

A Special SDTM Domain RELREC and its Application

Changhong Shi, Beilei Xu
Merck Sharpe & Dohme Corp.
Rahway, NJ

ABSTRACT

The Related Records (RELREC) domain is a Special-Purpose Relationship domain in the Study Data Tabulation Model (SDTM). It is used to identify relationships between records in two (or more) domains. The relationship is sometimes important and unique for analysis. This paper discusses the basic attributes and meaning of the RELREC domain, as well as how to use different types of relationships in RELREC to facilitate analysis.

Keywords: SDTM, Related Records, RELREC

INTRODUCTION

In SDTM, RELREC is a Special Purpose Relationship domain. It contains the following variables:

STUDYID, RDOMAIN, USUBJID, IDVAR, IDVARVAL, RELTYPE, and RELID

Among these variables, the key variables for the RELREC domain are:

STUDYID, RDOMAIN, USUBJID, IDVAR, IDVARVAL, and RELID

The meanings of the above variables are displayed in the table below:

Variable Name	Variable Description
STUDYID	Study Identifier
RDOMAIN	The two-letter domain code of the record in the relationship, such as AE, CM.
USUBJID	Unique Subject Identifier
IDVAR	Unique Record Identifier such as AESEQ
IDVARVAL	The value of the corresponding IDVAR
RELTYPE	Relationship Type. Values should be either ONE or MANY. However, values are only necessary when identifying a relationship between datasets
RELID	Relationship Identifier

The RELREC domain can be used to describe relationships between records for a subject or relationships between different domains. The relationships represented in RELREC are collected relationships, either by explicit references or check boxes on the Case Report Form (CRF). For example, concomitant medications taken as a result of an adverse event or inexplicitly by the design of the CRF - information collected on the Hypoglycemia (CAH) form is mapped to different domains: CF, LB etc.

In RELREC, a relationship is constructed by adding a record to RELREC for each record to be related and by assigning a unique character identifier value for that relationship. Each record in the RELREC domain contains keys that identify a record (or group of records) and the relationship identifier stored in the RELID variable. The value of RELID must be identical for all related records within each USUBJID. Records expressing a relationship are specified using the key variables mentioned above. Single records can be related by using a unique-record-identifier variable such as AESEQ in IDVAR. Groups of records can be related by using grouping variables such as LBGRPID in IDVAR. IDVARVAL would then contain the value of the variable described in IDVAR.

This paper will use some examples to illustrate how to use RELREC to facilitate analyses.

DISCUSSIONS

There are two types of relationships in RELREC.

One type is a collected relationship, either by explicit references or check boxes on the CRF. For example, concomitant medications are taken as a result of an adverse event.

The following datasets show how the collected relationship is captured in the RELREC domain:

Patient 12345 has a CM record with CMSEQ as '22222' resulting from a serious AE with AESEQ as '1111'.

In the CM domain, the CM record is shown as:

USUBJID	CMSEQ	CMSPID	CMTRT
12345	22222	2	Aspirin

In the AE domain the AE record is shown as:

USUBJID	AESEQ	AESPID	AETERM
12345	1111	1	Severe Headache

Their relationship is shown in the RELREC domain as below:

USUBJID	RDOMAIN	IDVAR	IDVARVAL	RELID
12345	AE	AESEQ	1111	1
12345	CM	CMSEQ	22222	1

The relationship is shown in the CM CRF with a section: "Complete entire section below for statin medication or if taken within 14 days of onset of SAE." If a site entered information into that section, it is then mapped to RELREC with the same RELID to link the CM and AE records in their respective domains.

The following is the SAS® code to retrieve the relationship from RELREC and link the two records.

```
*-----*;  
*- Step 01: Get the links from RELREC -*;  
*-----*;  
  
data ae cm;  
  set lptss.relrec;  
  if idvar='CMSEQ' then output cm;  
  else if idvar='AESEQ' then output ae;  
run;  
  
proc sort data=cm;  
  by usubjid relid;  
run;  
  
proc sort data=ae;  
  by usubjid relid;  
run;  
  
data cmae(drop=rdomain) ;  
  merge cm(in=cm rename=(idvarval=cmseq))  
        ae(in=ae rename=(idvarval=aeseq))  
  by usubjid relid;  
  if ae;  
run;
```

Using the above code based on RELREC, the concomitant medications due to a specific AE are identified.

The second type of relationship in RELREC is implicit by the design of the CRF and the SDTM mapping.

For example, the Hypoglycemia (CAH) form collects both clinical findings and the lab results of the glucose which are mapped to different domains: CF and LB, respectively.

Although different information is mapped to different domains, the relationship is established in RELREC which shows they are related – both coming from the same CRF at one assessment.

To perform the analysis, knowledge of which CAH record has what glucose value is required along with information to which that specific glucose value is associated with. For example, if that episode needs medical attention, use RELREC to tie the two pieces of information together. This will be illustrated by the example below.

The following is an example of raw collected CAH data for one patient's single episode before mapping is applied.

Table 1: Raw CAH data pre-mapping application

USUBJID	VISITNM	DT_ASSESSMENT	DT_EPISODE	TM_EPISODE	REQUIRE_MED_ATTENTION	GLUCOSE	GLUCOSE_UNIT
000010	CAH/FSG	9/14/2010	2/4/2010	11:05	N	160	mg/dL

Note that in the CAH form, other information is collected such as the reason for the episode, the time the patient was treated, etc. For the purposes of this paper, only the finger stick glucose value and its unit (i.e. GLUCOSE and GLUCOSE_UNIT in the above table), as well as the information on whether the patient requires medical attention for that specific episode (i.e. REQUIRE_MED_ATTENTION field) were kept.

From the raw data above, it is easily determined that this episode did not require medical attention, and that the finger stick glucose value was 160 mg/dL. However, per the mapping specification, the above one row record was mapped to two different domains: CF and LB, as displayed below.

Table 2: CF (Only with the records for requiring medical assistance and glucose information.)

USUBJID	DOMAIN	CFGRPID	CFSEQ	VISIT	CFDTC	CFTEST	CFORRES
000010	CF	CAH	22	V8 Week 34 UNSCHEDULED	2010-09-14	Receive Medical Attention for Symptoms	N

and

Table 3: LB

USUBJID	DOMAIN	LBGRPID	LBSEQ	LB DTC	LBTEST	LBORRES	LBORRESU
000010	LB	CAH	29	2010-09-14T11:05	Glucose	160	mg/dL

From the above CF and LB data, even though the records represent the same patient, i.e. USUBJID='000010', it cannot be determined whether the records are related or not. In order to establish if a relationship exists post mapping, reliance on the information in the RELREC domain below is required.

The records in RELREC for this patient are shown in Table 4. The record that has IDVAR='CFSEQ' means that record is based on the CF domain, similarly, the record that has IDVAR='LBSEQ' is based on the LB domain:

Table 4: RELREC records from CF and LB domains only

USUBJID	RDOMAIN	IDVAR	IDVARVAL	RELTYPE	RELID
000010	LB	LBSEQ	29		35

000010	CF	CFSEQ	22		35
--------	----	-------	----	--	----

From the RELREC records above, it is proven that the two records are related because they have the same value for the RELID variable. Once the relationship is established, the associated information from the individual domain is obtained along with the ability to produce the output tables/listings for analysis.

Due to limitation of space and for display purposes, the one row record with the associated information was divided into two tables.

Table 5: The associated information from the LB domain

USUBJID	LBSEQ	LBDC	LBTEST	LBORRES	LBORRESU	CFSEQ
000010	29	2010-09-14T11:05	Glucose	160	mg/dL	22

Table 6: The associated information from the CF domain

USUBJID	LBSEQ	CFSEQ	VISIT	CFDTC	CFTEST	CFORRES
000010	29	22	V8 Week 34 UNSCHEDULED	2010-09-14	Receive Medical Attention for Symptoms	N

The LBSEQ and CFSEQ fields were duplicated intentionally to show that the two tables represent one record, with the information from both LB and CF domains that are related to each other.

From this associated record, it is established that this episode did not require medical attention, and the finger stick glucose value was 160 mg/dL, which is exactly the same conclusion drawn based on raw data pre-mapping (Table 1).

For reference purposes, the programming code to obtain the final dataset with the associated information is shown in the appendix. The final dataset of cf_lb_assoc_info2 in the appendix contains information from both the CF and LB domain, as displayed in Table 5 and 6 above.

Note that the above are pseudo-code for display logic purposes and only assume one episode case. Therefore, code adjustments are needed for other cases.

CONCLUSION

In some situations, information collected on the same CRF needs to be mapped to different domains to comply with the SDTM IG, and RELREC has to be used to trace the relationships between these records (as in the CAH example above); in some other situations, the relationship is collected on the CRF, but the relation can only be mapped to RELREC (the AE and CM example above). In both cases, using RELREC is unavoidable. The example code above and the code in the appendix show the basic technique on how to tie the relation back into the analysis.

REFERENCES

CDISC Study Data Tabulation Model Implementation Guide: Human Clinical Trials
Version 3.1.1

ACKNOWLEDGEMENTS

The author would like to thank the management team for their review of this paper and help through the company publication process.

CONTACT INFORMATION

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

Changhong Shi
Merck Co. & Inc.
RY34-A320
P.O. Box 2000
Rahway, NJ 07065
(732)-594-1383
changhong_shi@merck.com

Beilei Xu
Merck Co. & Inc.
RY34-A320
P.O. Box 2000
Rahway, NJ 07065
(732)-594-9980
beilei_xu@merck.com

APPENDIX

```
/* Step 1: Obtain the CF data */

libname tt "C:\PharmSUG2011\examples";

/* Find the CF domain data. Since in the Inform CAH data,
   there are many fields mapped to CF, we only keep finger glucose
   data and whether the patient needs medical attention info from CF
   for simplicity */

*** CF data;

data cf;

    set tt.cf;
    WHERE cfstestcd='MEDATTN';

    if usubjid='000010';

    keep usubjid domain cfgrpdc cfseq visit cfdtc cfstest cforres;
run;

/* Step 2: Obtain the LB data from CAH for the same patient */

data lb;
    set tt.lb;

    where usubjid='000010' and lbgrpdc='CAH';

    keep usubjid domain lbseq lbdtc lborres lborresu lbtest lbgrpdc;
run;

/* Step 3: Obtain the relrec records for this pt, based on CF or LB domain relation only */

data relrec_cf_lb_all;
    set tt.relrec;

    where usubjid='000010' and rdomain in ('CF', 'LB');

run;

/* Step 4: Now link the LB and CF data together using RELREC info */

*** The Relrec CF record that corresponds to the CF record in step 1;

proc sort data=relrec_cf_lb_all(where=(rdomain='CF' and idvar='CFSEQ'))
    out=cf_relrec;
    by idvarval;
run;

proc sort data=cf_relrec;
    by usubjid idvarval;
run;

proc sort data=cf;
    by usubjid cfseq;
run;
```

```

data cf_relrec2;
  merge cf(in=cc) cf_relrec(in=rr rename=(idvarval=cfseq));
  by usubjid cfseq;
  if cc and rr;
run;

*** Relrec record that corresponds to the record in the LB domain data
    in step 2;

proc sort data=relrec_cf_lb_all(where=(rdomain='LB' and idvar='LBSEQ'))
  out=lb_relrec;
  by idvarval;
run;

proc sort data=lb_relrec;
  by usubjid idvarval;
run;

proc sort data=lb;
  by usubjid lbseq;
run;

data lb_relrec2;
  merge lb(in=cc) lb_relrec(in=rr rename=(idvarval=lbseq));
  by usubjid lbseq;
  if cc and rr;
run;

*** Now find the associated LB and CF records by checking if they share the same relid;

proc sort data=lb_relrec2;
  by usubjid relid;
run;

proc sort data=cf_relrec2;
  by usubjid relid;
run;

data cf_lb_related;
  merge lb_relrec2(in=l1) cf_relrec2(in=cc rename=(idvar=idvar_cf
                                                    rdomain=rdomain_cf));
  by usubjid relid;

  if l1 and cc;

run;

*** Now put the associated CF and LB info together for output tables;

proc sort data=cf_lb_related;
  by usubjid idvarval_cf;
run;

data cf_assoc_info;
  merge cf(in=cc) cf_lb_related(in=rr keep=usubjid idvarval_cf
                                rename=(idvarval_cf=cfseq));
  by usubjid cfseq;

  if cc and rr;
run;

```

```

*** associated lab info;

proc sort data=cf_lb_related;
  by usubjid idvarval;
run;

data lb_assoc_info;
  merge lb(in=cc) cf_lb_related(in=rr keep=usubjid idvarval
                               rename=(idvarval=lbseq)
                               );
  by usubjid lbseq;

  if cc and rr;
run;

*** merge the LB and CF data together;

data cf_lb_assoc_info;
  merge lb_assoc_info(in=ll) cf_assoc_info(in=cc);
  by usubjid;

run;

proc sql;
  create table cf_lb_assoc_info2 as
  select usubjid, lbseq format best22., lbdtc, lbtest, lborres,
         lborresu, cfseq format best22., visit, cfdtc,
         cfctest, cforres
  from cf_lb_assoc_info;
quit;

```