

How to validate SDTM SUPPQUAL Data Set

Kevin Lee, Covance (Periapproval), Conshohocken , PA

ABSTRACT

The paper is intended for Clinical Trial SAS programmers who are interested in validating SDTM SUPPQUAL data set. The paper will show the basic structures of SDTM SUPPQUAL, and it will discuss the issues and the resolution of the validation of SDTM SUPPQUAL.

INSTRUCTION OF SUPPQUAL

The SDTM does not allow any new variables beside ones assigned to each SDTM domain. So, the supplemental data set is introduced to supplement each SDTM domain. The supplemental data set is either a single SUPPQUAL dataset or separate supplementary data sets (SUPP) such as SUPPDM, SUPPAE, and SUPPEX.

STRUCTURE OF SUPPQUAL

The SUPPQUAL data set consists of the following variables.

Domain	Name	Label	Type	Role	Codelist/Format	Core	Origin
SUPPQUAL	STUDYID	Study Identifier	Text	Identifier	NONE	Req	CRF
SUPPQUAL	RDOMAIN	Related Domain Abbreviation	Text	Identifier	DOMAIN	Req	Derived
SUPPQUAL	USUBJID	Unique Subject Identifier	Text	Identifier	NONE	Req	Sponsor Defined
SUPPQUAL	IDVAR	Identifying Variable	Text	Identifier	NONE	Req	Sponsor Defined
SUPPQUAL	IDVARVAL	Identifying Variable Value	Text	Topic	NONE	Req	CRF / Sponsor Defined
SUPPQUAL	QNAM	Qualifier Variable Name	Text	Record Qualifier	NONE	Req	Derived
SUPPQUAL	QLABEL	Qualifier Variable Label	Text	Record Qualifier	NONE	Req	CRF / Sponsor Defined
SUPPQUAL	QVAL	Data Value	Text	Qualifier	NONE	Req	CRF / Sponsor Defined
SUPPQUAL	QORIG	Origin	Text	Qualifier	NONE	Req	Derived
SUPPQUAL	QEVAL	Evaluator	Text	Qualifier	NONE	Perm	Derived

RELATION OF SUPPQUAL TO EACH DOMAIN

The SUPPQUAL data set is linked back to a parent record using RDOMAIN, USUBJID, IDVAR and IDVARVAL. IDVAR and IDVARVAL is usually the sequence number (ex. AESEQ, VSSEQ and so on) except DM and SV.

ISSUES OF VALIDATING SUPPQUAL

There are a couple of issues in validating the SUPPQUAL data set.

1. QVAL is the character
2. SUPPQUAL data set is too large, so it takes too long to read and to validate the SUPPQUAL data set
3. It is really difficult to link the SUPPQUAL data back to a parent record for the review

THE RECOMMENDATION

The followings are the recommendations from our experiences.

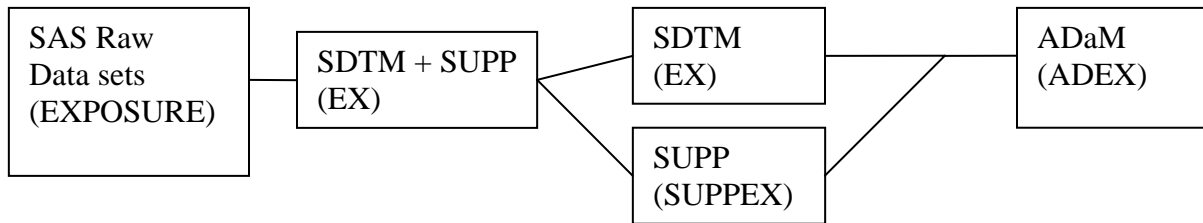
1. Create the separate SUPP for each domain for easier validation, not just a single SUPPQUAL.
2. Create three data sets for each domain. First, create SDTM+SUPP data set and from it, create SDTM data set and the corresponding supplementary (SUPP) data set.

The advantage to have individual SUPP data set is that supplementary data sets won't be as big as SUPPQUAL, so it takes less time to review, to update, and to validate.

The advantages to have SDTM+SUPP data set are

1. No need to open a parent SDTM data set and its supplementary data set to review since SDTM+SUPP keep all the related variables for the domain, including supplementary variables.
2. Much easier to link back to a parent record for the supplementary variables.
3. Much easier to validate, especially the numeric supplementary variables.
4. Validating only one data set, not 2 data sets, assuming that splitting into SDTM and SUPP are correct.
5. Much easier when creating ADaM Data set.

THE FLOWCHART OF THE PROCEDURE



According to flowchart, we have 3 data sets to validate, EX in SDTM+SUPP, EX and SUPPEX in SDTM. The followings are several ways to validate the supplementary variables.

1. Validate all 3 of them - EX in SDTM+SUPP, EX and SUPPEX.
2. Validate only 2 SDTM data sets – EX and SUPPEX
3. Validate only EX in SDTM+SUPP

Our recommendation is to validate all 3 of them. However, the best way to validate is to start validating EX in SDTM+SUPP in the initial stage and validating all 3 of them in the later stage of the trials.

The advantages to validate EX in SDTM+SUPP are

1. much easier to validate the numeric supplementary variables.
2. if there are issues in the supplementary variables, it is much easier to link back to parent records for the review.

In order to validate only one data set in SDTM+SUPP, the SAS programmer needs to create the automatic procedure that splits SDTM+SUPP into SDTM and SUPP – the suppqal flag document and the macro that reads the document and splits SDTM+SUPP into SDTM and SUPP.

THE AUTOMATIC PROCEDURE TO CREATE SDTM and SUPP

In order to create SDTM and SUPP from SDTM+SUPP, the SAS programmer needs to create macro and the flag document (the best option is to create the flag variable column in Define document) to indicate which variables are SDTM domain variables and supplementary variables.

THE DEFINE DOCUMENT to separate SDTM, SUPP, SDTM+SUPP

Domain	Name	Label	Type	Role	Codelist/Format	Core	Origin	Suppqual Flag
EX	STUDYID	Study Identifier	Char	Identifier	NONE	Req	CRF	
EX	DOMAIN	Domain Abbreviation	Char	Identifier	DOMAIN	Req	Derived	
EX	USUBJID	Unique Subject Identifier	Char	Identifier	NONE	Req	Sponsor Defined	
EX	EXSEQ	Sequence Number	Num	Identifier	NONE	Req	Derived	
EX	EXTRT	Name of Actual Treatment	Char	Topic	NONE	Req	Derived	
EX	EXDOSE	Dose per Administration	Num	Record Qualifier	NONE	Perm	CRF	
EX	EXDOSU	Dose Units	Char	Variable Qualifier	DOSU	Exp	CRF	
EX	EXDOSFRQ	Dosing Frequency Per Interval	Char	Variable Qualifier	DOSFRQ	Perm	Derived	
EX	EXROUTE	Route of Administration	Char	Variable Qualifier	ROUTE	Perm	CRF	
EX	EXSTDTC	Start Date/Time of Treatment	Char	Timing	ISO8601	Perm	CRF	
EX	VISIT	Visit Name	Char	Timing	NONE	Perm	Derived	
EX	VISITNUM	Visit Number	Num	Timing	NONE	Perm	Derived	
EX	TAETORD	Order of Element within Arm	Num	Timing	NONE	Perm	Derived	
EX	EPOCH	Trial Epoch	Char	Timing	EPOCH	Perm	Derived	
EX	EXSPID2	Vial or Component ID	Char	Plus	NONE	Perm	Derived	Y
EX	EXSTDTI	Start Date/Time of Treatment Imputed	Char	Plus	ISO8601	Perm	Derived	Y
EX	ELEMENT	Description of Element	Char	Plus	NONE	Perm	Derived	Y
EX	ETCD	Element Code	Char	Plus	NONE	Perm	Derived	Y
EX	STDYRLEP	Start Day Rel to Epoch	Num	Plus	NONE	Perm	Derived	Y

According to EX domain in Define Doc, EX in SDTM+SUPP will contain all the variables in Define Doc. The suppqual flagged variables (EXSPID2, EXSTDTI, ELEMENT, ETCD, STDYRLEP) will go to SUPPEX and the rest of variables will be kept in EX.

THE SAMPLE CODES to create SDTM and SUPP from SDTM+SUPP

```
**** Create SDTM EX domain;
data sdtm.ex;
    set sdtmsupp.ex;
    drop exspid2 exstdti exvacnum element etcd stdyrlep;
run;
**** Create SDTM SUPPEX domain;
proc transpose data=sdtmsupp.ex out=_suppex(rename=( _name_=qnam _label_=qlabel
coll=qval exseq=_exseq));
    by studyid domain usubjid exseq;
    var exspid2 exstdti exvacnum element etcd stdyrlep;
run;
data sdtm.suppex(drop=_exseq);
    set _suppex;
    exseq = trim(left(put(_exseq, best30.)));
    qorig = 'Derived'; **** all are derived, but it also can come from Define Doc;
run;
```

Above sample codes should be implemented into macro and the supplementary variables should be flagged from Define document.

THE OVERALL STEPS TO VALIDATE

The following steps are the recommendation for the validation of SUPPQUAL data sets.

1. Create the define document including the flags to indicate the supplementary variables
2. Create SDTM+SUPP and using the macro and Define document, create SDTM and SUPP from SDTM+SUPP
3. Validate SDTM+SUPP data set first and then SDTM and SUPP

OTHER CONCERNS

The SAS programmer might be requested by the sponsor to create a single SUPPQUAL data set. If the sponsor wants a single SUPPQUAL rather than separate SUPP, the SAS programmer needs to create SUPPQUAL in the end. Even at those occasions, it is recommended to create and validate SDTM+SUPP, and individual SUPP, then append all supplementary data sets to SUPPQUAL. When appending all the supplementary data sets, the SAS programmer needs to watch if there is any truncation due to the different length in the variables.

CONCLUSION

The validation of SUPPQUAL data set takes a lot of times and resource because of its size and complexity. However, through the automatic procedure, the validation will take less time and less resource. By creating SDTM+SUPP data sets and the automatic procedure, the SAS programmer can easily validate SDTM SUPPQUAL data set.

CONTACT INFORMATION

Your comments and questions are valued and welcomed. Please contact the author at

Kevin Lee
Covance Inc.
Conshohocken, PA
(610) 832 - 5751
Email:Kevin.lee@covance.com

TRADEMARKS

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries.

Oracle and Oracle Clinical are registered trademarks of Oracle Corporation, Redwood Shores, CA.

® indicates USA registration. Other brand and product names are registered trademarks or trademarks of their respective companies.