Paper DS22

Optimizing Trial Design Dataset/s Creation to Minimize P21 Compliance Issues

Rama Kudaravalli, Merck & Co., Inc., Rahway, NJ, USA

ABSTRACT

This paper presents key considerations for effectively mapping trial design domains to ensure compliance and reduce issues identified by P21 Enterprise checks. Trial design datasets encapsulate the planned structure of a clinical study, including treatment arms, protocol-defined schedules of subject visits, and other essential design elements. Accurate alignment of these datasets with the study protocol is critical to reflect the intended research objectives and operational plans. Once trial design domains are created, they are validated using P21 Enterprise, and any unresolved issues must be documented in the Clinical Study Data Reviewer's Guide (cSDRG). This paper outlines best practices and strategic measures for mapping trial design domains to ensure high-quality, compliant data and minimize the need for post-submission issue resolution.

.

INTRODUCTION

The trial design domains can be mapped before or after generating the study SDTM domains. If the trial domains (TDs) are generated before other SDTM domains, TDs can be used as input for other SDTM domains otherwise vice versa. In both methods, we need some important documents like latest protocol, related SDTM IG, latest CRF, specification template, Inclusion/Exclusion criteria including all versions of protocol amendments. After specification mappings are done, the trial domains are generated and checked for compliance by running the P21 enterprise tool, including the remaining SDTM domains.

Trial domains needed for a study

The trial domains are mapped as per the planned events in the protocol. Initially, 5 trial domains (TA, TE, TI, TS, TV) were designed in SDTM (up to SDTMIG v3.1.3). In SDTMIG v3.2 and v3.3 the TD (Trial Disease Assessments) and TM (Trial Disease Milestones) domains were added respectively. These TA, TE, TI, TS, TV, TD, TM trial domains provide a brief description of the overall plan and design of the trial. Missing some of these domains will cause compliance issues. P21 validation rules also vary depending on the SDTM IG version used in mapping the trial domains for a study. In this paper the validation rules are not specified based on agency name.

Trial Arms (TA) - Describes the sequences of elements in each epoch for each arm, and thus describes the complete sequence of elements in each arm. TAETORD is an integer. The values of TAETORD need not be sequential but their order must be in the correct order for the elements in the arm path.

Trial Elements (TE) - It contains the element code that is unique for each element, the element description, and the rules for starting and ending an element.

Trial Inclusion/Exclusion (TI) - Planned eligibility criteria are mapped.

Trial Summary (TS) - High-level study attributes like phase, indication, objectives, randomization, blinding are mapped.

Trial Visits (TV) - Planned visit schedule (names, days, windows) is mapped as per protocol.

Trial Disease Assessments (TD) - Planned disease assessment schedule (efficacy assessments related to the disease under study).

Trial Disease Milestones (TM) -Describe disease milestones, which are observations or activities anticipated to occur in the course of the disease under study, and which trigger the collection of data.

Core causes of P21 issues

The TA, TE, TS domains should be included in every submission for SDTM, missing these domains will trigger SD1112, SD1113, SD1115 issues in P21 report and applicable to data mapped with SDTMIG v3.1.2 and above. Missing trial summary (TS) domain will cause a reject issue (SD1115) in P21 report.

TA and TE domains

The TA and TE datasets are interrelated, and they provide the building blocks for the development of subject-level treatment information in DM and SE domains. Combination of ETCD and ELEMENT variables should be unique and match across the trial domains and other SDTM domains in a study.

The ETCD and ELEMENT variables should match in TA, TE and SE domains. If the ETCD/ELEMENT values do not match across TA, TE, and SE (except for UNPLAN element code in SE) domains, SD0067 issue is triggered. If the

value of ETCD is more than 8 character length then it leads to SD1009 issue. The ETCD and ELEMENT combination should match across TA, TE, and SE domains. If the combination is inconsistent (except for UNPLAN in SE) then SD1012 is triggered. The ETCD/ELEMNET must have unique values (TA, TE domains), otherwise SD1050 (ETCD), 1027 (ELEMENT) issues will be triggered. If ETCD/ELEMNET values are not unique then SD1064, SD1068 issues will be triggered respectively (See table 1). All these validation rules are applicable for STMIG v3.1.2 and above.

Rule	Requirement	Allowed Exception	Error if Violated	Domains
ETCD/ELEMENT match across TA, TE, SE	Must match	UNPLAN in SE	SD0067	TA
ETCD length	≤ 8 characters	None	SD1009	TA, TE
ETCD/ELEMENT combination consistency	Must match across domains	UNPLAN in SE	SD1012	TA
ETCD/ELEMENT uniqueness	Must be unique	None	SD1050, SD1027, **SD1064, SD1068**	TA, **TE**

Table 1. P21 issues related to ETCD/ELEMENT in TA and TE domains.

TA and TV domains

ARMCD and ARM values must maintain one-to-one mapping in TA and TV domains. If the value of Planned Arm Code (ARMCD) is more than 20 character length then SD1004 issue (DM, TA, TV) is triggered. For a given ARMCD, there must be exactly one corresponding ARM value within a domain. Inconsistent value for ARM and ARMCD will trigger SD1033, SD1034 issues respectively. Records for subjects, who failed a screening or were not assigned to study treatment (ARMCD is 'SCRNFAIL' or 'NOTASSGN') should not be included in the Trial Arms (TA) or Trial Visits (TV) datasets. Unexpected value for ARMCD variable triggers SD1053 issue. Arm Code (ARMCD) values defined in Trial Arms (TA) should be assigned to at least one subject (USUBJID) in Demographics (DM). If the ARMCD from TA domain is not in DM domain it triggers the SD1354 issue. TAETORD is planned order of elements within an arm. It must be unique within each ARMCD. Non-unique value for TAETORD within ARMCD triggers SD1052 issue. TAETORD must be a whole number. Invalid TAETORD value triggers SD1267 issue (See table 2). All the issues related to TA and TV domains mentioned below are applicable for the study data mapped in SDTMIG v3.1.2 and above except for SD1354 issue, which is applicable for SDTMIG v3.2 and above.

Rule	Requirement	Error if Violated	Domains
ARMCD length	≤ 20 characters	SD1004	TA, TV
ARM/ARMCD uniqueness	Must be unique	SD1033, SD1034	TA, TV
ARMCD values ('SCRNFAIL' or 'NOTASSGN') should not be included in the TA or TV	ARMCD variable should not contain unexpected values	SD1053	TA, TV
ARMCD values from TA domain should be assigned to at least one subject in DM	DM should have the ARMCD value from TA	SD1354	TA
TAETORD/ARMCD uniqueness	Must be unique	SD1052	TA
TAETORD value must be an integer	TAETORD should have valid value	SD1267	TA

Table 2. P21 issues related to ARMCD/ARM/TAETORD in TA and TV domains.

TE domain
TEDUR value must be greater than or equal to 0. Negative value in TEDUR triggers SD0015 issue. Either TEDUR or TEENRL values need to be populated in TE domain. Missing values for TEENRL and TEDUR will trigger SD0089 issue. If the combination of ELEMENT, TESTRL TEENRL, and TEDUR is not unique for each ETCD then SD1271 is triggered. If a ETCD value is missing in TA but present in TE will trigger SD1378 issue (See table 3).

Rule	Requirement	Error if Violated	Domains
TEDUR value must be greater than or equal to 0	TEDUR cannot have negative values	SD0015	TE
TEDUR or TEENRL need to be populated	TEENRL and TEDUR cannot be missing	SD0089	TE
Combination of ELEMENT, TESTRL TEENRL, and TEDUR unique for each ETCD	Combination of variables must be unique	SD1271	TE
ETCD value is missing in TA but present in TE	ETCD value must be present in TA and TE	SD1378	TE

Table 3. P21 issues related to TEDUR/TESTRL/TEENRL in TE domain.

TI domain
Trial Inclusion/Exclusion Criteria (TI) domain describes one record for each of the inclusion and exclusion criteria for the trial. The IETESTCD variable should be limited to 8 characters, cannot start with a number, and cannot contain characters other than letters in upper case, numbers, or underscores. Invalid value for IETESTCD variable triggers SD0018 issue. IETEST/IETESTCD values must be unique. Inconsistency in IETEST/IETESTCD values trigger SD1150, SD1151 issues. If IETESTCD is not unique within TIVERS then SD1286 issue is triggered (See table 4).

Rule	Requirement	Error if Violated	Domains
IETESTCD cannot start with a number, and cannot contain characters other than letters in upper case, numbers, or underscores	≤ 8 characters length	SD0018	TI
IETEST/IETESTCD uniqueness	Must be unique	SD1150, SD1151	TI
IETESTCD should be unique within TIVERS	Must be unique	SD1286	TI

Table 4. P21 issues related to IETESTCD/IETEST/TIVERS in TI domain.

TS domain
Trial Summary (TS) domain lists key facts (parameters) about the trial that are likely to appear in a registry of clinical trials. The trial summary parameters should describe the protocol.

The trial summary parameters should be included in TS domain as per the Technical Specifications Document, Study Data Technical Conformance Guide. Additional parameters can be added to the TS domain as per the study design and protocol. The TS parameters can be classified into different categories based on their characteristics. As per the FDA SD-TCG some of the trial summary parameters are required (See table 5) and others are conditionally required (See table 6) for submission.

Both the required and conditionally required parameters are mapped in TS domain. Additional parameters can be added to TS domain by the sponsor as per the protocol and study design. The trial summary parameter values are listed under CDISCT CT. The codelist for TS parameters is extensible.

Missing required trial summary parameter (TSPARMCD/TSPARMs) values will trigger SD22xx issues. For example, if

the TSPARMCD = 'COMPTRT' and TSVAL (preferred term)/TSVALCD (unique ingredient identifier) values are not populated as per the FDA Substance Registration System then SD2253, SD2254 issues are triggered respectively.

Category	TSPARMCD	TSPARM	P21 Rule
Treatment / Intervention	COMPTRT	Comparative Treatment Name (Invalid TSVAL and TSVALCD)	SD2253, SD2254
	THERAREA	Therapeutic Area	SD2282
	PDPSTIND	Pediatric Postmarket Study Indicator	SD2276
	PDSTIND	Pediatric Study Indicator	SD2277
	PIPIND	Pediatric Investigation Plan Indicator	SD2278
	RDIND	Rare Disease Indicator	SD2279
Objectives	OBJPRIM	Trial Primary Objective	SD2217
	OBJSEC	Trial Secondary Objective	SD2275
	OUTMSPRI	Primary Outcome Measure	SD2222
Study Design / Trial Characteristics	ADDON	Added on to Existing Treatments	SD2201
	LENGTH	Trial Length	SD2204
	PLANSUB	Planned Number of Subjects	SD2205
	SEXPOP	Sex of Participants	SD2207
	STOPRULE	Study Stop Rules	SD2208
	TBLIND	Trial Blinding Schema	SD2209
	TCNTRL	Control Type	SD2210
	TITLE	Trial Title	SD2213
	TPHASE	Trial Phase Classification	SD2214
	TTYPE	Trial Type	SD2215
	STYPE	Study Type	SD2230
	EXTTIND	Extension Trial Indicator	SD2273
	ADAPT	Adaptive Design	SD2225
Population / Enrollment	AGEMAX	Planned Maximum Age of Subjects	SD2202
	AGEMIN	Planned Minimum Age of Subjects	SD2203
	ACTSUB	Actual Number of Subjects	SD2234
	HLTSUBJI	Healthy Subject Indicator	SD2235
	NCOHORT	Number of Groups/Cohorts	SD2274
	NARMS	Planned Number of Arms	SD2229
Dates	DCUTDTC	Data Cutoff Date	SD2226
	DCUTDESC	Data Cutoff Description	SD2227
	SSTDTC	Study Start Date	SD2232
	SENDTC	Study End Date	SD2233
Standards Versions	SDTIGVER	SDTM IG Version	SD2280
	SDTMVER	SDTM Version	SD2281
Sponsor / Regulatory	SPONSOR	Clinical Study Sponsor	SD2218
	REGID	Registry Identifier	SD2221
Geography	FCNTRY	Planned Country of Investigational Sites	SD2224
Study Status	ONGOSIND	Ongoing Study Indicator	SD2287

Table 5. P21 issues related to required TSPARMCD/TSPARM values in TS domain.

Missing conditionally required parameters will trigger additional issues other than SD22xx issues in P21 report. The conditionally required parameters will depend on other trial summary parameters. If the values are missing for conditional parameters and other dependent parameters may trigger more than one P21 issue in the report.

For example, missing TSVAL when TSPARMCD = 'INTMODEL' will cause SD2228 issue. In addition to this, missing TSVAL when TSPARMCD = 'INTMODEL' and TSPARMCD = 'STYPE' with TSVAL = 'INTERVENTIONAL' combination will trigger SD1310 issue. Similar issues occur for TSPARMCD = 'INTTYPE'. When TSVAL is missing for the TSPARMCD = 'INTTYPE' and TSPARMCD = 'STYPE' with TSVAL = 'INTERVENTIONAL', this combination will trigger SD1311 issue. See table 6 for further details.

TSPARMCD	TSPARM	Dependency	P21 issue
CURTRT	Current Therapy or Treatment	ADDON = 'Y' and CURTRT is missing	SD2216
		TSVCDREF value must be 'UNII'	SD2241
		Invalid TSVAL value for CURTRT	SD2250
		Invalid TSVALCD value for CURTRT	SD2251
		TSVAL/TSVALCD value mismatch for CURTRT	SD2252
		ADDON = 'Y' and TSVAL is missing for CURTRT	SD1308
INDIC	Trial Disease/Condition Indication	TSVCDREF value must be 'SNOMED'	SD2240
		Invalid TSVAL value for INDIC	SD2257
		Invalid TSVALCD value for INDIC	SD2258
		TSVAL/TSVALCD value mismatch for INDIC	SD2259
PCLAS	Pharmacologic Class	TSVCDREF value must be 'MED-RT' or 'NDF-RT'	SD2243
		Invalid TSVAL value for PCLAS	SD2263
		Invalid TSVALCD value for PCLAS	SD2264
		TSVAL/TSVALCD value mismatch for PCLAS	SD2265
		STYPE = 'INTERVENTIONAL' and TSVAL is missing for PCLAS	SD1312
TDIGRP	Diagnosis Group	TSVCDREF value must be 'SNOMED'	SD2266
		Invalid TSVAL value for TDIGRP	SD2267
		Invalid TSVALCD value for TDIGRP	SD2268
		HLTSUBJI = 'Y' and TSVAL is not missing for TDIGRP	SD1306
		HLTSUBJI = 'N' and TSVAL is missing for TDIGRP	SD1307
TRT	Investigational Therapy or Treatment	TSVCDREF value must be 'UNII'	SD2242
		Invalid TSVAL value for TRT	SD2260
		Invalid TSVALCD value for TRT	SD2261
		STYPE = 'INTERVENTIONAL' and TSVAL is missing for TRT	SD1309
INTMODEL	Intervention Model	STYPE = 'INTERVENTIONAL' and TSVAL is missing for INTMODEL	SD1310
INTTYPE	Intervention Type	STYPE = 'INTERVENTIONAL' and TSVAL is missing for INTTYPE	SD1311
RANDQT	Randomization Quotient	TSVAL value must be a number between 0 and 1, with some cases where the ratio might be equal to 1	SD1269

Table 6. P21 issues related to conditionally required TSPARMCD/TSPARM values in TS domain.

Controlled Terminology

A list of values for TSPARM and TSPARMCD can be found in CDISC Controlled Terminology. The US FDA Data Standards Catalog includes terminologies to be used for certain trial summary parameters. For a particular trial summary parameter, responses (values in TSVAL) may be numeric, datetimes or amounts of time represented in

ISO8601 format, or text. For some parameters, textual responses may be taken from controlled terminology; for others, responses may be free text. TSVCDREF variable contains the reference terminology used for parameter. TSVCDREF values must be mapped as per DICTNAM extensible codelist.

CONCLUSION

The trial domains provide the design of a clinical trial, help reviewers to compare planned and actual treatments and scheduled visits for subjects in a trial. These domains are mapped as per the planned protocol and can be used as inputs for certain variables in generation of other SDTM domains in a study. Missing some of these trial domains may cause compliance issues especially with missing trial summary domain in submission package causes a reject issue as per P21 validation rules. Combination of certain variables like ARM/ARMCD, ELEMENT/ETCD must have one-to-one mapping and maintain the uniqueness in trial domains. There are required and conditional variables in trial summary domain as per the FDA SD-TCG and missing these variables in submission may trigger SD22xx and other (e.g., SD13xx) compliance issues. Appropriate controlled terminologies like CDISC-CT (extensible/non-extensible), UNII, SNOMED, MED-RT and other related dictionaries must be used in mapping these domains. Duration and dates must use ISO 8601 format. To map country of investigation sites, 3-letter country code must follow ISO 3166 dictionary. Applying the CDISC-CT, FDA SD-TCG, other dictionaries are helpful in creating compliant trial datasets for submission and minimizing P21 issues. Considering the required trial domains, required and conditional parameters, and use of appropriate controlled terminologies as per the FDA SD-TCG and CDISC rules are helpful in generating compliant datasets for submissions.

REFERENCES

1. sdTCG-v6.1-December-2025-FINAL
2. FDA Validator Rules v1.6 December 2022_0
3. CDISC SDTMIG v3.1.2 to v3.4

CONTACT INFORMATION

Rama Kudaravalli
Merck & Co., Inc., Rahway, NJ, USA
Rama.kudaravalli@merck.com