



OCCDS – Creating flags or records

*Rob Wartenhorst
PhUSE– Barcelona 2016*

-
- ADAE, ADCM and ADMH originally set up using ADAE structure.
 - After release of OCCDS guide attempted to implement the occurrence flags
 - Soon we realised there is the potential to create a huge number of occurrence flags

Occurrence flags



What are they (from the IG)

Variable Name	Variable Label	Type	Code List / Controlled Terms	Core	CDISC Notes
AOCCFL	1st Occurrence within Subject Flag	Char	Y	Perm	Character indicator for the first occurrence of any event/intervention/finding within the subject. Example derivation: Sort the data in the required order and flag the first treatment emergent record for each subject.
AOCCSFL	1st Occurrence of SOC Flag	Char	Y	Perm	Character indicator for the first occurrence of the system organ class within the subject. Example derivation: Sort the data in the required order and flag the first treatment emergent record for each body system for each subject
AOCCPFL	1st Occurrence of Preferred Term Flag	Char	Y	Perm	Character indicator for the first occurrence of the preferred term within the subject. Example derivation: Sort the data in the required order and flag the first treatment emergent record for each --DECOD for each subject.
AOCCIFL	1st Max Sev./Int. Occurrence Flag
AOCCSIFL	1st Max Sev./Int. Occur Within SOC Flag
AOCCPIFL	1st Max Sev./Int. Occur Within PT Flag

Sample data with occurrence flags



USUBJID	AEDECOD	AEBODSYS	AOCCFL	AOCCSFL	AOCCPFL
01	Diarrhoea	Gastrointestinal disorders	Y	Y	Y
01	Diarrhoea	Gastrointestinal disorders			
01	Vomiting	Gastrointestinal disorders			Y
01	Dizziness	Nervous system disorders		Y	Y
01	Headache	Nervous system disorders			Y
01	Headache	Nervous system disorders			
01	Presyncope	Nervous system disorders			Y

Occurrence flags used in a table



System Organ Class	Preferred Term	Vaccine A (N=xxx) n(%)	Vaccine B (N=xxx) n(%)
At Least one symptom	AOCCFL	x (x.x)	x (x.x)
Gastrointestinal disorders		x (x.x)	x (x.x)
	At Least one symptom	AOCCSFL	x (x.x)
	Diarrhoea	x (x.x)	x (x.x)
	Vomiting	AOCCPFL	x (x.x)
Nervous system disorders			
	At Least one symptom	AOCCSFL	x (x.x)
	Dizziness	x (x.x)	x (x.x)
	Headache	AOCCPFL	x (x.x)
	Presyncope	x (x.x)	x (x.x)

```
PROC freq DATA = adae;
  BY trta;
  TABLES aoccf1;
RUN;
```

```
PROC freq DATA = adae;
  BY trta;
  TABLES aebodsys aoccsf1;
RUN;
```

```
PROC freq DATA = adae;
  BY trta;
  TABLES aedecod*aebodsys aoccpf1;
RUN;
```

How many flags do we need?



- AE analysis in vaccines has often:
 - Analysis of AE in x-y days after vaccination. E.g.
 - First event after a vaccination
 - First 7 days after vaccination,
 - From 8 days after vacc. until next vacc.
 - Within the analysis selection of x-y days after vaccination, a selection of ae's of special interest. In our standard we have **11** of such categories. E.g.
 - Related AE's,
 - AE's leading to withdrawal,
 - serious AE's etc..

How many flags do we need?



For just the 1st Occurrence within Subject Flag

Variable	Description
AOCCFL	1st Occurrence within Subject Flag
AOCC01FL	1st Occurrence within Subject related Flag
AOCC02FL	1st Occurrence within Subject after vaccination flag
AOCC03FL	1st Occurrence within Subject after vaccination related flag
AOCC04FL	1st Occurrence within Subject first 7 days flag
AOCC05FL	1st Occurrence within Subject first 7 days related flag
Etc...	

- For our example with 3 occurrence type flags, 3 timing categories and 11 categories of special interest we would need:
 - $3 * 3 + 3 = 12$ flags
 - $12 * 11 = 132$ flags
- When including the intensity flags (AOCCIFL) this number doubles to **264!!!!**

Issues with a large number of flags



- AOCCzzFL only allows 99 variables
- A lot of work
- Error prone
- Not transparent
- Difficult to keep the variable name and meaning the same across studies and projects

1. Not use the occurrence flags
 - Lose traceability
 - More logic needed in AResM and TFL programming
2. For each set of categories create records
 - Introduce variables **ACAT1** and **ACAT2** to identify the categories
 - ACATy is a categorization variable. From the I.G.:

Variable Name	Variable Label	Type	Code List / Controlled Terms	Core	CDISC Notes
ACATy	Analysis Category y	Char		Perm	Category used in analysis. May be derived from --CAT and/or --SCAT. Examples include records of special interest like prohibited medications, concomitant medications taken during an infusion reaction, growth factors, antimicrobial medications, and other such categories not defined elsewhere or present in SDTM domains.

- Definite set of 6 occurrence variables that always have the same meaning, not only within a company, but across all companies applying the occurrence data structure

Creating records



- Sample data, Headache occurring twice

AEDECOD	ASTDT	REL	ACAT1	ACAT2
HEADACHE	11OCT15	OVERALL	OVERALL	OVERALL
HEADACHE	09OCT16	OVERALL	OVERALL	OVERALL

Last step. Create records for second category, the ae's of special interest, create records and assign ACAT2

Copy original records Assign value "OVERALL" to both ACAT1 and ACAT2
 Create records and assign time period to ACAT1. Note a record can be assigned to multiple time periods

- ADAE dataset

AEDECOD	ASTDT	ACAT1	REL	AOCCSFL	AOCCPFL
HEADACHE	11OCT15	OVERALL	OVERALL	Y	Y
HEADACHE	09OCT16	OVERALL	OVERALL		
HEADACHE	11OCT15	AFTER VAC 1	OVERALL	Y	Y
HEADACHE	09OCT16	AFTER VAC 2	OVERALL	Y	Y
HEADACHE	11OCT15	OVERALL	RELATED AE	Y	Y
HEADACHE	11OCT15	OVERALL	SERIOUS AE	Y	Y
HEADACHE	09OCT16	OVERALL	LEAD TO WD	Y	Y
HEADACHE	11OCT15	AFTER VAC 1	RELATED AE	Y	Y
HEADACHE	11OCT15	AFTER VAC 1	SERIOUS AE	Y	Y
HEADACHE	09OCT016	AFTER VAC 2	LEAD TO WD	Y	Y

1
2
3

Sample specifications



B	C	D	E	I	J
DSET	PARAMID	VAR	VAR_LABEL	SOURCE	COMP_DESC
Dataset Name	Parameter Name	Variable Name	Variable Label	Variable Source	Computation Description
ADAE	Ident	ACAT1	Analysis Category 1	ASSIGNED	<ol style="list-style-type: none"> 1. Copy all records from AE for ACAT1 = 'OVERALL'1. 2. For all records created for ACAT1 , copy the records where ASPER is not missing and assign value "BY VACCINATION" concatenated with ASPER to ACAT1. 3. From the records created for ACAT1 copy the records where ASTDY between >= 1 and <= 30. Assign value "DAY 1-30" to ACAT1
ADAE		ACAT2	Analysis Category 2	ASSIGNED	<ol style="list-style-type: none"> 1. For all records created for ACAT1 assign value "OVERALL" to ACAT2. 2. For all records created for ACAT1 , copy the records where AE.AEREL EQ "Y" for ACAT2 = "RELATED". 3. For all records created for ACAT1 , copy the records where AE.AESER EQ "Y" for ACAT2 = "SERIOUS".
ADAE		AOCCFL	1st Occurrence within Subject Flag	DERIVED	For data where TRTEMFL = 'Y', sort the data by ACAT1, ACAT2, USUBJID, ASTDT, AESEQ. For the first observation within each USUBJID set to "Y". Otherwise set to null.
ADAE		AOCCPFL	1st Occurrence of Preferred Term Flag	DERIVED	For data where TRTEMFL = 'Y', sort the data by ACAT1, ACAT2, USUBJID, AEDECOD, ASTDT, AESEQ. For the first observation within each AEDECOD set to "Y". Otherwise set to null.
ADAE		AOCCSFL	1st Occurrence of SOC Flag	DERIVED	For data where TRTEMFL = 'Y', sort the data by ACAT1, ACAT2, USUBJID, AEBODSYS, ASTDT, AESEQ. For the first observation within each AEBODSYS set to "Y". Otherwise set to null.

- Creating the data in 3 steps:
 - Data step to create records for OVERALL and each desired time period (ACAT1)
 - Study specific code
 - Data step to create records for categories of AE's of special interest (ACAT2)
 - Study specific code
 - Sort data by ACAT1, ACAT2, other relevant standard variables and create the 6 standard occurrence flags.
 - This step is standard and should not need to be modified from study to study
- Analysis very straight forward. A frequency count by ACAT1 and ACAT2 gives all the counts for all tables in the analysis.
- Checking counts through a data viewer very easy by being able to subset on ACAT1 and ACAT2, together with the appropriate occurrence flag.

- Occurrence flags are useful for traceability
- Only create one flag per unique purpose! I.e. Same flag used for tables that summarize the same data, but with different subset
- By creating records instead of many flags:
 - Logical naming of flags makes writing analysis results metadata and programming easier
 - Easier to use same variable across trials
 - Significantly less specification and programming effort required
 - Meaningful text in ACATy variables makes it easier to use the data
- Running time of program, shorter when creating records, than when creating flags due to fewer sort steps needed.
- Dataset is bigger, but not to the extent that the dataset has become too big to work with
- Easier to add or remove categories by adding/removing records than variables
- Comment from a sceptical programmer: It has become too easy to create the TFL. Flags are standard, all that is needed is to select the correct ACAT values which have a meaningful text.
- Overall, little effort is required to create the adverse event summary tables

Any Questions?

Thank you