Janssen SDTM Validation Process Manual

1. **INTRODUCTION**

The purpose of this document is to act as a guide for how to use the validation macros to validate the SDTM deliverables as well as review the validation outputs and implement the changes to aCRF, SDTM SPEC and SDTM program.

All the process mentioned in this guide is not mandatory per PAREXEL SOP, but this is required process in SDTM tasks in Janssen account. If you feel that this document is in conflict with an SOP or have any questions, please talk to your manager.

1. **SDTM validation flow chart**

All validation macros can be executed **AFTER** the statistical programming QC steps are done per SOP-WW-GDO-003, and it should be done **BEFORE** delivering SDTM package to Janssen.

Passed Statistical Programming QC?

Modify SDTM datasets QC program

Modify SDTM datasets Main program

No

Validation:

jjchkmetastd.sas

Jjchkmetadata.sas

jjchksdtmvalid.sas

jjchkother.sas

jjchkacrf.sas

Yes

1. **Introduction to each macro and output**
   1. jjchkmetastd.sas
      1. Purpose:

To compare study level metadata (datadef, valdef, cd, compmeth and vardef) with standard metadata

* + 1. Invoke:

Invoke the macro with no parameters for new study, library names meta (directory for study level metadata) and metastd (directory for standard metadata) from setup.sas should exist. The code is already set in the macro, no need to set any other parameters. Just put this macro into study folder and execute the macro, then get the validation output

* + 1. Report:

There are five compare result in txt format and one xml output. Those are generated correspond to each metadata dataset, which show the difference between study level metadata and standard metadata. Please note that some differences are acceptable.

* qc\_CD.txt - This file show the difference of code list (Controlled Terminology) between study level metadata and standard metadata
* qc\_COMPMETH.txt - This file show the difference of computational method (Computational Algorithms) between study level metadata and standard metadata.
* qc\_DATADEF.txt - This file show the difference of data definition (Dataset-Level Metadata) between study level metadata and standard metadata
* qc\_VALDEF.txt - This file show the difference of value definition (Value-Level Metadata) between study level metadata and standard metadata
* qc\_VARDEF.txt - This file show the difference of variable definition (Variable-Level Metadata) between study level metadata and standard metadata
* Metadata.xml - This file contains items included in study level metadata but not included in standard metadata
  + 1. Action:
* Spec should be reviewed and updated as needed
* Run metadata program to generate the updated study level metadata
* Run the macro again to see if all issues were resolved
  1. jjchkmetadata.sas
     1. Purpose:

To cross check study level metadata, check list:

* To check if the order of dataset in DATADEF is sorted by CLASSNM, DATASET
* To check if there is no duplicate DECOD(for JJ standard)
* To check if there is no duplicate CODEVAL(for JANSSEN standard)
* To check if variable REFERENC is not missing(for JJ standard)
* To check if variables CODEVAL and DECOD are not missing except MedDRA and WHODRUG
* To check if variables DICTNRY and VERSION of MedDRA and WHODRUG are not missing
* To check if variables CODEVAL and DECOD of MedDRA and WHODRUG are missing
* To check if the VALVAL in VALDEF is not missing
* To check if the ORIGIN is not missing
* To check if the CRFPAGE is not missing when ORIGIN contains CRF
* To check if the CRF page format is correct(the delimiter should be ", ")
* To check if the value list of IECAT (JJ standard), CCCAT, DSCAT, FTCAT, QSCAT, LBCAT, LBSPEC or LBMTTHOD is correct
* To check if the DECDIG of float point variable is not missing
* To check if the DECDIG of non-float point variable is missing
* To check if the variable order is correct
  + 1. Invoke:

Example: % jjchkmetadata( mlib = meta

, stdv = JANSSEN

, outdir = \_tglobal

, output = MetadataCheck

);

Explanation of each parameter in the macro

* MLIB: Library name of study metadata datasets (default meta)
* STDV: Flag for sponsor metadata version (JJ for JJ standard and JANSSEN for JANSSEN standard, default JANSSEN)
* OUTDIR: Full path specifying location of the output file (default \_tglobal)
* OUTPUT: File name of the output (default MetadataCheck)
  + 1. Report:

An xml file, named as “PXLTimeCode\_MetadataCheck\_yyyymmdd.xml”, generated with 5 sheet included. Details information of each sheet is:

* DatasetChk
  1. The order of dataset is not sorted by CLASSNM, DATASET
* CodelistChk:

1. There are duplicate CODEVAL attached to the code list XXX
2. There are duplicate DECOD attached to the code list XXX
3. Variable REFERENC in CD is missing (for JJ standard)
4. Variable CODEVAL and DECOD in code list is missing except MedDRA and WHODRUG
5. Variable DICTNRY and VERSION of MedDRA and WHODRUG in code list is missing
6. Variables CODEVAL and DECOD of MedDRA and WHODRUG in code list is not missing

* ValuelistChk

1. VALVAL in VALDEF is missing
2. The ORIGIN in is missing
3. ORIGIN format is not correct for multiple origin values. If there are multiple ORIGINS then the CRF should be in the end. For example: “Assigned, CRF” not “CRF, Assigned”.
4. CRFPAGE is missing when ORIGIN contains CRF
5. CRF page format is not correct (the delimiter should be ", ")
6. The value list of IECAT (JJ standard), CCCAT, DSCAT, FTCAT, QSCAT, LBCAT, LBSPEC or LBMTTHOD is not correct
7. The DECDIG of float-point variable is missing
8. The DECDIG of non-float-point variable is not missing

* VardefChk

1. The ORIGIN is missing
2. ORIGIN format is not correct for multiple origin values. If there are multiple ORIGINS then the CRF should be in the end. For example: “Assigned, CRF” not “CRF, Assigned”.
3. CRFPAGE is missing when ORIGIN contains CRF
4. CRF page format is not correct (the delimiter should be ", ")
5. The DECDIG of float-point variable is missing
6. The DECDIG of non-float-point variable is not missing
7. The variable order is not correct

* CrossChk

1. Code list/Value list/ Computational Algorithm Method in study metadata VARDEF or VALDEF but not in study metadata CD/VALDEF/COMPMETH
2. Action: add Code list/Value list/ Computational Algorithm Method to sheet CD/VALDEF/COMPMETH
3. Code list/Value list/ Computational Algorithm Method in study metadata CD/VALDEF/COMPMETH but not in study metadata VARDEF or VALDEF
4. Action: remove Code list/Value list/ Computational Algorithm Method from sheet CD/VALDEF/COMPMETH
5. ORIGIN of –ORRES/QVAL in study metadata VARDEF do not correspond with ORIGIN of –TESTCD/QNAM in study metadata VALDEF
6. Action: revise ORIGIN of –ORRES/QVAL in sheet VARDEF or ORIGIN of –TESTCD/QNAM in sheet VALDEF
7. There is mismatch between CRF page in study metadata VARDEF and CRF page in study metadata VALDEF
8. Action: revise CRF page in sheet VARDEF or VALDEF
9. The logical key order in study metadata VARDEF is not consistent with the keys in metadata DATADEF (for JJ standard)
10. Action: revise logical key order in sheet VARDEF
    1. jjchksdtmvalid.sas
       1. Purpose:

To validate SDTM datasets, check list:

|  |  |  |
| --- | --- | --- |
| CheckID | Check Description | Message |
| Check030010 | Check for variable or variable attribute in SDTM datasets is consistent with the attributes of variables in metadata VARDEF | 1. Variable XXX in SDTM dataset DOMAIN but not is metadata VARDEF or vice versa  2. The attributes of variables in SDTM datasets is not consistent with the attributes of  variables in VARDEF, the observation number is xxx |
| Check030010 | Check for each codelist related variable, that the value is found in the study-specific codelist attached to that variable | The value cannot be found in the codelist attached to the variable |
| Check030026 | Check for each value level metadata related variable, that the value is found in the value level metadata attached to that variable | The value cannot be found in the value level metadata attached to the variable |
| Check030270 | Check if --DY, --STDY or --ENDY is present in the dataset that a COMPMETHOD is provided in the comments | A STUDY DAY OF COLLECTION is present but there is no link to a COMPUTATIONAL ALGORITHM. |
| Check030714 | Check for each domain that a comment is attached to the variable --STTPT | Informative Check: A comment that describes the start of the protocol-specified reference period (--STTPT) is missing in define.xml. |
| Check030715 | Check for each domain that a comment is attached to the variable --ENTPT | Informative Check: A comment that describes the end of the protocol-specified reference period (--ENTPT) is missing in define.xml. |

* + 1. Invoke:

Example: % jjchksdtmvalid (slib = transfer

, mlib = meta

, stdv = JANSSEN

, outdir = \_tglobal

, output = SdtmValid

);

Explanation of each parameter in the macro

* SLIB: Library name of SDTM datasets (default transfer)
* MLIB: Library name of study metadata datasets(default meta)
* STDV: Flag for sponsor metadata version (JJ for JJ standard and JANSSEN for JANSSEN standard, default JANSSEN)
* OUTDIR: Full path specifying location of the output file (default \_tglobal)
* OUTPUT: File name of the output (default SdtmValid)
  + 1. Report:

An xml file, named as “PXLTimeCode\_ SdtmValid \_yyyymmdd.xml”, generated with 4 sheet included. Details information of each sheet is:

* Codelstchk (Check030010):

1. For codelist related variable, that the value cannot be found in the study-specific codelist attached to that variable
2. Action: Sheet CD in Spec should be updated or codelist related variable should be coded. E.g.: CODELST attached to variable LBSTAT is ND, so value ‘ NOT DONE’ should be assigned to this variable instead of ‘Not Done’

* Valuelstchk (Check030026):

1. For value level metadata related variable, that the value cannot be found in the value level metadata attached to that variable
2. Action: Sheet VALDEF in Spec should be updated or the value related variables are not assigned correctly

* Valuelstchk2 (Check030026\_2):

1. For each value level metadata related variable, that the value cannot be found in the study-specific codelist attached to that variable
2. Action: Sheet CD in Spec should be updated or codelist related to value of the variable should be coded. E.g.: codelst related to value ‘OCCUR’ of variable FATESTCD is NY, so the value of FAORRES for this TESTCD must be one of ‘N’, ‘NA’, ‘U’ and ‘Y’

* Comblstchk:

1. Variable in SDTM dataset but not is study metadata VARDEF or vice versa (Check030000)
3. A STUDY DAY OF COLLECTION is present but there is no link to a COMPUTATIONAL ALGORITHM (Check030270)
4. A comment that describes the start of the protocol-specified reference period --STTPT is missing in metadata (Check030714)
5. A comment that describes the end of the protocol-specified reference period --ENTPT is missing in metadata (Check030715)
6. Action: Spec or program should be updated until all issue are resolved.
   1. jjchkother.sas
      1. Purpose:

To validate SDTM datasets

* + 1. Invoke:

Example: % jjchkother (slib = transfer

, mlib = meta

, outdir = \_tglobal

, output = OtherChk

);

Explanation of each parameter in the macro

* SLIB: Library name of SDTM datasets (default transfer)
* MLIB: Library name of study metadata datasets (default meta)
* OUTDIR: Full path specifying location of the output file (default \_tglobal)
* OUTPUT: File name of the output (default SdtmValid)
  + 1. Report:

An xml file, named as “PXLTimeCode\_ OtherChk \_yyyymmdd.xml”, generated with 1 sheet included. Details information of each sheet is:

* Comblstchk (checkid included in below file).



* + 1. Action:

Spec or program should be updated

* 1. jjchkacrf.sas
     1. Purpose:

Check list:

* To check if the variable in aCRF is consistent with the variable in standard metadata VARDEF
* To check if the variable in study metadata dataset is consistent with the variable in aCRF
* To check if the variable in SDTM dataset is consistent with the variable in aCRF
* To check if the value in study metadata VALDEF is consistent with the value in aCRF
* To check if the CRF page in study metadata VALDEF is consistent with the page in aCRF
* To check if the CRF page in study metadata VARDEF is consistent with the page in aCRF
  + 1. Invoke:

To use this macro, we need to extract the information from aCRF so to cross check with the define.xml and SDTM datasets.

Prerequisite: creating the annotation summary file, Adobe Acrobat process:

1. Select ‘Comments 🡪 Summarize Comments’ from the main menu
2. A new box with the title “Summarize Options” appears. Choose layout ‘Comments only’
3. Press ‘OK’. Acrobat will create a new PDF with annotation summaries
4. Save this resulting PDF as a text file named comments at your location. Select ‘File 🡪 Save As’ from the main menu, and select save as type to be ‘Text (Plain) (\*.txt)’
5. Copy text file ‘comments.txt’ to kennet folder used to save DTMS

Example: % jjchkacrf (slib = transfer

, mlib = meta

, smlib = metastd

, outdir = \_tglobal

, output = CRFPageCheck

);

Explanation of each parameter in the macro

* SLIB: Library name of SDTM datasets (default transfer)
* MLIB: Library name of study metadata datasets (default meta)
* SMLIB: Library name of standard metadata datasets (default metastd)
* OUTDIR: Full path specifying location of the output file (default \_tglobal)
* OUTPUT: File name of the output (default SdtmValid)
  + 1. Report:

An xml file, named as “PXLTimeCode\_ CRFPageCheck \_yyyymmdd.xml”, generated with 6 sheet included. Details information of each sheet is:

* VarChk1:

1. Variable in aCRF but not in standard metadata VARDEF
2. Action: check spelling of variable name in aCRF

* VarChk2:

1. Variable in study metadata dataset which origin contains 'CRF' but not in aCRF
2. Variable in aCRF but the origin does not contain 'CRF' in study metadata dataset
3. Variable in aCRF but not in study metadata dataset
4. Action: origin of variable should be reviewed and updated as needed in corresponding domain; variable should be added to corresponding domain

* VarChk3:

1. Variable in SDTM dataset which origin contains 'CRF' but not in aCRF
2. Variable in aCRF but the origin does not contain 'CRF' in SDTM dataset
3. Variable in aCRF but not in SDTM dataset
4. Action: origin of variable should be reviewed and updated as needed in corresponding domain; variable should be added to corresponding domain

* ValChk:

1. Value in study metadata VALDEF which origin contains 'CRF' but not in aCRF
2. Value in aCRF but the origin does not contain 'CRF' in study metadata VALDEF
3. Value in aCRF but not in study metadata VALDEF
4. Action: origin of variable should be reviewed and updated as needed in sheet VALDEF; variable should be added to sheet VALDEF

* VarCrfPageChk:

1. The CRFPAGE attached to the variable in study metadata VARDEF is not correct
2. Action: revise the CRFPAGE in corresponding domain

* ValCrfPageChk:

1. The CRFPAGE attached to the value in study metadata VALDEF is not correct
2. Action: revise the CRFPAGE in sheet VALDEF