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.

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. For regular transfer, if there is no change on standard metadata and study level metadata, then we don’t need check this result file.

Note, the report needs to be re-checked when sponsor release the new standard metadata.

* + 1. Invoke:

Invoke the macro with no parameters for new study, library names meta (directory for study level metadata), metastd (directory for standard metadata) and global macro variables &\_tglobal (directory for TXT report) 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
* PXLTimeCode\_MetadataDiff.xml - This file:

1. The items included in study level metadata but not included in standard metadata (see Datadef, Valdef, CD, Compmeth, Vardef tabs).
2. The ‘Exp’ or ‘Req’ variables in standard metadata are not included in study level metadata (see Vardef\_Reqexp tab).
   * 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 comment or computational method of derived variable/VALVAL is not missing
* To check if the comment or computational method of non-derived variable/VALVAL is missing
* To check if the value list of IECAT (JJ standard), CCCAT, DSCAT, FTCAT, QSCAT, LBCAT, LBSPEC or LBMETHOD 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
* To check if --DY, --STDY or --ENDY is present in the vardef that a COMPMETHOD is provided in the comments (Check030270)
* To Check for each domain that a comment is attached to the variable –STTPT (Check030714)
* To Check for each domain that a comment is attached to the variable –ENTPT (Check030715)
* To check if the CODEVAL is sorted numerically for meaningful RNK CODELST
* To Check if the CODEVAL is sorted alphabetically for non-meaningful RNK CODELST
  + 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.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
7. The CODEVAL is not sorted numerically for meaningful RNK CODELST XXX

Action: for CODEVAL in standard metadata, assign sequential RNK starting from 1 per the relative order in standard CODELST. After these are well-sorted, sort the customized CODEVAL in study level not in standard CODELST alphabetically to assign sequential RNK.

1. The CODEVAL is not sorted alphabetically for non-meaningful RNK CODELST XXX

* ValdefChk

1. VALVAL in VALDEF is missing
2. The ORIGIN in VALDEF 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 comment or computational method of derived VALVAL is missing
7. The comment or computational method of non-derived VALVAL is not missing
8. The value list of IECAT (JJ standard), CCCAT, DSCAT, FTCAT, QSCAT, LBCAT, LBSPEC or LBMETHOD is not correct. Please note for nested value list, the VALUELST can only be missing when it’s the lowest level. Like for LB, LBCAT/LBSPEC/LBMETHOD the VALUELST should not be missing, but only the LBTESTCD should have missing VALUELST
9. The DECDIG of float-point variable is missing
10. 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 comment or computational method of derived variable is missing
6. The comment or computational method of non-derived variable is not missing
7. The DECDIG of float-point variable is missing
8. The DECDIG of non-float-point variable is not missing
9. The variable order is not correct
10. COMPUTATIONAL ALGORITHM attached to variable (--DY, --STDY or –ENDY) is missing.
11. A comment that describes the start of the protocol-specified reference period is missing
12. A comment that describes the end of the protocol-specified reference period is missing

* CrossChk

1. Code list/Value list/ Computational Algorithm Method in study metadata VARDEF or VALDEF but not in study metadata CD/VALDEF/COMPMETH

Action: add Code list/Value list/ Computational Algorithm Method to sheet CD/VALDEF/COMPMETH

1. Code list/Value list/ Computational Algorithm Method in study metadata CD/VALDEF/COMPMETH but not in study metadata VARDEF or VALDEF

Action: remove Code list/Value list/ Computational Algorithm Method from sheet CD/VALDEF/COMPMETH

ORIGIN of –ORRES/QVAL in study metadata VARDEF do not correspond with ORIGIN of –TESTCD/QNAM in study metadata VALDEF

Action: revise ORIGIN of –ORRES/QVAL in sheet VARDEF or ORIGIN of –TESTCD/QNAM in sheet VALDEF

1. There is mismatch between CRF page in study metadata VARDEF and CRF page in study metadata VALDEF

Action: revise CRF page in sheet VARDEF or VALDEF

1. The logical key order in study metadata VARDEF is not consistent with the keys in metadata DATADEF (for JJ standard)

Action: revise logical key order in sheet VARDEF

* 1. jjchksdtmvalid.sas
     1. Purpose:

To validate SDTM datasets, check list:

|  |  |  |
| --- | --- | --- |
| CheckID | Check Description | Message |
| Check030000 | Check for variable or variable attribute in SDTM datasets is consistent with the attributes of variables in metadata VARDEF | 1. Variable XX 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 variable list is XX |
| Check030010 | Check for each codelist related variable, that the value is found in the study-specific codelist attached to that variable | The value XX cannot be found in the codelist XX attached to the variable XX |
| 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 XX cannot be found in the value level metadata attached to the variable XX |

* + 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 .xml”, generated with 4 sheet included. Details information of each sheet is:

* Varattchk:

1. Variable in SDTM dataset but not is study metadata VARDEF or vice versa (Check030000)
2. The attributes of variables in SDTM datasets is not consistent with the attributes of variables in study metadata VARDEF (Check030000)

* Codelstchk (Check030010):

1. For codelist related variable, that the value cannot be found in the study-specific codelist attached to that variable

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

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

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’.

Spec or program should be updated until all issues are resolved.

* 1. jjchkacrfvar.sas
     1. Purpose:

Check list:

* 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
  + 1. Invoke:

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

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

1. Select ‘Create Comment Summary’ from the main menu
2. A new box with the title “Summarize Comments” 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)PDF to Text (Simple) (\*.txt)’
5. Copy text file ‘comments.txt’ to kennet folder used to save DTMS

Example: % jjchkacrfvar (slib = transfer

, mlib = meta

, smlib = metastd

, spath=SPECPATH

, outdir = \_tglobal

, output = aCRFVarCheck

);

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)
* SPATH: Library name of folder rawspec (default SPECPATH)
* OUTDIR: Full path specifying location of the output file (default \_tglobal)
* OUTPUT: File name of the output (default aCRFVarCheck )
  + 1. Report:

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

* VarChk1:

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

Action: Origin of variable should be reviewed and updated as needed in corresponding domain.

1. Variable in aCRF but not in study metadata dataset

Action: Variable should be reviewed and added it to corresponding domain in DTMS if confirmed.

* VarChk2:

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

Action: Origin of variable should be reviewed and updated as needed in corresponding domain.

1. Variable in aCRF but not in SDTM dataset

Action: Variable should be reviewed and added it to corresponding SDTM dataset if confirmed.

Note for VarChk1 and VarChk2

1. Only variables in standard metadata will be checked;
2. Per current SDTM process, SDTM dataset read variables from DTMS, these two check reports will be same.

* ValChk:

1. Value (VALVAL) in study metadata VALDEF which origin contains 'CRF' but not in aCRF

Action: Origin of Value (VALVAL) should be reviewed and updated as needed in corresponding domain. Note: no need update if the reported value in VALDEF contains ‘CRF’ but not in aCRF since the value is preprinted text on aCRF, such as when ‘ECOG’ is preprinted on CRF, we generally only annotate QSCAT beside ‘ECOG’, no need annotate QSCAT = ECOG.

* 1. jjchkacrfpage.sas
     1. Purpose:

To check SDTM aCRF annotation's page number with Define XML. For variable level metadata (VARDEF), all variables page numbers will be checked. For value level metadata (VALDEF), there will be two scenarios. 1) For nested domains, only check the 1st level page (--CAT). All other levels need to be checked manually. 2) For non-nested domain, all value lists page numbers will be checked. Note, QS domain is a special case, even QSTESTCD is nested by QSCAT, but both QSCAT and QSTESTCD will be checked since for all QSTESTCD except for QSALL are unique. For QSALL needs to be checked manually as it will depend on the different QSCAT.

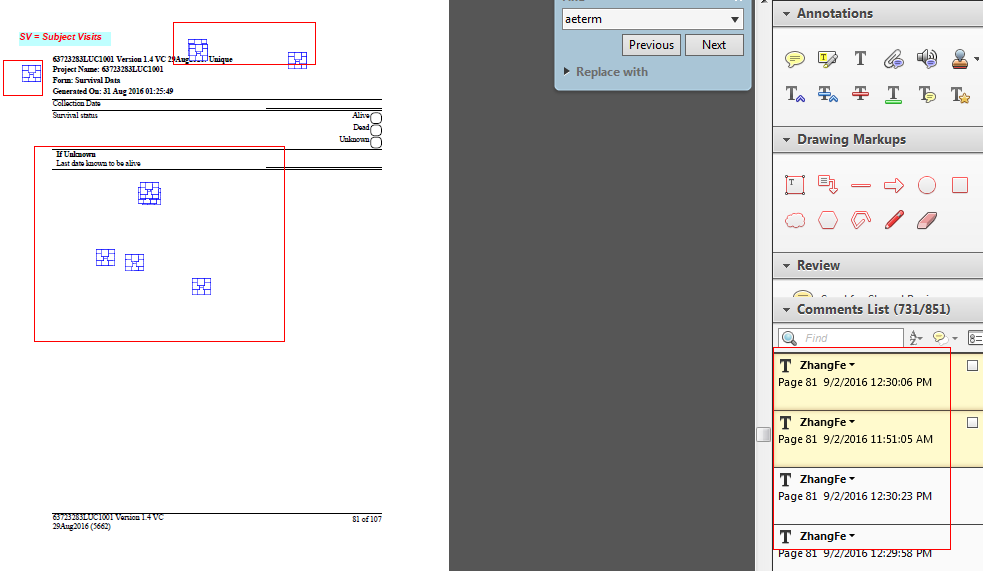
* + 1. Invoke:

To use this macro, we need to prepare 5 documents. 1) define.xml 2) definexml.map 3) acrf.xfdf 4) acrf.map 5) jjchkacrfpage.sas. The definexml.map and acrf.map and jjchkacrfpage.sas will be released together in the validation macro package. You need to follow below process to export the acrf.xfdf.

Open your acrf by Adobe Reader 🡪 Select ‘Comments’ 🡪 Select ‘Export All to Data File’ 🡪 Select the type to ‘Acrobat XFDF Files (\*.xfdf)’ and name it as acrf.xfdf

Note 1: You need to make sure your acrf doesn’t contain any blank text box. All text boxes should contain the annotations, Otherwise the macro will go wrong. To check this, you can follow below steps.

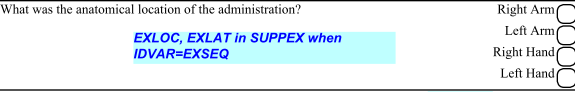
Open your acrf by Adobe Reader 🡪 Select ‘Comments’🡪In the Comments List those null text box should be removed. (See below the screenshot)



Note 2: Follow the Janssen Define Spec v1.0, in case multiple origins are defined, the CRF Page should be listed at the end (e.g. Assigned, CRF Pages 1, 2, 3). Please make sure define.xml follow this scenario, if not, please revise the spec firstly and then generate the updated study metadata and define.xml.

Note 3: When do the annotation, please don’t use Ctrl +Enter to change line. Please use space only and adjust the text box.

Note 4: Don’t put main domain variable and SUPP QNAM in the same text box.



Once you prepared all the 5 documents, uploaded them to the same folder on kennet.

Example: %jjchkacrfpage(root=/user1/shenb/jjcrfpage/226483/

, nest\_domain=CC|DS|FT|LB);

Note: ntest\_domain contain the default value, you can only mentioned the root.

Example: %jjchkacrfpage(root=/user1/shenb/jjcrfpage/226483/);

Explanation of each parameter in the macro

* root: Please put the folder path where you uploaded all the 5 files.
* nest\_domain: The default value is CC|DS|FT|LB based on latest Janssen metadata. For QS, it's also a common nested domain, however except for QSALL, in most of cases the QSTESTCD will be unique so we don't include QS here. In case you want to update this, please list all the nested domains here and separate them by pipe (|).
  + 1. Report:

Two txt files will be generated. 1) “VALDEF Page Potential Issue.txt” and 2) “VARDEF Page Potential Issue.txt”.

1. Action: Please review the reports carefully. In case define.xml is wrong, please update the SDTM mapping spec firstly and then generated the updated study metadata and define.xml and re-run the validation program.