CDISCPILOT02 – Updated for ADaMIG v1.1

Test Data Factory Project Team

September 2018

[1 Introduction 2](#_Toc525821947)

[2 Specific Comments 3](#_Toc525821948)

[2.1 Split datasets 3](#_Toc525821949)

[2.2 Renamed Datasets 3](#_Toc525821950)

[2.3 Missing Information in define.xml 3](#_Toc525821951)

[3 Data Conformance Summary – Explanation on Remaining Pinnacle 21 Findings 5](#_Toc525821952)

# Introduction

The PhUSE organization initiated the Test Data Factory (TDF) project to provide test data for CDISC-based processes and programs publicly available. The CDISCPILOT02 datasets from 2013 were updated to conform to newer standards. The TDF project team used the following approach:

* The team ran the datasets through Pinnacle 21 Community v2.2 using ADaM 1.0 as the Configuration and 2016-06-24 as the CDISC CT. Note that the team decided to follow the ADaM 1.1 IG though it is not yet supported by a corresponding Pinnacle 21 version.
* Based on the reported findings each dataset was updated so that no more unexplained findings (see section 3) are reported.
* Pinnacle 21 Community v2.2 was again used on all updated datasets to find additional issues resulting from inconsistencies between the datasets. The datasets were updated again.
* For updating the datasets, the team used SAS and R programs.

The team used Pinnacle 21 as a conformance checking tool because it is a commonly used tool to evaluate conformance and is openly available to everyone. The team is very well aware of shortcomings and knows that it is not perfect and comprehensive but decided that for lack of better alternatives, this would be the right choice. To resolve some of the findings, the team needed to make some decisions on how the datasets should be updated. These changes should not be construed as definitive approaches for resolving conformance issues in general. Users of the updated CDISCPILOT datasets should be aware that depending on the situation, several options can and need to be considered.

Note that the rather large ADLBC dataset was split (see section 2.1) to accommodate limitations of the repository that was used during the work. For consistency, the team decided to split the ADLBH as well using the same rules. This was not done because of any guidance or other requirement to split these domains. Three datasets were renamed to follow CDISC guidance (see section 2.2).

Section 3 of this document provides explanations for the Pinnacle 21 findings that are still seen in the report.

Note that this document is not intended to represent a full-fledged Analysis Data Reviewer’s Guide (ADRG). Instead, its purpose is to serve as documentation for the updated test data, and to explain some of the decisions that were made during the update process. In turn, the test data is intended to be used for the development and testing of standard reporting and analysis scripts and is not meant to represent a complete regulatory submission package. The original CDISCPILOT02 data may be downloaded from <https://www.cdisc.org/sdtmadam-pilot-project>, if desired.

# Specific Comments

## Split datasets

The ADLBC and ADLBH datasets were split based on the values of the indicated variable. Note that this splitting was done to reduce the size of the resulting datasets and to demonstrate split datasets and not because of any guidance or other requirement to split these domains.

|  |  |  |
| --- | --- | --- |
| Split Dataset | Value of split Variable | Comment about Split Variable |
| adlbc.xpt | Preceding “\_” in PARAMCD | The dataset contains two sets of observations, one is marked as “change from previous visit, relative to normal range” and is marked in PARAMCD using a preceding “\_”.  Using the PARAMCD value the dataset was split into adlbc.xpt and adlbcpv.xpt. |
| adlbh.xpt | Preceding “\_” in PARAMCD | The dataset contains two sets of observations, one is marked as “change from previous visit, relative to normal range” and is marked in PARAMCD using a preceding “\_”.  Using the PARAMCD value the dataset was split into adlbh.xpt and adlbhpv.xpt. |

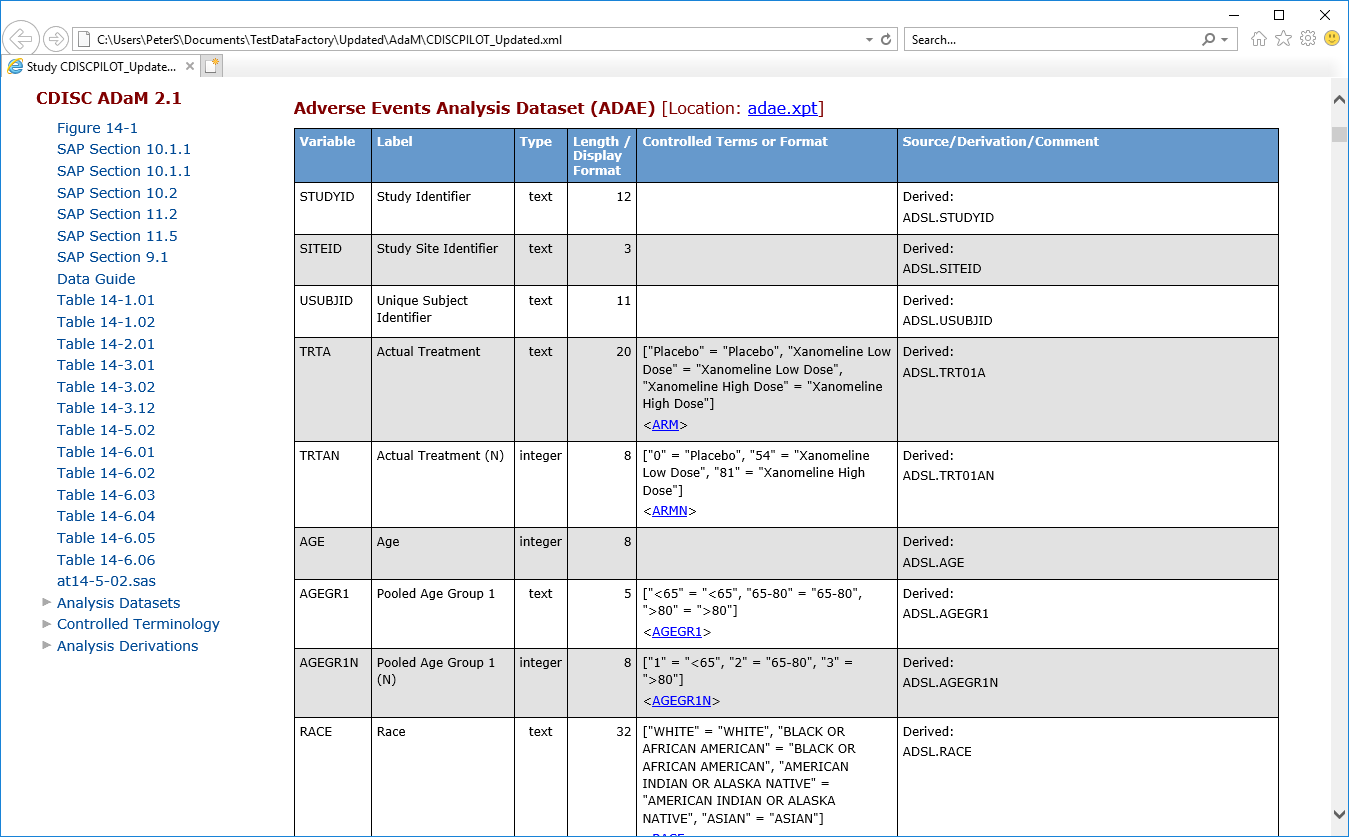
## Renamed Datasets

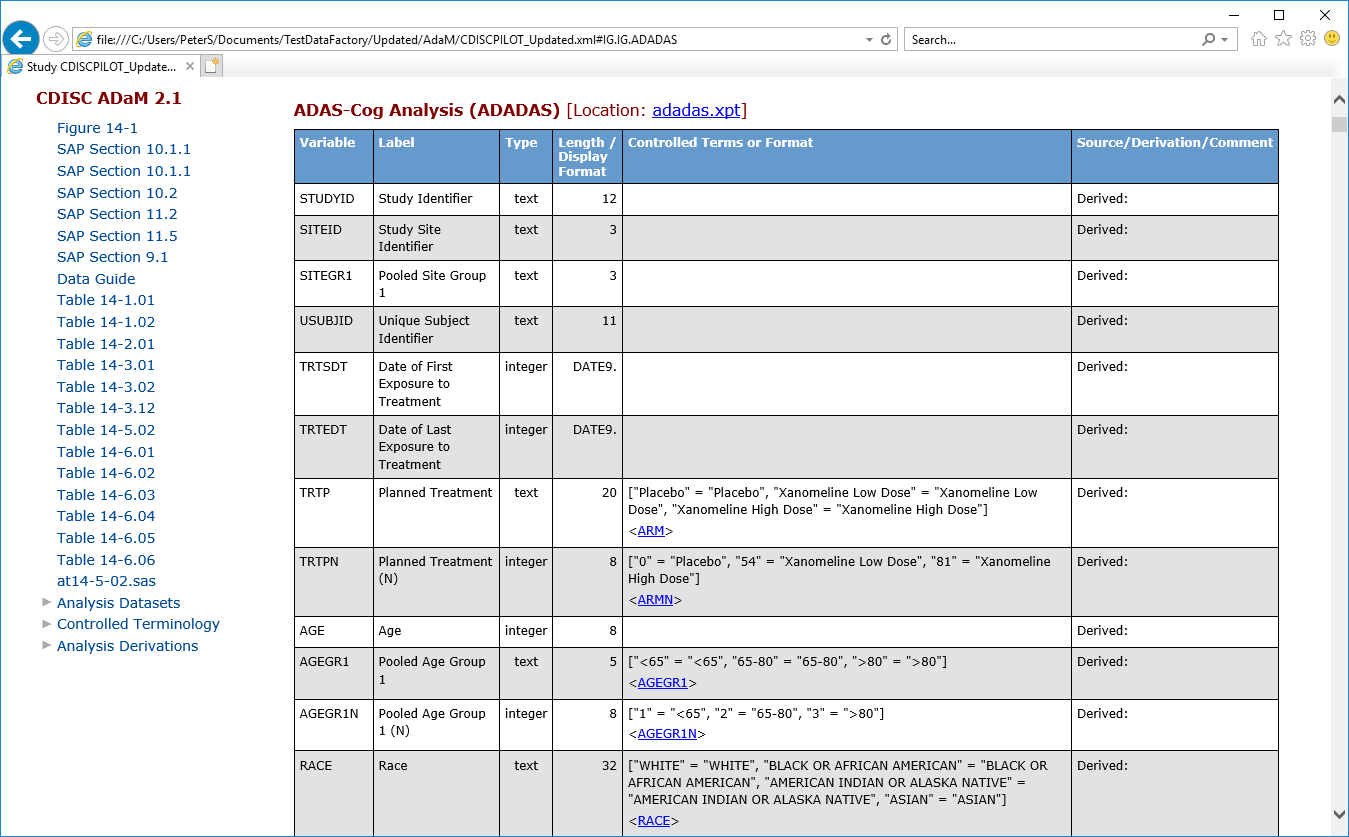
The following table shows a list of datasets that were renamed to confirm to CDISC guidance.

|  |  |
| --- | --- |
| Original Dataset Name | New Dataset Name |
| adqscibc.xpt | adcibc.xpt |
| adqsadas.xpt | adadas.xpt |
| adqsnpix.xpt | adnpix.xpt |

## Missing Information in define.xml

An updated define.xml document is included with the updated ADaM datasets. This define.xml document has been created using the ‘Generate Define’ feature of the Pinnacle 21 Community edition from an Excel specification document. Note that there is incomplete information in the define.xml file. More specifically, Source/Derivation/Comment information is not included for all datasets and variables. The TDF team believes that the information included in the define.xml is sufficient to show how this information would be provided.

  
Screenshot of ADAE.xpt details with Source/Derivation/Comment information included

  
Screenshot of ADADAS.xpt details where Source/Derivation/Comment information is missing

# Data Conformance Summary – Explanation on Remaining Pinnacle 21 Findings

The following table explains the remaining warnings from a Pinnacle 21 Community Edition validation. This list represents the status after the update of the define.xml file and is included so that users of the updated CDISC Pilot dataset can understand the extent of changes that were applied to the datasets. A final validation report from Pinnacle 21 Community Edition is included with the datasets or can easily be created by running the tool against the datasets.

| Dataset | Variables | Values | Message | Severity | Explanation |
| --- | --- | --- | --- | --- | --- |
| ADADAS | AWU, AWLO, AWHI | DAYS, null, 1 or DAYS, 141, null | AWU is populated but both AWLO and AWHI are not populated | Error | A one-sided window is incorrectly marked by the current Pinnacle 21 version as an error. |
| ADLBC | VARIABLE, LABEL | ANL01FL, Analysis Record Flag 1 | Variable label mismatch between dataset and ADaM standard | Error | Pinnacle 21 Community Edition validation uses ADaMIGv1.0. The value for the variable label conforms to ADaMIGv1.1 as required. |
| ADLBCPV | VARIABLE, LABEL | ANL01FL, Analysis Record Flag 1 | Variable label mismatch between dataset and ADaM standard | Error | Pinnacle 21 Community Edition validation uses ADaMIGv1.0. The value for the variable label conforms to ADaMIGv1.1 as required. |
| ADLBH | VARIABLE, LABEL | ANL01FL, Analysis Record Flag 1 | Variable label mismatch between dataset and ADaM standard | Error | Pinnacle 21 Community Edition validation uses ADaMIGv1.0. The value for the variable label conforms to ADaMIGv1.1 as required. |
| ADLBHPV | VARIABLE, LABEL | ANL01FL, Analysis Record Flag 1 | Variable label mismatch between dataset and ADaM standard | Error | Pinnacle 21 Community Edition validation uses ADaMIGv1.0. The value for the variable label conforms to ADaMIGv1.1 as required. |
| ADNPIX | AWU, AWLO, AWHI | DAYS, null, 1 or DAYS, 176, null | AWU is populated but both AWLO and AWHI are not populated | Error | A one-sided window is incorrectly marked by the current Pinnacle 21 version as an error. |
| ADVS | VARIABLE, LABEL | ANL01FL, Analysis Flag 01 | Variable label mismatch between dataset and ADaM standard | Error | Pinnacle 21 Community Edition validation uses ADaMIGv1.0. The value for the variable label conforms to ADaMIGv1.1 as required. |