PharmaSUG 2015 - Paper DS08

Proper Parenting: A Guide in Using ADaM Flag/Criterion Variables and When to Create a Child Dataset

Richann Watson, Experis, Batavia, OH
Karl Miller, inVentiv Health, Lincoln, NE
Paul Slagle, inVentiv Health, Ann Arbor, MI

ABSTRACT

There has always been confusion as when to use some of the various flags (ANLzzFL, CRITyFL) or category variables (AVALCATy) reserved for ADaM basic data structure (BDS). Although some of these variables can be used interchangeably it might help to come up with rules to help keep a consistency of when to use these variables. Furthermore, there may be some situations where the creation of a new data set would be the better solution. These data sets are referred to as parent-child data sets throughout the paper. This paper will focus on rules that the authors follow when developing ADaM data set specifications for the proper use of ANLzzFL, CRITy/CRITyFL, AVALCATy or when a parent-child data set option is more feasible.

INTRODUCTION

During the development of ADaM data set specifications, the developer will make use of the Statistical Analysis Plan (SAP) and table, listing and figures (TLF) mock-ups to get a better understanding of how to structure the ADaM data sets. Without these documents it is difficult to adequately create data set specifications that will be both CDISC compliant and perform the necessary analysis. The SAP and TLF mock-ups will tell the developer key information that will be used for analysis. It can inform the developer if there are certain criterion that need to be met for specific parameters, how to handle multiple records at a visit, how to determine if a subject is in a specific population. All this information will then be taken to develop the specifications and during that development, decisions need to be made as to what variables are best for handling the analysis while maintaining CDISC compliance or if creating a new data set is the better option.

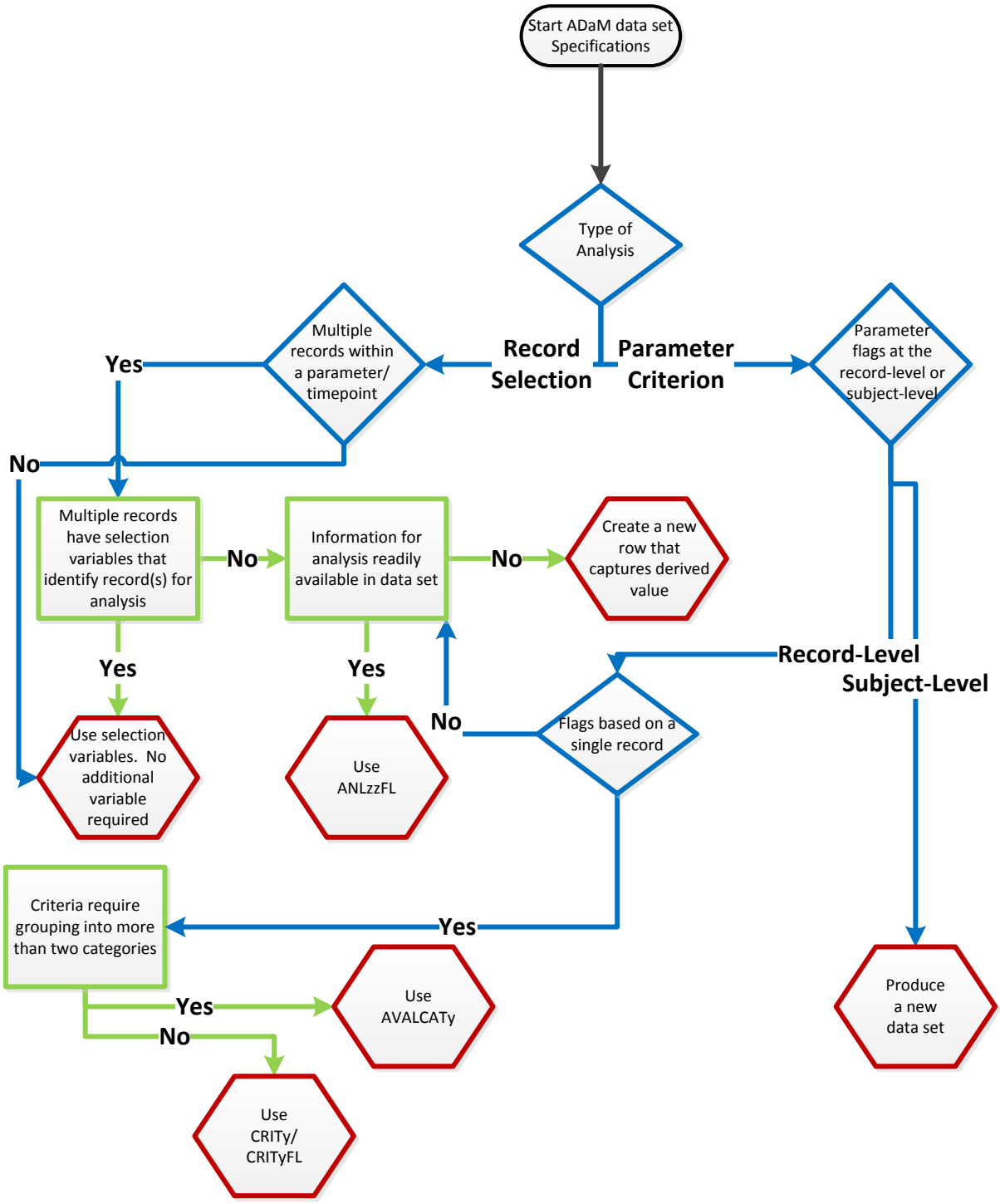
The ADaM Implementation Guide (ADaM IG) indicates that the analysis criterion variable and its corresponding flag (CRITy and CRITyFL) are to be used when determining if a specific criterion is met. However, the ADaM IG does not indicate if the criterion has to be based on one record, which can lead to confusion as to should the CRITy/CRITyFL be used or should ANLzzFL be used and what about using AVALCATy. A criterion is essentially grouping the data into two categories: criterion met or criterion not met. So establishing basic rules to follow will help to ensure consistency.

DETERMINING BEST OPTION

Although the ADaM IG allows for flexibility in how a data set is set up, sometimes determining the best option is not always easy and making a less than ideal decision can lead to disastrous results which can cause more work later on. When determining the best option, it is sometimes easier to just ask a series of questions to make the decision.

- What is the type of analysis?
- Are there multiple records within a parameter/timepoint?
- Do the multiple records have selection variables that identify the necessary record(s) needed for analysis?
- Is the information needed for analysis readily available in the data set?
- Are these parameters criterion flags at the record-level or subject-level?
- Are the record-level flags based on a single record?
- Does the criterion require grouping into more than two categories?

The response to each question will lead down a different path in a decision tree (Display 1), with each path leading to what is possibly the best option for the situation.



Display 1: Decision Tree

The rest of this paper will focus on providing examples and traversing the different paths of the decision tree, with the exception of the path that leads to 'No additional variable required'. This path is the most straight forward path and essentially indicates that all the necessary information/selection variables needed for analysis are already in the data set.

PATH TO NEW ROW

Regardless if you have a single parameter, or multiple parameters, that need flagging if the data is not readily available in the data set, then the creation of a new row is the best option. The creation of a new row can be in the form of either creating derived value with an existing parameter or creating a new parameter.

In this scenario, the ECG data is collected in triplicate at each visit, so that there are three results per visit in the SDTM EG domain. Using the EG along with the rules set forth in the ADaM IG, the ECG analysis data set (ADEG) will also have three results per visit. (Display 2).

	PARAMCD	PARAM	ADT	ATM	ADTM	ADY	AVISIT	AVISITN	AVAL
1	QTcB	QT interval using Bazett's correction(msec)	25-Feb-14	8:30	25FEB2014:08:30:24	1	Baseline	-2	449
2	QTcB	QT interval using Bazett's correction(msec)	25-Feb-14	8:31	25FEB2014:08:31:07	1	Baseline	-2	474
3	QTcB	QT interval using Bazett's correction(msec)	25-Feb-14	8:31	25FEB2014:08:31:41	1	Baseline	-2	477
4	QTcB	QT interval using Bazett's correction(msec)	27-Feb-14	9:13	27FEB2014:09:13:55	3	Day 3	0	457
5	QTcB	QT interval using Bazett's correction(msec)	27-Feb-14	9:14	27FEB2014:09:14:28	3	Day 3	0	469
6	QTcB	QT interval using Bazett's correction(msec)	27-Feb-14	9:14	27FEB2014:09:14:55	3	Day 3	0	456
7	QTcB	QT interval using Bazett's correction(msec)	13-Mar-14	9:29	13MAR2014:09:29:35	17	Week 2	2	500
8	QTcB	QT interval using Bazett's correction(msec)	13-Mar-14	9:30	13MAR2014:09:30:04	17	Week 2	2	495
9	QTcB	QT interval using Bazett's correction(msec)	13-Mar-14	9:30	13MAR2014:09:30:45	17	Week 2	2	480
10	QTcB	QT interval using Bazett's correction(msec)	3-Apr-14	9:18	03APR2014:09:18:26	38	Week 5	5	449
11	QTcB	QT interval using Bazett's correction(msec)	3-Apr-14	9:19	03APR2014:09:19:05	38	Week 5	5	460
12	QTcB	QT interval using Bazett's correction(msec)	3-Apr-14	9:19	03APR2014:09:19:32	38	Week 5	5	460

Display 2: ECG Example

The desire for the analysis is to use the average value of the three results at each visit. It may be tempting to just create a new variable QTCBAVG that contains the average value for each visit and then flag the first datetime record as the one to use in the analysis (Display 3). However, adding a new variable QTCBAVG violates the ADaM BDS rule #1 for adding a new variable (refer to section 4.2 of ADaM IG).

	PARAMCD	PARAM	ADT	ATM	ADTM	ADY	AVISIT	AVISITN	AVAL	QTCBAVG	ANL01FL
1	QTcB	QT interval using Bazett's correction(msec)	25-Feb-14	8:30	25FEB2014:08:30:24	1	Baseline	-2	449	466.7	Y
2	QTcB	QT interval using Bazett's correction(msec)	25-Feb-14	8:31	25FEB2014:08:31:07	1	Baseline	-2	474	466.7	
3	QTcB	QT interval using Bazett's correction(msec)	25-Feb-14	8:31	25FEB2014:08:31:41	1	Baseline	-2	477	466.7	
5	QTcB	QT interval using Bazett's correction(msec)	27-Feb-14	9:13	27FEB2014:09:13:55	3	Day 3	0	457	460.7	Y
6	QTcB	QT interval using Bazett's correction(msec)	27-Feb-14	9:14	27FEB2014:09:14:28	3	Day 3	0	469	460.7	
7	QTcB	QT interval using Bazett's correction(msec)	27-Feb-14	9:14	27FEB2014:09:14:55	3	Day 3	0	456	460.7	
9	QTcB	QT interval using Bazett's correction(msec)	13-Mar-14	9:29	13MAR2014:09:29:35	17	Week 2	2	500	491.7	Y
10	QTcB	QT interval using Bazett's correction(msec)	13-Mar-14	9:30	13MAR2014:09:30:04	17	Week 2	2	495	491.7	
11	QTcB	QT interval using Bazett's correction(msec)	13-Mar-14	9:30	13MAR2014:09:30:45	17	Week 2	2	480	491.7	
13	QTcB	QT interval using Bazett's correction(msec)	3-Apr-14	9:18	03APR2014:09:18:26	38	Week 5	5	449	456.3	Y
14	QTcB	QT interval using Bazett's correction(msec)	3-Apr-14	9:19	03APR2014:09:19:05	38	Week 5	5	460	456.3	
15	QTcB	QT interval using Bazett's correction(msec)	3-Apr-14	9:19	03APR2014:09:19:32	38	Week 5	5	460	456.3	

Display 3: Incorrect Way to Add Average Value - Adding New Variable (or column)

Since the average value of the three results at each visit is not readily available and adding a new variable is not CDISC compliant, a new record should be created that will capture the average result at each visit. Since the record is created based on the average, then DTYPE should be populated accordingly (Display 4). DTYPE can then be used as a selection variable when determining which record(s) should be analyzed.

	PARAMCD	PARAM	ADT	ATM	ADTM	ADY	AVISIT	AVISITN	AVAL	DTYPE
1	QTcB	QT interval using Bazett's correction(msec)	25-Feb-14	8:30	25FEB2014:08:30:24	1	Baseline	-2	449	
2	QTcB	QT interval using Bazett's correction(msec)	25-Feb-14	8:31	25FEB2014:08:31:07	1	Baseline	-2	474	
3	QTcB	QT interval using Bazett's correction(msec)	25-Feb-14	8:31	25FEB2014:08:31:41	1	Baseline	-2	477	
4	QTcB	QT interval using Bazett's correction(msec)	25-Feb-14			1	Baseline	-2	466.7	AVERAGE
5	QTcB	QT interval using Bazett's correction(msec)	27-Feb-14	9:13	27FEB2014:09:13:55	3	Day 3	0	457	
6	QTcB	QT interval using Bazett's correction(msec)	27-Feb-14	9:14	27FEB2014:09:14:28	3	Day 3	0	469	
7	QTcB	QT interval using Bazett's correction(msec)	27-Feb-14	9:14	27FEB2014:09:14:55	3	Day 3	0	456	
8	QTcB	QT interval using Bazett's correction(msec)	27-Feb-14			3	Day 3	0	460.7	AVERAGE
9	QTcB	QT interval using Bazett's correction(msec)	13-Mar-14	9:29	13MAR2014:09:29:35	17	Week 2	2	500	
10	QTcB	QT interval using Bazett's correction(msec)	13-Mar-14	9:30	13MAR2014:09:30:04	17	Week 2	2	495	
11	QTcB	QT interval using Bazett's correction(msec)	13-Mar-14	9:30	13MAR2014:09:30:45	17	Week 2	2	480	
12	QTcB	QT interval using Bazett's correction(msec)	13-Mar-14			17	Week 2	2	491.7	AVERAGE
13	QTcB	QT interval using Bazett's correction(msec)	3-Apr-14	9:18	03APR2014:09:18:26	38	Week 5	5	449	
14	QTcB	QT interval using Bazett's correction(msec)	3-Apr-14	9:19	03APR2014:09:19:05	38	Week 5	5	460	
15	QTcB	QT interval using Bazett's correction(msec)	3-Apr-14	9:19	03APR2014:09:19:32	38	Week 5	5	460	
16	QTcB	QT interval using Bazett's correction(msec)	3-Apr-14			38	Week 5	5	456.3	AVERAGE

Display 4: Correct Way to Add Average Value - Adding New Row

With the addition of the new record with DTYPE = “AVERAGE” the data needed for analysis is now readily available and the selection variables are sufficient to identify the correct record(s) to analyze.

PATH TO ANLZZFL

Some people see the analysis level flag (i.e. ANLzzFL) as a ‘tie breaker that is used only to determine which record should be used for analysis purposes. However, the ANLzzFL by definition is used to help identify a record when selection variables are not enough to adequately determine a record that is to be used for analysis. This implies that ANLzzFL can be used to select records where a specific criterion is met, which appears to be similar to the definition for CRITy/CRITyFL. There is a slight difference between these two types of variables and will be discussed in a following section.

So when should ANLzzFL be used over CRITy/CRITyFL?

Below are some questions to ask when determining if ANLzzFL is the right choice:

- Is the type of analysis for selecting a specific record?
  - If “Yes”, are there multiple records within a parameter/timepoint?
    - If “Yes”, do the multiple records have selection variables that identify record(s) for analysis?
      - If “No”, is the information needed for analysis readily available in the data set? If “Yes”, then ANLzzFL is the best option

Going back to the original ECG example (Display 2), the use of ANLzzFL is illustrated. In this scenario, the desire for the analysis is to use the maximum value from each of the three results at each visit and if there are two results that tie for the maximum, then the first result is used. Although there are unique identifiers for each record, these identifiers do not select the records that are to be used for analysis. This scenario could be handled in two ways:

1. creating a new parameter that specifies the maximum value, or
2. use ANLzzFL to select the correct record.

Option 1, would require the creation of a new row of data for each parameter/timepoint, which could be problematic if there are already a lot of data in the data set. Furthermore, the new row of data does not provide any additional information that is not readily available in the data set.

So if there are multiple records within the same parameter and/or timepoint and the data is readily available in the data set, but the current selection variables within the data set do not identify a record that is to be included in an analysis, then using ANLzzFL helps to select the correct record(s).

	PARAMCD	PARAM	ADT	ATM	ADTM	ADY	AVISIT	AVISITN	AVAL	ANL01FL
1	QTcB	QT interval using Bazett's correction(msec)	25-Feb-14	8:30	25FEB2014:08:30:24	1	Baseline	-2	449	
2	QTcB	QT interval using Bazett's correction(msec)	25-Feb-14	8:31	25FEB2014:08:31:07	1	Baseline	-2	474	
3	QTcB	QT interval using Bazett's correction(msec)	25-Feb-14	8:31	25FEB2014:08:31:41	1	Baseline	-2	477	Y
4	QTcB	QT interval using Bazett's correction(msec)	27-Feb-14	9:13	27FEB2014:09:13:55	3	Day 3	0	457	
5	QTcB	QT interval using Bazett's correction(msec)	27-Feb-14	9:14	27FEB2014:09:14:28	3	Day 3	0	469	Y
6	QTcB	QT interval using Bazett's correction(msec)	27-Feb-14	9:14	27FEB2014:09:14:55	3	Day 3	0	456	
7	QTcB	QT interval using Bazett's correction(msec)	13-Mar-14	9:29	13MAR2014:09:29:35	17	Week 2	2	500	Y
8	QTcB	QT interval using Bazett's correction(msec)	13-Mar-14	9:30	13MAR2014:09:30:04	17	Week 2	2	495	
9	QTcB	QT interval using Bazett's correction(msec)	13-Mar-14	9:30	13MAR2014:09:30:45	17	Week 2	2	480	
10	QTcB	QT interval using Bazett's correction(msec)	3-Apr-14	9:18	03APR2014:09:18:26	38	Week 5	5	449	
11	QTcB	QT interval using Bazett's correction(msec)	3-Apr-14	9:19	03APR2014:09:19:05	38	Week 5	5	460	Y
12	QTcB	QT interval using Bazett's correction(msec)	3-Apr-14	9:19	03APR2014:09:19:32	38	Week 5	5	460	

Display 5: Correct Use of ANLzzFL

In Display 5, since there are three records for each visit and only one should be used for analysis and it is not consistently first, second or third result then ANLzzFL will help to identify which record should be used. The algorithm for ANLzzFL should clearly indicate how to select the appropriate record which would be used for analysis by flagging that record by populating ANLzzFL = 'Y'.

Although the ANLzzFL can be used to identify records needed for analysis it can be misused. It should not be used to flag records that met a specific criterion but rather it should be used to select a record that is used for analysis. It is used to uniquely identify a record when the selection criteria are not enough. For example, if the user wanted to identify all the records where QTcB > 450 ms, then using ANL01FL would be considered inappropriate since a subject can still have multiple records that meet the criterion; therefore, a record for use in the analysis is not uniquely identified (Display 6). The records can still be identified but would require the use of additional ADaM approved variables.

	PARAMCD	PARAM	ADT	ATM	ADTM	ADY	AVISIT	AVISITN	AVAL	ANL01FL
1	QTcB	QT interval using Bazett's correction(msec)	25-Feb-14	8:30	25FEB2014:08:30:24	1	Baseline	-2	449	
2	QTcB	QT interval using Bazett's correction(msec)	25-Feb-14	8:31	25FEB2014:08:31:07	1	Baseline	-2	474	Y
3	QTcB	QT interval using Bazett's correction(msec)	25-Feb-14	8:31	25FEB2014:08:31:41	1	Baseline	-2	477	Y
4	QTcB	QT interval using Bazett's correction(msec)	27-Feb-14	9:13	27FEB2014:09:13:55	3	Day 3	0	457	Y
5	QTcB	QT interval using Bazett's correction(msec)	27-Feb-14	9:14	27FEB2014:09:14:28	3	Day 3	0	469	Y
6	QTcB	QT interval using Bazett's correction(msec)	27-Feb-14	9:14	27FEB2014:09:14:55	3	Day 3	0	456	Y
7	QTcB	QT interval using Bazett's correction(msec)	13-Mar-14	9:29	13MAR2014:09:29:35	17	Week 2	2	500	Y
8	QTcB	QT interval using Bazett's correction(msec)	13-Mar-14	9:30	13MAR2014:09:30:04	17	Week 2	2	495	Y
9	QTcB	QT interval using Bazett's correction(msec)	13-Mar-14	9:30	13MAR2014:09:30:45	17	Week 2	2	480	Y
10	QTcB	QT interval using Bazett's correction(msec)	3-Apr-14	9:18	03APR2014:09:18:26	38	Week 5	5	449	
11	QTcB	QT interval using Bazett's correction(msec)	3-Apr-14	9:19	03APR2014:09:19:05	38	Week 5	5	460	Y
12	QTcB	QT interval using Bazett's correction(msec)	3-Apr-14	9:19	03APR2014:09:19:32	38	Week 5	5	460	Y

Display 6: Incorrect Use of ANLzzFL

PATH TO CRITY AND CRITYFL

The ADaM IG indicates that "for analysis purpose, criteria are often defined to group results based on the collected value's relationship" (ADaM IG 2009, p. 76), that is, it indicates that the analysis criterion variable (CRITy) and its corresponding flag (CRITyFL) are to be used when determining if a specific criterion is met. As indicated previously, this is very similar to the definition for ANLzzFL. However, the ADaM IG indicates in Example 3 of section 4.7.1 that if the data needed to assess the criterion is on multiple rows, then a new row should be added. It is determined that CRITy and CRITyFL are to be used to group results but that is similar to AVALCATy. Again, this begs to ask which set of flags is the correct one to use in the data set.

Below are some questions to ask when determining if CRITy/CRITyFL is the right choice:

- Is the type of analysis for assessing a specific parameter criterion?
  - Is the parameter criterion flag at the record-level?
    - If "Yes", is criterion based on a single record?
      - If "Yes", does criterion require grouping into more than two categories? If "No", then CRITy/CRITyFL is the best option

Continuing with the original ECG example (Display 2), the use of CRITy/CRITyFL is illustrated. For analysis purposes the statistician wants to know if the QTcB value was greater than 450 ms, or not, as both are of interest for the analysis. Since the data should be considered analysis ready when producing the tables and figures, then it is best to create a variable that will group the results.

	PARAMCD	PARAM	ADT	ATM	ADTM	ADY	AVISIT	AVISITN	AVAL	CRIT1	CRIT1FL
1	QTcB	QT interval using Bazett's correction(msec)	25-Feb-14	8:30	25FEB2014:08:30:24	1	Baseline	-2	449	QTcB > 450ms	N
2	QTcB	QT interval using Bazett's correction(msec)	25-Feb-14	8:31	25FEB2014:08:31:07	1	Baseline	-2	474	QTcB > 450ms	Y
3	QTcB	QT interval using Bazett's correction(msec)	25-Feb-14	8:31	25FEB2014:08:31:41	1	Baseline	-2	477	QTcB > 450ms	Y
4	QTcB	QT interval using Bazett's correction(msec)	27-Feb-14	9:13	27FEB2014:09:13:55	3	Day 3	0	457	QTcB > 450ms	Y
5	QTcB	QT interval using Bazett's correction(msec)	27-Feb-14	9:14	27FEB2014:09:14:28	3	Day 3	0	469	QTcB > 450ms	Y
6	QTcB	QT interval using Bazett's correction(msec)	27-Feb-14	9:14	27FEB2014:09:14:55	3	Day 3	0	456	QTcB > 450ms	Y
7	QTcB	QT interval using Bazett's correction(msec)	13-Mar-14	9:29	13MAR2014:09:29:35	17	Week 2	2	500	QTcB > 450ms	Y
8	QTcB	QT interval using Bazett's correction(msec)	13-Mar-14	9:30	13MAR2014:09:30:04	17	Week 2	2	495	QTcB > 450ms	Y
9	QTcB	QT interval using Bazett's correction(msec)	13-Mar-14	9:30	13MAR2014:09:30:45	17	Week 2	2	480	QTcB > 450ms	Y
10	QTcB	QT interval using Bazett's correction(msec)	3-Apr-14	9:18	03APR2014:09:18:26	38	Week 5	5	449	QTcB > 450ms	N
11	QTcB	QT interval using Bazett's correction(msec)	3-Apr-14	9:19	03APR2014:09:19:05	38	Week 5	5	460	QTcB > 450ms	Y
12	QTcB	QT interval using Bazett's correction(msec)	3-Apr-14	9:19	03APR2014:09:19:32	38	Week 5	5	460	QTcB > 450ms	Y

Display 7: Correct Use of CRITy/CRITyFL

In Display 7, since the criterion is "QTcB > 450 ms" then either the record meets the criterion, CRITyFL = "Y", or it does not meet the criterion, CRITyFL = "N". Note that if only records greater than 450 ms was of interest for the analysis then it is not necessary to populate CRITy/CRITyFL for subjects that did not meet that criterion. The incorporation of these variables into the data set to flag the data accordingly, these variables can then be used as selection variables when producing the output.

Although the CRITy/CRITyFL variables are used to assess a specific criterion for a specific parameter, it should only look at the individual row and be based off of the analysis value on that row. To reiterate if the data that needs to be assessed resides across multiple rows, then CRITy/CRITyFL should not be used and other options should be considered.

	PARAMCD	PARAM	ADT	ATM	ADTM	ADY	AVISIT	AVISITN	AVAL	CRIT1	CRIT1FL
1	QTcB	QT interval using Bazett's correction(msec)	25-Feb-14	8:30	25FEB2014:08:30:24	1	Baseline	-2	449	Max QTcB	N
2	QTcB	QT interval using Bazett's correction(msec)	25-Feb-14	8:31	25FEB2014:08:31:07	1	Baseline	-2	474	Max QTcB	N
3	QTcB	QT interval using Bazett's correction(msec)	25-Feb-14	8:31	25FEB2014:08:31:41	1	Baseline	-2	477	Max QTcB	Y
4	QTcB	QT interval using Bazett's correction(msec)	27-Feb-14	9:13	27FEB2014:09:13:55	3	Day 3	0	457	Max QTcB	N
5	QTcB	QT interval using Bazett's correction(msec)	27-Feb-14	9:14	27FEB2014:09:14:28	3	Day 3	0	469	Max QTcB	Y
6	QTcB	QT interval using Bazett's correction(msec)	27-Feb-14	9:14	27FEB2014:09:14:55	3	Day 3	0	456	Max QTcB	N
7	QTcB	QT interval using Bazett's correction(msec)	13-Mar-14	9:29	13MAR2014:09:29:35	17	Week 2	2	500	Max QTcB	Y
8	QTcB	QT interval using Bazett's correction(msec)	13-Mar-14	9:30	13MAR2014:09:30:04	17	Week 2	2	495	Max QTcB	N
9	QTcB	QT interval using Bazett's correction(msec)	13-Mar-14	9:30	13MAR2014:09:30:45	17	Week 2	2	480	Max QTcB	N
10	QTcB	QT interval using Bazett's correction(msec)	3-Apr-14	9:18	03APR2014:09:18:26	38	Week 5	5	449	Max QTcB	N
11	QTcB	QT interval using Bazett's correction(msec)	3-Apr-14	9:19	03APR2014:09:19:05	38	Week 5	5	460	Max QTcB	Y
12	QTcB	QT interval using Bazett's correction(msec)	3-Apr-14	9:19	03APR2014:09:19:32	38	Week 5	5	460	Max QTcB	Y

Display 8: Incorrect Use of CRITy/CRITyFL

As illustrated in Display 8, the CRITy/CRITyFL variables were used to flag the maximum value for each visit. This is not the proper use of CRITy/CRITyFL because multiple records needed to be looked at in order to determine which record had the maximum value. Other options such as using ANLzzFL or creating a new record would be a more viable solution.

Note that there are additional rules and guidelines for how to populate CRITy/CRITyFL but the purpose of the paper is to only illustrate when it is best to use CRITy/CRITyFL and not how to populate the variables. For further details on populating the types of variables refer to the ADaM IG section 4.7.

PATH TO AVALCATY

It was previously determined that both CRITy/CRITyFL and AVALCATy can be used to group results into categories. Using the decision tree (Display 1), it was decided that if there were only two categories that CRITy/CRITyFL are the best solutions. Therefore, if there are more than two categories that the analysis value needs to be grouped into, then AVALCATy would be the ideal choice.

	PARAMCD	PARAM	ADT	ATM	ADTM	ADY	AVISIT	AVISITN	AVAL	AVALCAT1
1	QTcB	QT interval using Bazett's correction(msec)	25-Feb-14	8:30	25FEB2014:08:30:24	1	Baseline	-2	449	<= 450
2	QTcB	QT interval using Bazett's correction(msec)	25-Feb-14	8:31	25FEB2014:08:31:07	1	Baseline	-2	474	>450-480
3	QTcB	QT interval using Bazett's correction(msec)	25-Feb-14	8:31	25FEB2014:08:31:41	1	Baseline	-2	477	>450-480
4	QTcB	QT interval using Bazett's correction(msec)	27-Feb-14	9:13	27FEB2014:09:13:55	3	Day 3	0	457	>450-480
5	QTcB	QT interval using Bazett's correction(msec)	27-Feb-14	9:14	27FEB2014:09:14:28	3	Day 3	0	469	>450-480
6	QTcB	QT interval using Bazett's correction(msec)	27-Feb-14	9:14	27FEB2014:09:14:55	3	Day 3	0	456	>450-480
7	QTcB	QT interval using Bazett's correction(msec)	13-Mar-14	9:29	13MAR2014:09:29:35	17	Week 2	2	500	>480-500
8	QTcB	QT interval using Bazett's correction(msec)	13-Mar-14	9:30	13MAR2014:09:30:04	17	Week 2	2	495	>480-500
9	QTcB	QT interval using Bazett's correction(msec)	13-Mar-14	9:30	13MAR2014:09:30:45	17	Week 2	2	480	>450-480
10	QTcB	QT interval using Bazett's correction(msec)	3-Apr-14	9:18	03APR2014:09:18:26	38	Week 5	5	449	<= 450
11	QTcB	QT interval using Bazett's correction(msec)	3-Apr-14	9:19	03APR2014:09:19:05	38	Week 5	5	460	>450-480
12	QTcB	QT interval using Bazett's correction(msec)	3-Apr-14	9:19	03APR2014:09:19:32	38	Week 5	5	460	>450-480

Display 9: Correct Use of AVALCATy

In Display 9**Display 7**, the QTcB needed to be categorized into one of four categories:

- QTcB less than or equal to 450 (<= 450)
- QTcB greater than 450 but less than or equal to 480 (>450-480)
- QTcB greater than 480 but less than or equal to 500 (>480-500)
- QTcB greater than 500 (>500)

Because there were four groups in which the QTcB needed to be categorized, AVALCATy is the best option for preparing the data to be analysis ready. With the incorporation of this variable into the data, AVALCATy can then be used as a selection variable when producing the output.

AVALCATy should only be used when the analysis value (AVAL or AVALC) needs to be categorized. If a category is based off of AVAL/AVALC and/or some other variable, then AVALCATy should not be used.

	PARAMCD	PARAM	ADT	ATM	ADTM	ADY	AVISIT	AVISITN	AVAL	ABLFL	BASE	CHG	AVALCAT1
1	QTcB	QT interval using Bazett's correction(msec)	25-Feb-14	8:30	25FEB2014:08:30:24	1	Baseline	-2	449				
2	QTcB	QT interval using Bazett's correction(msec)	25-Feb-14	8:31	25FEB2014:08:31:07	1	Baseline	-2	474				
3	QTcB	QT interval using Bazett's correction(msec)	25-Feb-14	8:31	25FEB2014:08:31:41	1	Baseline	-2	477	Y			
4	QTcB	QT interval using Bazett's correction(msec)	27-Feb-14	9:13	27FEB2014:09:13:55	3	Day 3	0	457		477	-20	QTcB <= 480 and CHG <= 20
5	QTcB	QT interval using Bazett's correction(msec)	27-Feb-14	9:14	27FEB2014:09:14:28	3	Day 3	0	469		477	-8	QTcB <= 480 and CHG <= 20
6	QTcB	QT interval using Bazett's correction(msec)	27-Feb-14	9:14	27FEB2014:09:14:55	3	Day 3	0	456		477	-21	QTcB <= 480 and CHG <= 20
7	QTcB	QT interval using Bazett's correction(msec)	13-Mar-14	9:29	13MAR2014:09:29:35	17	Week 2	2	500		477	23	QTcB > 480 or CHG > 20
8	QTcB	QT interval using Bazett's correction(msec)	13-Mar-14	9:30	13MAR2014:09:30:04	17	Week 2	2	495		477	18	QTcB > 480 or CHG > 20
9	QTcB	QT interval using Bazett's correction(msec)	13-Mar-14	9:30	13MAR2014:09:30:45	17	Week 2	2	480		477	3	QTcB <= 480 and CHG <= 20
10	QTcB	QT interval using Bazett's correction(msec)	3-Apr-14	9:18	03APR2014:09:18:26	38	Week 5	5	449		477	-28	QTcB <= 480 and CHG <= 20
11	QTcB	QT interval using Bazett's correction(msec)	3-Apr-14	9:19	03APR2014:09:19:05	38	Week 5	5	460		477	-17	QTcB <= 480 and CHG <= 20
12	QTcB	QT interval using Bazett's correction(msec)	3-Apr-14	9:19	03APR2014:09:19:32	38	Week 5	5	460		477	-17	QTcB <= 480 and CHG <= 20

Display 10: Incorrect Use of AVALCATy

In Display 10, we show an example of an improper use of AVALCATy. The QTcB records needed to be categorized into one of two categories:

- QTcB less than or equal to 480 with change from baseline less than or equal to 20 (AVAL <= 450 and CHG <= 20)
- Either QTcB greater than 480 or change from baseline greater than 20 (AVAL > 480 or CHG > 20)

Since the grouping criterion is also based on CHG, then AVALCATy should not be used. One of the other methods would be more appropriate, such as using CRITy/CRITyFL.

PATH TO NEW DATA SET

Although some clients may prefer to keep the number of ADaM data sets to a minimum, it may not always be the ideal solution to force additional information or flags into an existing data set. Sometimes when trying to prepare the data to be analysis ready, the best option may to just produce a new data set. Here are some things to consider, when determining if a new data set is the most feasible choice:

- Are there multiple parameters that need to be flagged?
- Are these parameter flags at the record-level or subject-level?
- If record-level, then one of the other paths previously discussed may be ideal. Note that depending on the number of parameters to flag it may still be ideal to create a new data set, but exploring the other paths should be the first option.
- If subject-level, then producing a new data set is the best option.

When parameter flag is at the record-level, then individual results are flagged according to a criterion and within a subject and parameter there can be different values, as illustrated in Display 7and Display 9. When a parameter flag is at the subject-level there is only one flag per subject for that parameter, this is what will be illustrated in this section.

In a hypothetical situation, a study required several analyses that were not easily obtained from the existing ADaM data sets.

1. Determine if a subject was eligible for assessment for specific parameters across various data sets.
2. Determine if a subject ever met a specific criterion for certain parameters across various data sets. Whether or not the criterion was met for eligible subjects would be used as covariates in a model.
3. Produce a listing of all the subjects that were eligible for assessment and whether or not they met the criterion by end of study for each parameter.

This information could have been incorporated into the main/parent data set (i.e. ADEG, ADLB and ADVS) by using various flags; however, this could lead to several issues:

- Too many flags being added to an existing data set especially if there were numerous parameters that were to be flagged. With too many flags, the data set gets harder to understand and all the flags could lead to confusion when viewing the data.
- If the flags are coming from numerous sources when trying to incorporate as covariates in a model, this can lead to having to merge numerous data sets together during the development of the output, which can make it difficult to review when the data is coming from multiple sources.
- How do we know if the subject was even eligible for assessment for that criterion? For example, if flags are only set to when a subject met the criterion at a post-baseline visit, then how do we determine the subjects that did have post-baseline visits but never met the criterion?

Because of these issues, it was decided to create a 'child' data set that is built off of ADEG, ADLB and ADVS (parent data sets). This allows all the flags from the various sources to be stored in one data set and this one data set can then be used to bring in all the necessary covariates for the modeling. In addition, the data set contains information as to whether or not the criterion was met for each parameter for each subject if it is met it contains information as to when it the criterion was first met.

	PARAMCD	PARAM	ADT	ADY	AVISIT	AVISITN	AVALC
1	QTCB1FL	QTcB > 450	27-Feb-14	3	Day 3	0	Y
2	QTCB2FL	QTcB > 480	13-Mar-14	17	Week 2	2	Y
3	QTCB3FL	QTcB > 500					N
4	QTCF1FL	QTcB > 450	27-Feb-14	3	Day 3	0	Y
5	QTCF2FL	QTcB > 480	13-Mar-14	17	Week 2	2	Y
6	QTCF3FL	QTcB > 500					N
7	HBA1C1FL	HbA1c < 6.5%	13-Mar-14	17	Week 2	2	Y
8	HBA1C2FL	HbA1c < 7%	13-Mar-14	17	Week 2	2	Y
9	HBA1C3FL	HbA1c < 8.5%	13-Mar-14	17	Week 2	2	Y
10	BMI1FL	BMI < 30 kg/m2					N
11	BMI2FL	BMI < 35 kg/m2					N

Display 11: Creating a New Data set

In Display 11, the data set is being created off parent data sets: ADEG, ADLB and ADVS. If the subject met the specific criterion indicated post-baseline, then the date and visit in which the criterion is met is captured and AVALC is set to "Y" to indicate that the subject met the criterion post-baseline. If by the end of the study a subject has had at least one post-baseline assessment for the parameter, but does not met the specific criterion, then a record is created with AVALC = "N" to indicate that subject was eligible for assessment but did not meet the criterion. This approach allows for the quick determination of which subjects were eligible for assessment for that particular parameter and it allows for the easy incorporation of covariates into a model and it allows for a listing to be produced from that lists all the required information.

CONCLUSION

The ADaM BDS provides us with several options to help with the generation of TFLs for analysis. These variables have different uses and at times can be confusing as to when it is appropriate to use one type of variable over another. Case in point, using ANLzzFL can help select the correct record for analysis but may not be the ideal choice. Other times, the analysis may be based off of a specific criterion or the analysis may involve the categorization of a result; in these cases CRITy/CRITyFL or AVALCATy would be the better choice. However, there are still situations where the creation of a new row within an existing ADaM data set, or the creation of a child data set that is built off of other ADaM data sets, are better suited for the analysis.

With all these options, developing ADaM data sets that are both useful for analyses, while still being CDISC compliant, does not have to be a daunting task. Although it may not always be easy to determine the best option, making the right choice for building ADaM data sets and stepping back and asking "What is the end goal? What am I trying to accomplish?" can help lead to the right decision. Through the use of the decision tree in creating ADaM specifications it can help remove the guess work on which variables would best fit your need by laying out the path based on the individual analysis need.

REFERENCES

- Analysis Data Model (ADaM) version 2.1
  http://www.cdisc.org/adam
- Analysis Data Model (ADaM) Implementation Guide version 1.0
  http://www.cdisc.org/adam

ACKNOWLEDGMENTS

Thanks to Debbie Nunner for walking through the weeds with her studies on these implementation issues.

RECOMMENDED READING

Analysis Data Model (ADaM) Implementation Guide version 1.0
http://www.cdisc.org/adam

CONTACT INFORMATION

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

Richann Watson
Experis
richann.watson@experis.com

Karl Miller
inVentiv Health
karl.miller@inventivhealth.com

Paul Slagle
inVentiv Health
paul.slagle@inventivhealth.com

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. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.