PharmaSUG 2013 - Paper PO02

# Derived observations and associated variables in ADaM datasets.

Arun Raj Vidhyadharan, inVentiv Health Clinical, Somerset, NJ

## ABSTRACT

This paper focuses on the observations that we derive in ADaM datasets for various reporting and statistical requirements. It also gives an insight to the derived variables that are associated with these derived observations and explains how they are related to each other.

## INTRODUCTION

We often produce ADaM datasets from their parent SDTM datasets. And in this process, we see that the resultant ADaM datasets grow in size. It seems like an average sized human being turned into a professional bodybuilder on steroids!

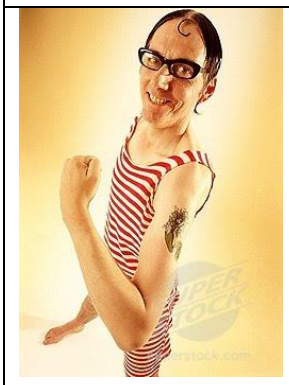
| SDTM dataset | Derivation Process | ADaM Dataset |
|---|---|---|
|  |  |  |

Figure 1

So how does the ADaM datasets grow in size? Obviously not steroids!

When the datasets grow in width, we know that there are new variables that were not present in the parent SDTM datasets and we call them as “Derived Variables”. And when the datasets grow in length, we know that there are new observations that came in and we call them as “Derived Observations”. Some derived variables are associated with the derived observations and their purpose is solely to give more information on the derived observations. The focus of this paper is on the observations that we derive in ADaM datasets and the variables associated with those derived observations.

The scope of this topic is well beyond the contents of this paper. There are limitless possibilities of deriving observations in ADaM datasets based on the requirements for reporting and other statistical analysis. The handful of scenarios described in this paper is based on my experience in my studies and the pharmaceutical companies that I have worked with.

First, let’s take a look at some common derived variables associated with derived observations for providing additional information.

## ANALYSIS DESCRIPTOR VARIABLES

### 1) PARAM:

PARAM is the unique identifier for the analysis parameter. It provides the unique description of the parameter with the

unit when applicable.

**2) PARAMCD:**

PARAMCD is its corresponding short name.

**3) PARAMN:**

PARAMN is the numeric form of the parameter.

*Note that neither PARAMCD nor PARAMN are CDISC required variables like PARAM. If they are present, a 1-to-1 correspondence with PARAM is required.*

**4) PARAMTYP:**

Codelist PT: PARAMTYP

Codelist Description: Parameter Type

Codelist Definition: Indicates whether the parameter is derived as a function of one or more other parameters

Extensible? No

| CDISC PT | Definition | Synonym |
|---|---|---|
| DERIVED | Indicates that a parameter is derived as a function of one or more other parameters. | Derived |

Table 1

This derived variable tells if an observation in a dataset is derived or observed (that came from the raw dataset). If an observation is derived, then the value for PARAMTYP will reflect “DERIVED” or else it will be blank. This variable is pretty straight forward and easy to understand as it typically holds one of the two values mentioned above.

**5) DTYPE:**

Codelist PT: DTYPE

Codelist Description: Derivation type

Codelist Definition: Analysis value derivation method

Extensible? Yes

| CDISC PT | Definition | Synonym |
|---|---|---|
| LOCF | Last Observation Carried Forward is a data imputation technique which populates missing values with the subject's previous nonmissing value | Last Observation Carried Forward |
| WOCF | Worst Observation Carried Forward is a data imputation technique which populates missing values with the subject's worst-case nonmissing value | Worst Observation Carried Forward |
| SOCF | Screening Observation Carried Forward is a data imputation technique which populates missing values with the subject's nonmissing screening observation | Screening Observation Carried Forward |
| BC | Best Case: A data imputation technique which populates missing values with the best possible outcome. | Best Case Imputation Technique |
| BLOCF | Baseline Observation Carried Forward is a data imputation technique which populates missing values with the subject's nonmissing baseline observation | Baseline Observation Carried Forward |

| | | |
|---|---|---|
| BOCF | Best Observation Carried Forward: A data imputation technique which populates missing values with the subject's best-case nonmissing value. | Best Observation Carried Forward Imputation Technique |
| ML | Maximum Likelihood is a data imputation technique which populates missing values with estimates that maximize the probability of observing what has in fact been observed | Maximum Likelihood |
| MI | Multiple Imputation is a data imputation technique which populates missing values with the average value from multiple imputed datasets which were derived from the nonmissing observed data. Bootstrapping is an example of multiple imputation. | Multiple Imputation |
| WC | Worst Case is a data imputation technique which populates missing values with the worst possible outcome | Worst Case |
| MOTH | Mean of Other Group is a data imputation technique which populates missing values with the mean value from a comparator or reference group | Mean of Other Group |
| POCF | Penultimate Observation Carried Forward is a data imputation technique which populates missing values with the subject's next-to-last nonmissing value | Penultimate Observation Carried Forward |
| WOV | Worst Observed Value in a Group is a data imputation technique which populates missing values with the worst value observed in a group of subjects | Worst Observed Value in a Group |
| MOV | Mean Observed Value in a Group is a data imputation technique which populates missing values with the mean value observed in a group of subjects | Mean Observed Value in a Group |
| INTERP | Interpolation is a method of imputation involving a missing value that is between known values and is estimated by a function of those known values. | Interpolation |
| ENDPOINT | Endpoint is a data derivation technique which calculates a subject's analysis end point value. | Endpoint |
| MINIMUM | Minimum is a data derivation technique which calculates a subject's minimum value over a defined set of records. | Minimum |
| MAXIMUM | Maximum is a data derivation technique which calculates a subject's maximum value over a defined set of records. | Maximum |

| | | |
|---|---|---|
| AVERAGE | Average is a data derivation technique which calculates a subject's average value over a defined set of records. | Average |

Table 2

Now, let's see the various derived observations and how the above mentioned variables describe these derived observations.

## DERIVED OBSRVATIONS

**Derived Observation for BASELINE**

| USUBJID | AVISIT | ABLFL | PARAM | AVAL | BASE | DTYPE | PARAMTYP |
|---|---|---|---|---|---|---|---|
| 1001 | Week -2 | | Heart Rate | 70 | | | |
| 1001 | Week -1 | | Heart Rate | 72 | | | |
| 1001 | Baseline | Y | Heart Rate | 71 | 71 | AVERAGE | DERIVED |

Table 3

Baseline derived observation can be found in ADEGM, ADVS, ADLB, ADEX and other efficacy datasets.

**Derived Observation for ANY VISIT AFTER START OF TREATMENT**

Some pharmaceutical companies prefer to have a derived observation for "Any Visit after start of treatment" for reporting purposes. The example shown below is based on a worst observation carried forward approach.

| USUBJID | AVISIT | ABLFL | PARAM | AVAL | BASE | DTYPE | PARAMTYP |
|---|---|---|---|---|---|---|---|
| 1001 | Week -2 | | LPH SCORE TOTAL | 70 | | | |
| 1001 | Week -1 | | LPH SCORE TOTAL | 72 | | | |
| 1001 | Baseline | Y | LPH SCORE TOTAL | 71 | 71 | AVERAGE | DERIVED |
| 1001 | Week 2 | | LPH SCORE TOTAL | 70 | 71 | | |
| 1001 | Week 3 | | LPH SCORE TOTAL | 65 | 71 | | |
| 1001 | Week 4 | | LPH SCORE TOTAL | 79 | 71 | | |
| 1001 | Any Visit after start of treatment | | LPH SCORE TOTAL | 65 | 71 | WOCF | DERIVED |

Table 4

**Derived Observation for LAST VISIT**

| USUBJID | AVISITN | AVISIT | PARAM | AVAL | ANL01FL | DTYPE | PARAMTYP |
|---|---|---|---|---|---|---|---|
| 1001 | 1 | Week 1 | Osteopontin | 70 | Y | | |
| 1001 | 2 | Week 2 | Osteopontin | 70 | Y | | |
| 1001 | 3 | Week 3 | Osteopontin | 70 | Y | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1001 | 4 | Week 4 | Osteopontin | 70 | Y | | |
| 1001 | 5 | Week 5 | Osteopontin | 72 | Y | | |
| 1001 | 999 | Last Visit | Osteopontin | 72 | Y | LOCF | DERIVED |

Table 5

A variant of the above scenario can be observed when the definition of Last Visit differs between safety reporting and efficacy reporting. The derived variable AVISITN can be utilized to accommodate such requirements as shown below.

| USUBJID | AVISITN | AVISIT | PARAM | AVAL | ANL01FL | ANL02FL | DTYPE | PARAMTYP |
|---|---|---|---|---|---|---|---|---|
| 1001 | 1 | Week 1 | Osteopontin | 70 | Y | Y | | |
| 1001 | 2 | Week 2 | Osteopontin | 70 | Y | Y | | |
| 1001 | 3 | Week 3 | Osteopontin | 70 | Y | Y | | |
| 1001 | 4 | Week 4 | Osteopontin | 70 | Y | Y | | |
| 1001 | 5 | Week 5 | Osteopontin | 72 | Y | Y | | |
| 1001 | 998 | Last Visit | Osteopontin | 72 | N | Y | ENDPOINT | DERIVED |
| 1001 | 999 | Last Visit | Osteopontin | 72 | Y | N | LOCF | DERIVED |

Table 6

The observations with AVISITN value 998 is removed while using the dataset for safety reporting and observations with AVISITN value 999 is removed while using the dataset for efficacy reporting. We can utilize the analysis variables ANL01FL and ANL02FL for implementing this.

**Derived Observation for Derived Tests**

Derivation of new tests in Lab dataset based on values of existing tests and other parameters. A classic example is the derivation of test Estimated Glomerular Filtration Rate (EGFR) based on MDRD method or Cockroft and Gault method in Cardio Vascular studies. Derivation of this test value uses the value of Creatinine test, age, sex and race of subjects.

MDRD method uses the formula:

EGFR value = 186*(Creatinine value**-1.154)*(Age**-0.203)*Sex factor*Race factor.

Where:

Sex factor is 1 for Male and 0.742 for Female

Race factor is 1.212 for Black and 1 for others

| USUBJID | AVISIT | PARAM | AVAL | AVALU | LBMETHOD | DTYPE | PARAMTYP |
|---|---|---|---|---|---|---|---|
| 1001 | Week 1 | Creatinine | 0.86 | MG/DL | | | |
| 1001 | Week 1 | Estimated Glomerular Filtration Rate | 72.33 | ML/MIN/1.73M2 | CALCULATED (MDRD) | CALCULATED | DERIVED |

Table 7

**Derived Observations for EXTENT OF EXPOSURE and TREATMENT DURATION**

These 2 derived observations are specific to the analysis dataset for Exposure domain. First, let's take a look at the definitions of "Extend of Exposure" and "Treatment Duration" for a better understanding of the case.

Extend of Exposure: The overall period during which the subject received medication. This is usually obtained by

counting the number of days between the first medication date and the last medication date without considering non-medication days in between.

Treatment Duration: This is the actual number of days a subject took medication.

A prototype of these derived observation looks like this:

| USUBJID | EX SEQ | PARAM | EXDOSE | EXDOSU | EXSTDTC | EXENDTC | EXDUR | DTYPE | PARAMTYP |
|---|---|---|---|---|---|---|---|---|---|
| 1001 | 1 | DRUG A | 30 | mg | 11-Jan-2012 | 01-Feb-2012 | 22 | | |
| 1001 | 2 | DRUG A | 30 | mg | 02-Feb-2012 | 15-Feb-2012 | 14 | | |
| 1001 | 3 | DRUG A | 30 | mg | 21-Feb-2012 | 06-Mar-2012 | 15 | | |
| 1001 | 4 | DRUG A | 30 | mg | 07-Mar-2012 | 14-Mar-2012 | 8 | | |
| 1001 | | Extent of exposure | | | 11-Jan-2012 | 14-Mar-2002 | 64 | CALCULATED | DERIVED |
| 1001 | | Treatment Duration | | | | | 59 | CALCULATED | DERIVED |

Table 8

***A word of caution:***

In theory, Treatment Duration will always be equal to or less than Extend of Exposure as we see in the ideal example above. However, the worst enemy of any statistical programmer “Data Error” could play a wicked role here and contradict this theory. Let’s see how this happens:

In an ideal scenario, the start date of medication in a medication record will be greater than the end date of medication in the previous medication record. But sometimes this is violated due to data error which could result in overlapping of two or more medication periods as shown in the below example.

| USUBJID | EX SEQ | PARAM | EXDOSE | EXDOSU | EXSTDTC | EXENDTC | EXDUR | DTYPE | PARAMTYP |
|---|---|---|---|---|---|---|---|---|---|
| 1001 | 1 | DRUG A | 30 | mg | 11-Jan-2012 | 01-Feb-2012 | 22 | | |
| 1001 | 2 | DRUG A | 30 | mg | 02-Feb-2012 | 15-Feb-2012 | 14 | | |
| 1001 | 3 | DRUG A | 30 | mg | 13-Feb-2012 | 06-Mar-2012 | 23 | | |
| 1001 | 4 | DRUG A | 30 | mg | 07-Mar-2012 | 14-Mar-2012 | 8 | | |
| 1001 | | Extent of exposure | | | 11-Jan-2012 | 14-Mar-2002 | 64 | CALCULATED | DERIVED |
| 1001 | | Treatment Duration | | | | | 67 | CALCULATED | DERIVED |

Table 9

An additional programming logic is required in this case to detect any overlap and then compute the Treatment Duration. A sample piece of code that implements this is given below:

```
data EX;
    set ADEX;
    by usubjud sm_id;
    retain TRTDUR;
    pre_edt=lag(sm_edt);
    if first.usubjid then do;
        pre_edt=.;
        _TRTDUR=sm_edt-sm_sdt+1;
    end;
    else do;
        if sm_sdt eq pre_edt then _TRTDUR=sm_edt-sm_sdt;
        else if sm_sdt gt pre_edt then _TRTDUR=sm_edt-sm_sdt+1;
        else if sm_sdt lt pre_edt then _TRTDUR=sm_edt-pre_edt;
    end;
    if first.usubjid then TRTDUR=_TRTDUR;
    else TRTDUR+_TRTDUR;
    EXDUR=TRT1ENDT-TRT1STDT+1;
run;
```

## CONCLUSION

This paper outlines the possibilities of deriving observations in ADaM datasets for reporting and other statistical purposes. The rules and relationships between the derived observations and associated variables provide great flexibility for statistical programmers to represent derived data in a consistent and meaningful manner.

## REFERENCES


- Author name: Freimark, Nate. “Common Misunderstandings about ADaM Implementation.”*PharmaSUG 2012 – Paper DS16*. Available at http://www.pharmasug.org/proceedings/2012/DS/PharmaSUG-2012-DS16.pdf
- Author name: Teng, Christine. “ADaM Standard Naming Conventions are Good to Have.”*PharmaSUG 2011 – Paper CD12*. Available at http://www.pharmasug.org/proceedings/2011/CD/PharmaSUG-2011-CD12.pdf
- Author name: last name, first name>. “<Title>.” <*Source*>. <Date>. Available at http://evs.nci.nih.gov/ftp1/CDISC/ADaM/ADaM%20Terminology.html
- Author name: Xu, Beilei. “A Taste of ADaM.” Available at http://analytics.ncsu.edu/sesug/2010/SDA03.Beilei.pdf


## ACKNOWLEDGEMENTS

The author would like to thank John Durski and Nancy Brucken for their review of this paper.

## CONTACT INFORMATION

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

Name: Arun Raj Vidhyadharan
Enterprise: inVentiv Health Clinical
Address: 500 Atrium Drive
City, State ZIP: Somerset, NJ 08873
Work Phone: 732.652.3490
E-mail: arunraj.vidhyadharan@inventivhealth.com