do;
          flag='no tie';
      end;
       keep usubjid subjid epoch1-epoch4 sestdti1-sestdti4 seendti1-
            seendti4 startdt flag epoch &domain.stdti CMSTRF;
    %end;
    %else %if %upcase(&domain)=AE
    |%upcase(&domain)=DS|%upcase(&domain)=EX %then %do;
        if startdt='' then flag='';
        else do;
             if starttime=. or seentime1=. then do;
                if startdate=seendate1 then flag='tie at epoch1';
            end;
             else do;
               if startdt = seendti1 then flag='tie at epoch1';
                else flag='no tie';
            end;
```

```
        end;
    %end;
    %else %do;
       flag='no tie';
        keep usubjid subjid epoch1-epoch4 sestdti1-sestdti4 seendti1-
            seendti4 startdt flag epoch &domain.stdti;
    %end;
```

(6) Derive *myepoch* :

```
data myepoch;
  length myepoch $200.;
  set check_dom;
     if flag='tie at epoch1' then do;
        if epoch2 ne ' ' then myepoch=epoch2;
     end;
     else do;
        if nmiss(seendti1, seendti2, seendti3) =0 then do;
           if sestdti1< &domain.start <=seendti1 then myepoch=epoch1;
            else if sestdti2< &domain.start <=seendti2 then
                  myepoch=epoch2;
           else if sestdti3< &domain.start <=seendti3 then
                  myepoch=epoch3;
                else if &domain.start > sestdti4 then myepoch=epoch4;
               end;
              …
               if &domain.epoch ne myepoch then flag1="mis-match EPOCH in
       &domain. domain";
 run;
```

(7) The following code is comparing the CDR derivations and self derivations. The difference will be listed in diff_domain.

```
data diff_dom1;
  set myepoch;
  where flag1="mis-match EPOCH in &domain.domain";
  keep usubjid subjid myepoch &domain.epoch sestdti seendti
       &domain.stdti;
run;

data diff_dom2(rename=epoch=&domain.epoch);
  set nose_frm_domain;
  where flag1="mis-match EPOCH in &domain. domain";
  keep usubjid subjid myepoch epoch &domain.stdti;
run;

data diff_domain(drop=p);
    set diff_domain1 diff_domain2;
    by USUBJID;
    if USUBJID^=' ' then p+1;
    call symput("num",put(p,12.));
    label myepoch="EPOCH that is expected"
      &domain.epoch="EPOCH in &domain. domain";
run;
```

CONCLUSION

Incorporating %CHKPHASERULE into checking the business rules for the phasing derivations in the CDR has proven to be a positive step in ensuring phasing derivatives follow the Metadata Setup Specifications and for consistency with the information provided by the Subject Element domain. It alleviates the amount of effort required to validate epochs derived from domain to domain.

REFERENCES

Clinical Data Repository Standards Business Rules/Database Derivations Business Version: 5.12
Clinical Data Repository User Interface

ACKNOWLEDGMENTS

Thanks for John Vlattas to review and modify this paper.

CONTACT INFORMATION

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

Ellen Xiao
Merck & Co., Inc.,
Rahway, NJ 07065
E-mail: hong_xiao2@merck.com