PharmaSUG China 2018 - Paper 79

My Understandings to SDTM and ADaM

Chunpeng Zhao, Boehringer-Ingelheim, Shanghai, China

ABSTRACT

Why are some diseases mapped into MH or AE, but others into CE? When we annotate data from CRF into SDTM variables, sometimes we are uncertain which domain a CRF field should be mapped into, Supplemental Qualifier Domain or Finding about Domain. How is clinical data grouped into different ADaM datasets? What should we consider when we try to create a new ADaM dataset? I will share my understandings in the paper.

TIMING AND PURPOSE IN SDTM

If a patient got cough, which domain the cough should be mapped into? It depends on timing.

1. It should be mapped into MH, if the cough started before Informed Consent (RFICDTC). The blue bar represents the period in which the cough that may start.

2. It should be mapped into AE, if the cough started after Informed Consent.

If a patient got cough, which domain the cough should be mapped into? It also depends on analysis purposes.

1. If the cough is only taken as a safety concern, it should be mapped into AE.
2. But it the cough is taken as an efficacy endpoint, e.g., study indication, it should be mapped into CE.

SUPPLEMENTAL QUALIFIER DOMAIN OR FINDING ABOUT DOMAIN

1. SUPPLEMENTAL QUALIFIER DOMAIN

For *Adverse Event of Special Interest in CRF,* Which domain is it should be mapped into?

Adverse Event of Special Interest (If event is an adverse event of special interest as defined in the protocol, complete an SAE form)	No ☐ Yes ☐

The answer must be Supplemental Qualifier Domain without hesitation.

Adverse Event of Special Interest (If event is an adverse event of special interest as defined in the protocol, complete an SAE form)	AESIGN in SUPPAE	No ☐ Yes ☐

Do you think why not map it into Finding about Domain?

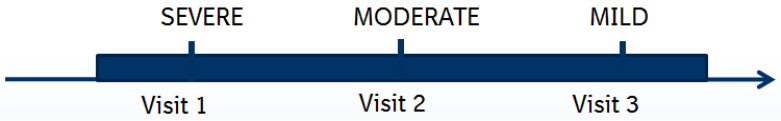
Adverse Event of Special Interest (If event is an adverse event of special interest as defined in the protocol, complete an SAE form)	FAORRES when FATESTCD="OCCUR" and FAOBJ="Adverse Event of Special Interest"	No ☐ Yes ☐

From a pragmatic point of view, it is much simpler and more straightforward to map into Supplemental Qualifier Domain instead of Finding about Domain. And *Adverse Event of Special Interest* is a qualifier to AETERM. Since there is no corresponding CDISC variable in AE main domain, it should be mapped into Supplemental Qualifier Domain, as the name suggests. AE main domain and AE Supplemental Qualifier Domain still belong to one domain. It is very easy to merge data from Supplemental Qualifier Domain back Main Domain into one dataset. However, Finding about Domain is another story. FAAE and AE are two separate domains. If we merge them into one dataset, we need additional information to link corresponding records, i.e., FAREFID & AEREFID. The relationship between

Main Domain and Finding about Domain is not so close-knit compared with that between Main Domain and Supplemental Qualifier Domain.

2. FINDING ABOUT DOMAIN

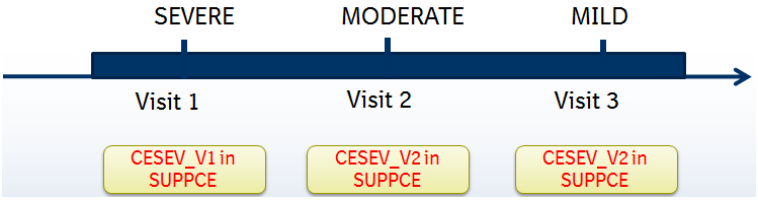
Cough is a clinical event of special interest. Which domain do you think its severity by visit should be mapped into?



Intuition may tell you the severity by visit should be mapped into FA.

VISIT	FATEST	FAOBJ	FAORRES
Visit 1	Severity/Intensity	Cough	SEVERE
Visit 2	Severity/Intensity	Cough	MODERATE
Visit 3	Severity/Intensity	Cough	MILD

Did you try to map it into Supplemental Qualifier Domain? It seems work too.



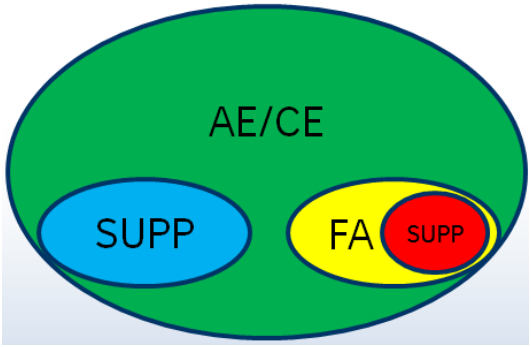
What is difference between these 2 mapping methods, FA and Supplemental Qualifier Domain?

1) FA: Cough severities across all visits are grouped by FATEST and FAOBJ. And Severity results display clearly by visit in vertical structure. Relationship of cough severities across visits is strengthened. It is much easier to review cough severity trend over time. But it is not straightforward to link cough severity in each visit with CETERM=Cough, because CETERM=Cough and its severity in each visit are mapped into separate domains, CE and FA.

2) Supplemental Qualifier Domain: Cough severity is each visit is mapped into corresponding Supplemental Qualifier variable. Their relationships with CETERM=Cough are strengthened, since they are in the same domain CE. However, Cough severities across visits, i.e., corresponding Supplemental Qualifier variables are not grouped together explicitly like FA. Relationship of cough severities across visits is weakened.

3. Relationships between Main Domain, Supplemental Qualifier Domain and Finding about Domain

1) Supplemental Qualifier Domain or Finding about Domain is not standalone topic domain. They are always linked with main domain under the same topic, even if corresponding records in main domain or the whole main domain dataset are not be created. Main domain, at least its concept itself should be decided in mind first in CRF SDTM annotation.



2) Relationship between main domain and Supplemental Qualifier Domain is much closer than that between main domain and FA, since Main Domain and Supplemental Qualifier Domain belong to the same domain logically. Some CRF data cannot be mapped into main domain, but into Supplemental Qualifier Domain. The only reason

is that there are not corresponding SDTM variables. Actually, Main Domain and Supplemental Qualifier Domain are always taken as one dataset, when we review data or do analysis.

3) Relationship between data within FA will be strengthened, both at variable level and at record level. For example, data are grouped by Variable FACAT, FAOBJ vertically. Relationships between variables are also self-explanatory through SDTM variable names, such as FAORRES, FAORRESC, and FADTC etc. horizontally.

FACAT	FAOBJ	FAORRES	FAORRESU	FADTC

METHODOLOGY OF CLINICAL DATA ORGANIZATION IN SDTM

1. SDTM data is organized by Topic first, then Data Structure.

1) Adverse Events and Concomitant Medications are different topics, so they are mapped into different domains AE and CM respectively. The same events occur in different periods. It could also be taken as different topics, e.g., MH and AE. A new analysis purpose could trigger new domain creation too. For example, an event is mapped into CE instead of AE, due to its efficacy analysis usage.

2) When clinical data cannot be mapped into SDTM main domain, data structure should be considered. If you want to strengthen its relationship with main domain, it should be mapped into Supplemental Qualifier Domain. If you want to strengthen the relationship between different parts itself, it should be mapped into FA.

2. SDTM annotation strategy

If you are uncertain where clinical data should be mapped into, I recommend following priorities:

1st Choice: Main Domain

2nd Choice: SUPP

3rd Choice: FA

4th Choice: SUPPFA

If you think Supplemental Qualifier Domain or Finding about Domain either is fine for an item in CRF, I prefer Supplemental Qualifier Domain with one non-SDTM variable. Otherwise, you have to create multiple variables, e.g. FATEST, FAOBJ, FAORRES, and FADTC etc. for it in FA. And its relationship with main domain would be weakened. It is not convenient to review data or to do analysis together with main domain, because it is not so straightforward to merge it back main domain as Supplemental Qualifier Domain.

PRACTICAL APPLICATION IN SDTM ANNOTATION

1. Randomization Number and Randomization Strata factors

Which domain should randomization number be mapped into? From the perspective of data structure, i.e., randomization number is at patient level, it should be mapped into SUPPDM. But from the perspective of topic, it should be mapped into SUPPDS, since randomization is a milestone in DS. SDTM data is organized by topic first, then data structure. So Randomization number should be mapped into SUPPDS. And Randomization Strata factors should follow the same rule to be mapped into SUPPDS too.

2. **Which do you think Butterfly-Stroke should be grouped with together, Butterfly or Freestyle-Stoke?**

It depends on which relationship you want to emphasize, the relationship between Butterfly-Stoke and Butterfly, or one between Butterfly-Stroke and Freestyle-Stroke.

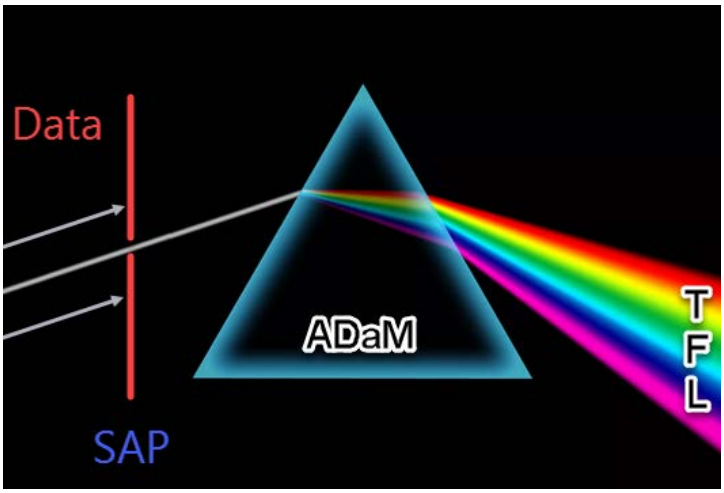
METHODOLOGY OF CLINICAL DATA ORGANIZATION IN ADAM

1. ADaM data is organized by Data Structure first, then Topic.

Is ADaM data also grouped into different datasets by topic as same as SDTM? When we read ADaM guides, we will find ADaM data is grouped by data structure first, ADSL, OCCDS and BDS (including ADTTE). All Time-to-Event parameters could be derived in one dataset ADTTE, no matter whether parameters are under the same topic. In practice, we create ADLB, ADVS and ADQS etc. following data structure BDS. We also create ADAE, ADCM and ADMH etc. following data structure OCCDS. So ADaM data is organized by Data Structure first, then Topic.

2. The Final Purpose of ADaM

We map each endpoint or derivation from SAP into ADaM. Then convert SDTM data into TFLs, i.e., Tables, Figures and Listings through ADaM. Two principles should be followed, Ready-for-Analysis and Traceability, for crystal transparent communication.



3. What could trigger new ADaM dataset creation?

Different source data structures under the same topic, e.g. e-diary data and CRF diary data, different analysis methods of the same endpoint, e.g. different imputation rules, a special table shell, even different programming logics or task assignments, etc. any of them may legally trigger our impulse to create a new ADaM dataset. But none of them should be a main criterion to group clinical data into a new ADaM dataset. We should still stick to the rule “ADaM data is organized by Data Structure first, then Topic”. Otherwise, it may undermine ADaM transparency. For example, if we group parameters into different ADaM datasets by programming logic instead of topic, it would not be easy to review or track all related parameters under the same topic.

4. Examples to create new ADaM datasets

1) Normally we create ADCM with OCCDS structure. We also could create another analysis concomitant dataset with BDS structure, if we need to summarize doses of some concomitant medications of special interest by period.

2) Normally ADLB is used for safety analysis. If some lab tests are used as biomarkers for efficacy analysis with special imputation rules, we also could create another ADaM dataset with the same BDS structure. Due

to different analysis purpose, biomarker can be taken as a new topic. And It will also facilitate data review and validation.

3) Sometimes we try to put all patient-level data into ADSL. ADSL including so many variables will become huge. It is better to separate some variables under the same topic into another ADaM dataset. For example, create BDS structure dataset without timing variables for disease characteristics analysis. Create BDS structure dataset ADEX for study treatment compliance analysis.

CONCLUSION

SDTM data is organized by Topic first, then Data Structure. The other way around, ADaM data is organized by Data Structure first, then Topic. Methodology of clinical data organization in SDTM and ADaM are driven by Data Clinical meanings and analysis needs for crystal transparent communication.

REFERENCES

Analysis Data Model Implementation Guide Version 1.1

Study Data Tabulation Model Implementation Guide: Human Clinical Trials Version 3.2

CONTACT INFORMATION

Your comments and questions are valued and encouraged. The SAS® code can be sent upon request. Contact the author at:

Name: Chunpeng Zhao
Enterprise: Boehringer Ingelheim
Address: 04/F, BenBen Mansion 300 Xikang
City, State ZIP: Shanghai, China
Work Phone: +86 (21) 5288-0209
E-mail: chunpeng.zhao@boehringer-ingelheim.com