Paper DS06

Mastering Pinnacle 21 for Oncology SDTM

Daryna Kholololovych, Intego Group, Kyiv, Ukraine

ABSTRACT

While virtually all programmers are working with CDISC standards, it is also vital to fully understand Pinnacle 21. With a focus on the common therapeutic area of Oncology, and particularly SDTM domains.It is useful to understand the specific details and nuances regarding Oncology such as understanding compliance checks more deeply and helping to stay focused while working with large reports. Also, common issues and challenges will be explained with real world examples. As well as saving time and resources it will add quality to the work produced. Also looking at outstanding scenarios which are preferable to document in the SDRG rather than to program around these issues.

INTRODUCTION

Pinnacle 21 is utilized by both sponsors and the FDA to ensure adherence to FDA business rules and CDISC standards. It serves as an effective diagnostic tool for identifying and reporting non-compliance issues in study data. Findings marked with "Error" or "Warning" messages can be classified as either data problems or SDTM mapping issues. Sponsors must address the data issues or clarify any discrepancies in the Study Data Reviewer's Guide (SDRG), or correct SDTM mapping errors prior to FDA submission. As a result, it does not enable sponsors to automatically resolve data conformance issues, even if used early in the SDTM programming process. Additionally, many of these non-compliant data issues can be quite "costly" and may be too late to fix if discovered at a later stage.

In general, the whole process could be described like this:

1) Raw data collection (eCRF);
2) Data issue reports (Data Management);
3) SDTM data and define.xml (specification and programming), early, long before submission;
4) Controlled Terminology and mismatches;
5) Resizing SDTMs and adapting data for define.xml;
6) Splitting into subfolders etc.

And for SDTMs and ADaMs it goes like this:

SDTM Package:

- SDTM SAS Transport files
- cSDRG.pdf
- define.xml
- aCRF.pdf

ADaM Package:

- ADaM datasets in xpt format
- ADRG.pdf
- define.xml
- aCRF.pdf
- Analysis Result Metadata
- SAS programs in text formats

ONCOLOGY DOMAINS

There are 3 oncology domains: TR, TU and RS.

1) **TU (Tumor Identification): the unique identification of tumors for each patient.**

- A single record is created for each distinct tumor identified.
- Tumors are typically identified only at baseline, unless a new tumor arises during a post-baseline record, necessitating its identification.
- If tumors are split or merged, additional post-baseline records are needed to clearly identify the new resultant tumors.
- The TULNKID variable serves as the key unique identifier, providing a distinct code for each identified tumor.
- According to the Findings domains within the core SDTM safety domains, controlled terminology is applied to the TUTESTCD and TUTEST variables. Examples include: TUMIDENT, TUMERGE, and TUSPLIT.

2) **TR (Tumor Results): This involves quantitative measurements and qualitative assessments of the tumors identified in the TU domain, evaluated during follow-up visits.**

- The TRLNKID variable links back to the TU domain.
- The TRLNKGRP variable connects the records in TR that correspond to a response assessment record in the RS domain. RELREC relationships should support this connection.
- In accordance with the Findings domains within the core SDTM safety domains, controlled terminology is applied to the TRTESTCD and TRTEST variables. Examples include: AREA, DIAMETER, and VOLUME.

3) **RS (Disease Response): This encompasses clinical response evaluations derived from TR data and other SDTM domains.**

- The RSLNKID variable links back to the TU domain.
- The RSLNKGRP variable connects the records in TR that support a response assessment record in the RS domain. RELREC relationships should reinforce this connection.
- In line with the Findings domains within the core SDTM safety domains, controlled terminology is applied to the RSTESTCD and RSTEST variables. Examples include: BESTRESP, OVRLRESP, and TRGRESP.
- Any data contributing to the response assessment should be linked to the RS record through RELREC, even if the supportive data is located outside the SDTM Oncology domains (e.g., lab results in LB).

In the following table you can see an overview of the FDA Validator Rules, which serve as essential guideline for ensuring compliance in various regulatory processes. Each row in the table represents a specific rule, detailing the criteria and requirements that must be met for successful validation. The columns provide additional context, including the rule's unique identifier, a brief description, and examples of applicable scenarios. This structured format facilitates easy reference and understanding of complex regulations. By adhering to these rules, organizations can enhance their operational integrity and maintain compliance with FDA standards. The rules not only serve as a resource for regulatory professionals but also promote transparency and consistency within the validating process. In this paper the rules which will be useful for SDTMs in oncology will be provided, as well as the first part – general rules which are important to check in general for oncology.

FDA Validator Rule ID	FDA Validator Rule Description
FDAB003	Reported Term for the Adverse Event (AETERM) is expected to be coded using the Medical Dictionary for Regulatory Affairs (MedDRA). Sponsors should code each term to the MedDRA-specified hierarchy of terms. These terms/codes are represented in SDTM variables that are subject to MedDRA: AELLT, AELLTCD, AEDECOD, AEPTCD, AEHLT, AEHLTCD, AEHLGT, AEHLGTCD, AEBODSYS, AEBDSYCD, AESOC, AESOCCD.

FDA Validator Rule ID	FDA Validator Rule Description
FDAB004	Adverse Events (AE) dataset should be included in every submission. Lab Test Results (LB) dataset should be included in every submission. Vital Signs (VS) dataset should be included in every submission. Subject Elements (SE) dataset should be included in every submission. Exposure (EX) dataset should be included in every submission. Disposition (DS) dataset should be included in every submission.
FDAB006	All death information should be populated for subjects that died during the study including any post treatment follow-up.
FDAB007	All deaths should be an independent row in the adverse event dataset.
FDAB010	All serious adverse events should be flagged.
FDAB028	Screen Failure subjects should have records in Inclusion/Exclusion dataset.
FDAB055	Trial participants should self-report race and ethnicity and they should not be assigned by the study team.

Next blocks of the rules are essential for all Findings and describing focus on baseline, CT, dates etc. The approach to reviewing your data is:

1) Carefully examine the errors and warning presented;
2) Determine which issues can be addressed and which cannot;
3) If data is coming from various sources – the issues could be poor data qulaity, incorrect or inconsistent mapping, programming mistakes;
4) If the study is in progress – some issues could stay temporarily;
5) Check with data management if issues that are left are related to data and could not be fix in any other way;
6) If issues still are present – they all should be explained in Reviewer's Guide (SDRG).

FDA Validator Rule ID	FDA Validator Rule Message	FDA Validator Rule Description	Type
FDAB026	Multiple BASELINE records for the same test	By its definition clinical baseline flags (e.g., last non-missing value prior to first dose) should have only a single record for the same test for subject. Therefore, each subject (USUBJID) should have no more than one record with Baseline Flag (--BLFL) equals 'Y' for the same assessment which is defined by Test (--TESTCD), Category (--CAT), Subcategory (--SCAT), Method (--METHOD), Specimen (--SPEC), Location (--LOC), Laterality (--LAT) and Directionality (--DIR).	Baseline
FDAB030	Inconsistent value for Standard Units	Standard Units (--STRESU) must be consistent for all records with the same Short Name of Measurement, Test or Examination (--TESTCD), Object of the Observation (--OBJ), Measurement,Test, or Examination Detail (--TSTDTL), Category (--CAT), Subcategory (--SCAT), Specimen Type (--SPEC) and Method of Test or Examination (--METHOD).	CT
FDAB016	Missing --DY variable, when --DTC variable is present	Collection Study Day (--DY) variable should be included into dataset, when Collection Study Date/Time (--DTC) variable is present.	Dates
FDAB016	--DY variable value is not populated	Collection Study Day (--DY) variable value should be populated, when Collection Study Date/Time (--DTC) and Subject Reference Start Date/Time (RFSTDTC)	Dates

FDA Validator Rule ID	FDA Validator Rule Message	FDA Validator Rule Description	Type
		variables values are provided, and both of them include complete date part.	
FDAB032	Missing values for both --DTC and --DY variables	Date/Time of Observation (--DTC) or Study Day of Observation (--DY) should be populated in the Findings observation class where data are usually collected at multiple study days.	Dates
FDAB034	--DTC is after --ENDTC	Date/Time of Specimen Collection (--DTC) must be less or equal to End Date/Time of Specimen Collection (--ENDTC).	Dates
FDAB034	Missing --STDY variable, when --STDTC variable is present	Study Day of Start (--STDY) variable should be included into dataset, when Start Study Date/Time (--STDTC) variable is present.	Dates
FDAB034	Missing --ENDY variable, when --ENDTC variable is present	Study Day of End (--ENDY) variable should be included into dataset, when End Study Date/Time (--ENDTC) variable is present.	Dates
FDAB034	--STDTC date is after RFPENDTC	Start Date/Time (--STDTC) variable value must be less than or equal to Date/Time of End of Participation (RFPENDTC).	Dates
FDAB034	--DTC date is after RFPENDTC	Date/Time of Collection (--DTC) variable value must be less than or equal to Date/Time of End of Participation (RFPENDTC).	Dates
FDAB034	--ENDTC date is after RFPENDTC	End Date/Time (--ENDTC) variable value must be less than or equal to Date/Time of End of Participation (RFPENDTC).	Dates
FDAB034	--DY is after --ENDY	Study Day of Observation (--DY) must be less or equal to Study Day of End of Observation (--ENDY)	Dates
FDAB038	Missing --DTC variable, when --DY variable is present	Collection Study Date/Time (--DTC) variable should be included into dataset, when Collection Study Day (--DY) variable is present.	Dates
FDAB038	Missing --STDTC variable, when --STDY variable is present	Start Study Date/Time (--STDTC) variable should be included into dataset, when Study Day of Start (--STDY) variable is present.	Dates
FDAB038	Missing --ENDTC variable, when --ENDY variable is present	End Study Date/Time (--ENDTC) variable should be included into dataset, when Study Day of End (--ENDY) variable is present.	Dates
FDAB040	Negative value for --DUR	Non-missing Duration of Event, Exposure or Observation (--DUR) value must be greater than or equal to 0.	Dates

FDA Validator Rule ID	FDA Validator Rule Message	FDA Validator Rule Description	Type
FDAB043	Missing value for --DTC, when --ENDTC is provided	Date/Time of Collection (--DTC) should not be NULL, when End Date/Time of Observation (--ENDTC) is not NULL.	Dates
FDAB022	Missing EPOCH value, when a start or observation date is provided	The EPOCH variable is not populated, however either a start date or a collection date is provided. The presence of a start date or collection date for a record should allow for the EPOCH variable to be populated. Variables requested by the relevant regulatory agency in policy documents (e.g., EPOCH) should be included in the dataset and populated.	Epoch
FDAB026	Multiple --LOBXFL records for the same test	By its definition Last Observation Before Exposure Flag (--LOBXFL) should have only a single record for the same test for subject. Therefore, each subject (USUBJID) should have no more than one record with Last Observation Before Exposure Flag (--LOBXFL) equals 'Y' for the same assessment which is defined by Test (--TESTCD), Category (--CAT), Subcategory (--SCAT), Position of Subject During Observation (--POS), Method (--METHOD), Specimen (--SPEC), Location (--LOC) and Laterality (--LAT).	Flags
FDAB018	Inconsistent variable length within split datasets	Split datasets should have matching variable lengths for future merges. Datasets should be resized to the maximum length used prior to splitting.	Meta
FDAB009	Inconsistent value for --TEST within --TESTCD	All values of Name of Measurement, Test or Examination (--TEST) should be the same for a given value of Short Name of Measurement, Test or Examination (--TESTCD).	Pair
FDAB009	Inconsistent value for --TESTCD within --TEST	All values of Short Name of Measurement, Test or Examination (--TESTCD) should be the same for a given value of Name of Measurement, Test or Examination (--TEST).	Pair
FDAB026	Missing --STRESC value for Baseline record	A Standard Result value (--STRESC) is expected to be populated for Baseline records (--BLFL='Y').	Results
FDAB031	Missing value for --STRESN	Standardized Result in Numeric Format (--STRESN) variable value should be populated, when Standardized Result in Character Format (--STRESC) variable value represents a numeric value.	Results
FDAB031	--STRESN does not equal --STRESC	Standardized Result in Numeric Format (--STRESN) variable value should be equal Standardized Result in Character Format (--STRESC) variable value, when Standardized Result in Character Format (--STRESC) variable value represents a numeric value.	Results
FDAB031	--STRESN is populated when --STRESC value is not numeric	Standardized Result in Numeric Format (--STRESN) variable value should be null when Standardized Result in Character Format (--STRESC) variable value does not represent a numeric value, (e.g. 'BLQ', value contains '<' or '>', etc.).	Results
FDAB039	Value for --STNRHI is less than value for --STNRLO	Reference Range Upper Limit-Std Units (--STNRHI) value must be greater than or equal to Reference Range Lower Limit-Std Units (--STNRLO) value.	Results

FDA Validator Rule ID	FDA Validator Rule Message	FDA Validator Rule Description	Type
FDAB042	--ORRES value is populated, when --STAT is 'NOT DONE'	Value of Original Result (--ORRES) variable is expected to be missing, when observation was not performed (Status (--STAT) is 'NOT DONE').	Results
FDAB042	Missing value for --REASND, when --STAT is 'NOT DONE'	Value of Reason Not Done (--REASND) variable should be populated, when Status (--STAT) variable value is 'NOT DONE'.	Status
FDAB042	Value for --STRESC is populated, when --STAT is 'NOT DONE'	Status (--STAT) should be NULL, when Standardized Result in Character Format (--STRESC) is provided.	Status
FDAB044	Missing value for --STTPT, when --STRTPT is populated	Start Reference Time Point (--STTPT) must not be NULL, when Start Relative to Reference Time Point (--STRTPT) is provided.	Time point
FDAB044	Missing value for --ENTPT, when --ENRTPT is populated	End Reference Time Point (--ENTPT) must not be NULL, when End Relative to Reference Time Point (--ENRTPT) is provided.	Time point
FDAB047	Inconsistent value for --ELTM within --TPT, --TPTREF, --CAT and/or --SCAT	All values of Planned Elapsed Time (--ELTM) variable should be the same for a given value of Planned Time Point Name (--TPT) variable, Time Point Reference (--TPTREF) and if present, Category and/or Subcategory (--CAT/--SCAT).	Time point
FDAB047	Inconsistent value for --TPT within --TPTNUM, --DY, --CAT and/or --SCAT	All values of Planned Time Point Name (--TPT) variable should be the same for a given value of Planned Time Point Number (--TPTNUM) variable. Uniqueness for these time points is determined by a combination of the domain, study day, and, if present, category and/or subcategory (--CAT/--SCAT).	Time point
FDAB047	Inconsistent value for --TPTNUM within --TPT, --DY, --CAT and/or --SCAT	All values of Planned Time Point Number (--TPTNUM) variable should be the same for a given value of Planned Time Point Name (--TPT) variable. Uniqueness for these time points is determined by a combination of the domain, Study Day (--DY), and, if present, category and/or subcategory (--CAT/--SCAT).	Time point

CONCLUSION

This paper has highlighted the main issues on which you could concentrate and check your data, starting from eCRF, defining type of errors and warnings, checking if they could be fixed with data management and if yes – addressing them and making your data clean, pointed out the main problems for variables that should be looked into, and if no rules could be applied and the issue still exist – everything what is left should be explained in details in SDRG, so the whole process is transparent and clear.

REFERENCES

https://www.cdisc.org/standards/foundational/sdtmig/sdtmig-v3-4
https://www.lexjansen.com/pharmasug/2018/DS/PharmaSUG-2018-DS06.pdf
https://www.lexjansen.com/phuse/2018/ds/DS03.pdf

CONTACT INFORMATION

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

Daryna Khololovych
Intego Group
35B Borychiv Tik Street
Kyiv 04070, Ukraine
Work Phone: +38 044 500 7020 ext. 2464
Email: darina.hololovich@gmail.com
Website: www.intego-group.com