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 with standard metadata

* + 1. Invoke:

Invoke the macro with no parameters for new study

There are five dataset are compared with standard metadata, Datadef, valdef, cd, compmeth and vardef. Are code are 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.

* 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

* 1. jjchkmetadata.sas
     1. Purpose:

To cross check study level metadata

* + 1. Invoke:

Example:

%metadatachk( mlib = meta

, stdv = JANSSEN

, outdir = \_tglobal

, output = MetadataCheck

);

Explanation of each parameter in the macro

* MLIB:
* STDV:
* OUTDIR:
* OUTPUT:
  + 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
2. Variables CODEVAL and DECOD is missing except MedDRA and WHODRUG
3. V variables DICTNRY and VERSION of MedDRA and WHODRUG is missing
4. Variables CODEVAL and DECOD of MedDRA and WHODRUG is missing

* ValuelistChk and VardefChk

1. VALVAL in VALDEF is missing
2. The ORIGIN is missing
3. CRFPAGE is missing when ORIGIN contains CRF
4. CRF page format is not correct (the delimiter should be ", ")
5. The comment or computer method of derived variables is missing
6. The value list of IECAT and QSCAT is not correct, shoulb be catx(".", VALUEOID, VALVAL, 'IETESTCD') and catx(".", VALUEOID, VALVAL, 'QSTESTCD')
7. The DECDIG of float-point variable is missing
8. The DECDIG of non-float-point variable is missing
9. The variable order is correct

* CrossChk

1. There is mismatch between CODELST in VARDEF and VALDEF and CODELST in CD
2. There is mismatch between VALUELST in VARDEF and VALDEF and VALUEOID in VALDEF
3. The logical key order in VARDEF is not consistent with the keys in DATADEF (for JJ standard)
   * 1. Action:

Spec should be updated

* 1. jjchksdtmvalid.sas
     1. Purpose:

To validate SDTM datasets

* + 1. Invoke:

Example:

%sdtmvalid(slib = transfer

, mlib = meta

, stdv = JANSSEN

, outdir = \_tglobal

, output = SdtmValid

);

Explanation of each parameter in the macro

* SLIB:
* MLIB
* STDV:
* OUTDIR:
* OUTPUT:
  + 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.: Adverse Event 🡪 AE

* 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

* 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 variable should be coded. E.g.: Adverse Event 🡪 AE

* Comblstchk (Check030000, Check030267, Check030270, Check030272, Check030714, Check030715):

1. The attributes of variables in SDTM datasets is not consistent with the attributes of variables in metadata VARDEF
2. There is codelist assigned to --ORRES (except for IEORRES)
3. A STUDY DAY OF COLLECTION is present but there is no link to a COMPUTATIONAL ALGORITHM.
4. There is codelist assigned to QVAL
5. A comment that describes the start of the protocol-specified reference period --STTPT is missing in metadata.
6. A comment that describes the end of the protocol-specified reference period --ENTPT is missing in metadata
7. Action: Spec or program should be updated
   1. jjchkother.sas
      1. Purpose:

To validate SDTM datasets

* + 1. Invoke:

Example:

%sdtmvalid(slib = transfer

, mlib = meta

, outdir = \_tglobal

, output = OtherChk

);

Explanation of each parameter in the macro

* SLIB:
* MLIB
* OUTDIR:
* OUTPUT:
  + 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:

To check CRF page number

* + 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:

%crfpagechk(slib = transfer

, mlib = meta

, smlib = metastd

, outdir = \_tglobal

, output = CRFPageCheck

);

* + 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 is not consistent with the variable in standard metadata VARDEF

* VarChk2:

1. Variable in study metadata dataset is not consistent with the variable in aCRF

* VarChk3:

1. Variable in SDTM dataset is not consistent with the variable in aCRF

* ValChk:

1. Variable in VALDEF is not consistent with the value in aCRF

* VarCrfPageChk:

1. CRF page in VALDEF is not consistent with the page in aCRF

* ValCrfPageChk:

1. CRF page in VARDEF is not consistent with the page in aCRF
   * 1. Action:

aCRF or Spec should be updated