PharmaSUG China 2024 – Paper DS-10031

CDASH-SDTM Combo Model in SDTM Automation

Chunpeng Zhao, BeiGene

ABSTRACT

SDTM mapping is typically conducted during CRF design following CDASH standards. This paper proposes the CDASH-SDTM Combo Model, which introduces a machine-readable data model for SDTM mapping rules integrated into CRF design specifications. By embedding machine-readable SDTM mapping rules, the transformation cost of CRF data to SDTM format can be minimized without incurring significant additional expenses during CRF design. Furthermore, incorporating machine-readable SDTM specifications during CRF design allows for CDISC compliance checks to be performed during CRF review. This approach enhances CDISC compliance and further reduces SDTM transformation costs.

INTRODUCTION

In traditional workflows, CRF design and SDTM transformation are treated as separate steps. When CDASH standards are followed, SDTM mapping rules are already established during CRF design, making it redundant to recreate SDTM mappings during the SDTM transformation step. Since SDTM transformation is not data-driven, SDTM specifications and computer programs can be prepared before the First Patient In (Chunpeng Zhao et al., 2022). By integrating SDTM specifications into CRF design specifications, the SDTM transformation rules can be incorporated into the CRF design phase. This means that once the CRF is finalized, SDTM specifications and SAS programs can be completed concurrently, eliminating the need to repeat SDTM mapping work.

TRADITIONAL PROCESS: SDTM MAPPING FROM CRF TO SDTM

In the traditional CRF to SDTM mapping process, there are two separate phases: the CRF design phase and the SDTM mapping phase. During the CRF design phase, biomedical concepts are mapped to CRF forms, field labels, and field options, with minimal consideration for SDTM mappings. In the SDTM mapping phase, CRF forms, field labels, and field options are mapped to SDTM domains, variables, and variable values.

Typically, two different teams handle these phases. For example, the data management team works on CRF design, and afterward, the statistical programming team handles the SDTM mapping. As a result, CRF designs and wording often do not fully comply with CDISC standards, leading to numerous wording changes during the SDTM mapping phase. This lack of consistency in terminologies across CRF, SDTM, and TFLs increases communication costs and the risk of misunderstandings.

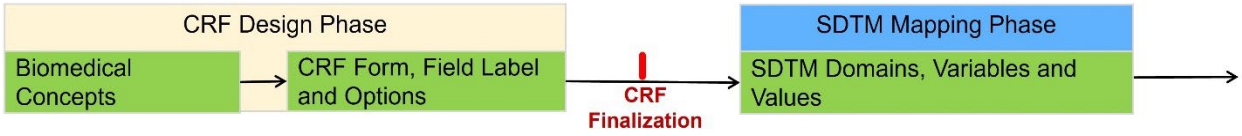


Figure 1. Traditional Process: CRF Design Phase and SDTM Mapping Phase

CDASH PROCESS: SDTM MAPPING FROM BIOMEDICAL CONCEPTS TO SDTM

In the CDASH process, the CRF design phase and the SDTM mapping phase are integrated into a single phase. Initially, biomedical concepts are mapped to SDTM domains, variables, and variable values. Subsequently, these SDTM domains, variables, and variable values are mapped to CRF forms, field labels, and field options. This approach ensures that the CRF design and terminology fully comply with CDISC standards, and that the CRF, SDTM, and TFLs are consistent in their use of terminology.

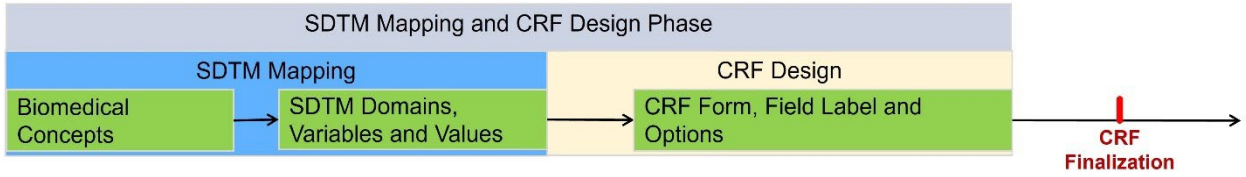
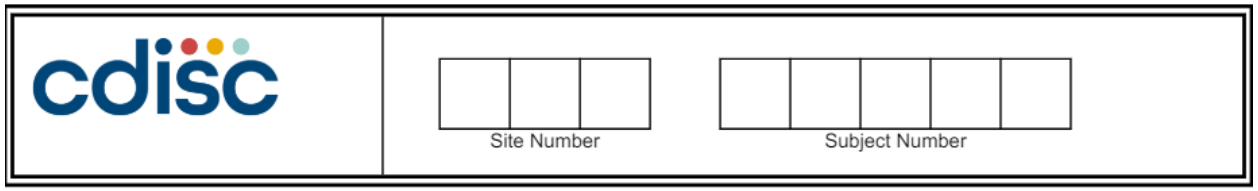


Figure 2. CDISC Process: Integrated SDTM Mapping and CRF Design Phase

CDASH PROCESS EXAMPLE: VITAL SIGNS FORM DESIGN



Form VS - Vital Signs HorizontalGeneric			
1.1	Were vital signs performed?	[N] No [Y] Yes	VSPERF
1.2	Date (DD-MMM-YYYY)		VSDAT
2 Vital Signs Details			
2.1	Systolic Blood Pressure		SYSBP_VSORRES
2.2	Systolic Blood Pressure Unit	mmHg	SYSBP_VSORRESU
2.3	Diastolic Blood Pressure		DIABP_VSORRES
2.4	Diastolic Blood Pressure Unit	mmHg	DIABP_VSORRESU
2.5	Blood Pressure Position	[PRONE] Prone [SEMI-RECUMBENT] Semi-Recumbent [SITTING] Sitting [STANDING] Standing [SUPINE] Supine	BP_VSPOS

Figure 3: [CDISC eCRF Portal Vital Signs]

Take an example of a CRF design for Vital Signs from CDISC to illustrate the process of mapping biomedical concepts to SDTM. Then, demonstrate how to map SDTM to the CRF.

According to the biomedical request form protocol, 'Vital signs will include measurements of blood pressure (both systolic and diastolic) along with

the patient's corresponding positions.' These biomedical concepts are mapped to the SDTM standard. Subsequently, a CRF is designed based on this SDTM mapping, as shown in Table 1

Table 1 Sample of Vital Signs Form Design

Biomedical concepts	SDTM Mapping	CRF Design
Vital signs	Map the topic to VS Domain	1. Map Domain Name “VS” and Domain Label “Vital Signs” to Form Name “Form VS - Vital Signs HorizontalGeneric”. 2. Map Domain Name “VS” to Form OID “VS_HORIZONTAL”.
Systolic Blood Pressure	Map the biomedical concept to corresponding values of VSTEST and VSTESTCD in SDTM Control Terminology: 1. VSTESTCD = SYSBP 2. VSTEST = Systolic Blood Pressure And map the result to SDTM variable VSORRES.	1. Map VSTEST value Systolic Blood Pressure to Field Label. 2. Map VSTESTCD Value “SYSBP” and SDTM variable VSORRES to Field OID “SYSBP_VSORRES”.
Blood Pressure Position	Map the biomedical concept to corresponding values of VSPOS in SDTM Control Terminology: 1. Map the result to SDTM variable VSPOS. 2. Map the code list to corresponding Control Terminology “POSITION” of VSPOS. a) Prone b) Semi-Recumbent c) Sitting d) Standing e) Supine	1. Map and specify Variable Label of SDTM Variable VSPOS “Vital Signs Position of Subject” to Field Label “Blood Pressure Position”. 2. Map and specify SDTM Variable Name VSPOS to Field OID “BP_VSPOS”. 3. Map VSPOS Control Thermology POSITION’s values to Field Options: a) Prone b) Semi-Recumbent c) Sitting d) Standing e) Supine

CDASH MODEL

Based on the CDASH model, the CRF specification can be designed as shown in Table 2. In this specification, the SDTM mapping rules are clearly outlined in the columns labeled "Domain," "SDTMIG Target," and "Mapping Instructions." However, the CDASH model alone is insufficient

for SDTM automation because the mapping rules are not machine-readable. Programmers must manually convert these SDTM mapping rules into programming code.

Table 2 Sample of CRF Specification for Vital Signs Form with CDASH Model

Domain	Variable Order	CDASHIG Variable	Question Text	SDTMIG Target	Mapping Instructions
VS	1.1	VSPERF	Were vital signs performed?	VSSTAT	This does not map directly to an SDTMIG variable. May be used to derive a value into the SDTM variable VSSTAT. If VSPERF="N", the value of VSSTAT will be "NOT DONE". If VSPERF="Y", VSSTAT should be null. A combination of SDTMIG variables (e.g., VSCAT and VSSCAT, VSTPT) is used to indicate that multiple tests were not done. In this situation, the SDTMIG variable VSTESTCD would be populated as VSALL and an appropriate test name VSTEST provided. See SDTMIG for additional information.
VS	1.2	VSDAT	What was the date of the vital signs measurement?	VSDTC	This does not map directly to an SDTMIG variable. For the SDTM submission dataset, concatenate all collected CDASH DATE and TIME components and populate the SDTMIG variable VSDTC in ISO 8601 format.
VS	2.1	SYSBP_VSORRES	What was the result of the Systolic Blood Pressure measurement?	VSORRES; VSTEST; VSTESTCD	Maps directly to the SDTMIG variable listed in the SDTMIG Target column. In addition to the SDTMIG variable VSORRES, create VSTESTCD from the CDASH variable name and determine the value of VSTEST from VSTESTCD. The CDASH prompt may also contain the VSTEST. Use appropriate CDISC Controlled Terminology for the test and test code.
VS	2.2	SYSBP_VSORRESU	What was the unit of the Systolic Blood Pressure measurement?	VSORRESU	Maps directly to the SDTMIG variable listed in the SDTMIG Target column.
VS	2.3	DIABP_VSORRES	What was the result of the Diastolic Blood Pressure measurement?	VSORRES; VSTEST; VSTESTCD	Maps directly to the SDTMIG variable listed in the SDTMIG Target column. In addition to the SDTMIG variable VSORRES, create VSTESTCD from the CDASH variable name and determine the value of VSTEST from VSTESTCD. The CDASH prompt may also contain the VSTEST. Use appropriate CDISC Controlled Terminology for the test and test code.
VS	2.4	DIABP_VSORRESU	What was the unit of the Diastolic Blood Pressure measurement?	VSORRESU	Maps directly to the SDTMIG variable listed in the SDTMIG Target column.
VS	2.5	BP_VSPOS	What was the position of the subject during the Blood Pressure measurement?	VSPOS	Maps directly to the SDTMIG variable listed in the SDTMIG Target column.

CDASH-SDTM COMBO MODEL

The CDASH-SDTM Combo Model provides a machine-readable structure at the variable value level for CRF design and SDTM mapping, helping to facilitate SDTM automation.

Table 3 CDASH-SDTM Combo Model

One Record per Field Option per Field OID per Form OID per SDTM Variable Value Per SDTM Variable per SDTM Domain.

Column	Description
Form OID	Dataset name in Electronic Data Capture (EDC) systems.
Form Name	CRF Form Name
Field Order	Field Sequence Number is CRF.
Field OID	Variable Name in Electronic Data Capture (EDC) systems.
Field Label	Filed Label or Question Text in CRF.
Field Option Order	Field Option Sequence Number in CRF.
Field Option	Field Option in CRF.
SDTM Domain	SDTM Domain Abbreviation.
SDTM Variable	SDTM Variable Name.
SDTM Variable Value	SDTM Variable Value.
SDTMRecordID	All SDTM Variable Values with the Same SDTMRecordID Values are mapped to One SDTM Record per Form per Subject.
Algorithm	Algorithm of Mapping Each Field Option to its Corresponding SDTM Variable Value

There are three types of Columns in CDASH-SDTM Combo Model：

1. CRF Design Columns: These include three hierarchical levels:
    a) Form Level: Form OID and Form Name.
    b) Field Level: Field Order, Field OID and Field Label.
    c) Field Option Level: Field Option Order, Field Option.
2. SDTM Target Columns: These consist of three hierarchical levels and one grouping variable for SDTM records:
    a) Three Hierarchical Levels.
        i. Domain Level: SDTM Domain.
        ii. Variable Level: SDTM Variable.
        iii. Variable Value Level: SDTM Variable Value.
    b) Grouping variable for SDTM records: SDTMRecordID.
3. Mapping Algorithm: This maps each field option to its corresponding SDTM variable value.

Example for Vital Signs

Based on Vital Signs Form in Figure 3, CDASH-SDTM Combo Model is implemented as Table 4.

1. The option "Yes" for the question "Were vital signs performed?" is not submitted, indicating Algorithm = NotSubmitted. This is because the information can be derived from other specific vital signs tests in the form, leaving the SDTM-related columns blank.
2. When the option "No" is selected for the question "Were vital signs performed?", the information is mapped to VSTAT = NOT DONE where VSTESTCD = VSALL and VSTEST = Vital Signs. This triggers an SDTM record identified by SDTMRecordID = VSPERF. The algorithm "IfAssigned" means that if "No" is ticked, the SDTM record will be created, and the corresponding variable values will be assigned.
3. The field "Date (DD-MMM-YYYY)" is mapped to VSDTC using the "DT2DTC" algorithm, which converts the date format to ISO8601 (e.g., YYYY-MM-DD). The SDTMRecordID is left blank, indicating that the date will be applied to all Vital Signs tests.
4. The field "Systolic Blood Pressure" is mapped to VSORRES where VSTESTCD = SYSBP and VSTEST = Systolic Blood Pressure. This triggers an SDTM record identified by SDTMRecordID = SYSBP_VSORRES. The VSORRES value is copied from the field value using the "Copy" algorithm, and VSTESTCD = SYSBP and VSTEST = Systolic Blood Pressure are assigned using the "Assigned" algorithm.
5. The field "Systolic Blood Pressure Unit" is a qualifier for Systolic Blood Pressure. Its value is copied to VSORRESU using the "Copy" algorithm in the same SDTM record as Systolic Blood Pressure. Therefore, its SDTMRecordID value "SYSBP_VSORRES" is the same as the field "Systolic Blood Pressure".
6. The field "Blood Pressure Position" has its values copied to VSPOS, with the case changed to uppercase for consistency with CDISC Submission Value in SDTM Terminology. The algorithm used is "CopyUpcase". The field "Blood Pressure Position" is only applied to Blood Pressure tests, and this is controlled by SDTMRecordID = SYSBP_VSORRES | DIABP_VSORRES, with the two SDTM Record IDs separated by "|".

Table 4 CDASH-SDTM Combo Model Specifications for Vital Signs

Form OID	Form Name	Field Order	Field OID	Field Label	Field Option Order	Field Option	SDTM Domain	SDTM Variable	SDTM Variable Value	SDTMRecordID	Algorithm
VS_HORIZONTAL	Vital Signs Horizontal Generic	1.1	VSPERF	Were vital signs performed?	1	Yes					NotSubmitted
VS_HORIZONTAL	Vital Signs Horizontal Generic	1.1	VSPERF	Were vital signs performed?	2	No	VS	VSTESTCD	VSALL	VSPERF	IfAssigned
VS_HORIZONTAL	Vital Signs Horizontal Generic	1.1	VSPERF	Were vital signs performed?	2	No	VS	VSTEST	Vital Signs	VSPERF	IfAssigned

VS_HORIZONTAL	Vital Signs Horizontal Generic	1.1	VSPERF	Were vital signs performed?	2	No	VS	VSSTAT	NOT DONE	VSPERF	IfAssigned
VS_HORIZONTAL	Vital Signs Horizontal Generic	1.2	VSDAT	Date (DD-MMM-YYYY)			VS	VSDTC	YYYY-MM-DD		DT2DTC
VS_HORIZONTAL	Vital Signs Horizontal Generic	2.1	SYSBP_VSORRES	Systolic Blood Pressure			VS	VSTESTCD	SYSBP	SYSBP_VSORRES	Assigned
VS_HORIZONTAL	Vital Signs Horizontal Generic	2.1	SYSBP_VSORRES	Systolic Blood Pressure			VS	VSTEST	Systolic Blood Pressure	SYSBP_VSORRES	Assigned
VS_HORIZONTAL	Vital Signs Horizontal Generic	2.1	SYSBP_VSORRES	Systolic Blood Pressure			VS	VSORRES		SYSBP_VSORRES	Copy
VS_HORIZONTAL	Vital Signs Horizontal Generic	2.2	SYSBP_VSORRESU	Systolic Blood Pressure Unit			VS	VSORRESU		SYSBP_VSORRES	Copy
VS_HORIZONTAL	Vital Signs Horizontal Generic	2.3	DIABP_VSORRES	Diastolic Blood Pressure			VS	VSTESTCD	DIABP	DIABP_VSORRES	Assigned
VS_HORIZONTAL	Vital Signs Horizontal Generic	2.3	DIABP_VSORRES	Diastolic Blood Pressure			VS	VSTEST	Diastolic Blood Pressure	DIABP_VSORRES	Assigned
VS_HORIZONTAL	Vital Signs Horizontal Generic	2.3	DIABP_VSORRES	Diastolic Blood Pressure			VS	VSORRES		DIABP_VSORRES	Copy
VS_HORIZONTAL	Vital Signs Horizontal Generic	2.4	DIABP_VSORRESU	Diastolic Blood Pressure Unit			VS	VSORRESU		DIABP_VSORRES	Copy
VS_HORIZONTAL	Vital Signs Horizontal Generic	2.5	BP_VSPOS	Blood Pressure Position	1	Prone	VS	VSPOS	PRONE	SYSBP_VSORRES| DIABP_VSORRES	CopyUpcase

VS_HORIZONTAL	Vital Signs Horizontal Generic	2.5	BP_VSPOS	Blood Pressure Position	2	Semi-Recumbent	VS	VSPOS	SEMI-RECUMBENT	SYSBP_VSORRES| DIABP_VSORRES	CopyUpcase
VS_HORIZONTAL	Vital Signs Horizontal Generic	2.5	BP_VSPOS	Blood Pressure Position	3	Sitting	VS	VSPOS	SITTING	SYSBP_VSORRES| DIABP_VSORRES	CopyUpcase
VS_HORIZONTAL	Vital Signs Horizontal Generic	2.5	BP_VSPOS	Blood Pressure Position	4	Standing	VS	VSPOS	STANDING	SYSBP_VSORRES| DIABP_VSORRES	CopyUpcase
VS_HORIZONTAL	Vital Signs Horizontal Generic	2.5	BP_VSPOS	Blood Pressure Position	5	Supine	VS	VSPOS	SUPINE	SYSBP_VSORRES| DIABP_VSORRES	CopyUpcase

T PROGRAMMING LANGUAGE

CDASH-SDTM Combo Model Specifications, as shown in Table 4, contains all the necessary information to transform CRF data into SDTM. All information in the Specifications is machine-readable, which means that the CDASH-SDTM Combo Model Specifications can be considered a type of programming language. This programming language is written in a structured data format, specifically a table, and can be referred to as the T Programming Language, with "T" standing for "Table."

The advantages of the T Programming Language are as follows:

1. The design of CRF and the mapping to SDTM are digitized into structured data. This information is modularized at the field level, and even at the field option level, making the CRF design (including CRF Completion Guides (CCG) and Edit Check) and SDTM mapping recyclable by field or field option.
2. The T Programming Language, being in structured data, can be automatically converted to other programming languages such as SAS, R, and Python.
3. The T Programming Language can also be executed directly by other programming languages without needing to be transformed into corresponding code.
4. If the CDASH-SDTM Combo Model Specifications are integrated into Electronic Data Capture (EDC) systems, SDTM data, excluding derived SDTM data, and SDTM aCRF can be directly exported from the EDC.

The T Programming Language is applicable not only to SDTM automation but also to table automation (Chunpeng Zhao, 2023).

CONCLUSION

The CDASH-SDTM Combo Model integrates CRF design and SDTM mapping into a single step, facilitating the automation of both processes. This model enhances compliance with CDISC standards by incorporating a pre-load CDISC compliance check during the CRF design phase. This ensures consistency across the CRF, TLF shell, and SDTM data, thereby minimizing the need for iterative changes in wording. Additionally, it stabilizes SDTM mapping rules and prevents costly late changes that could impact downstream ADaM/TFLs. ADaM and TFL work can be initiated during the CRF design phase because the structure of the SDTM data is already known at that stage. This enables real-time analysis to begin as soon as the first record is collected if everything is prepared in advance.

REFERENCES

Chunpeng Zhao etc., 2022. “Real-time SDTM” https://www.lexjansen.com/pharmasug-cn/2022/AT/Pharmasug-China-2022-AT105.pdf

CDISC eCRF Portal Vital Signs https://www.cdisc.org/kb/ecrf/vital-signs

Chunpeng Zhao, 2023. “Meta-programming in Table Automation” https://www.lexjansen.com/pharmasug-cn/2023/AP/Pharmasug-China-2023-AP103.pdf

CONTACT INFORMATION

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

Chunpeng Zhao
BeiGene
Chunpeng.zhao@beigene.com