CERTARA

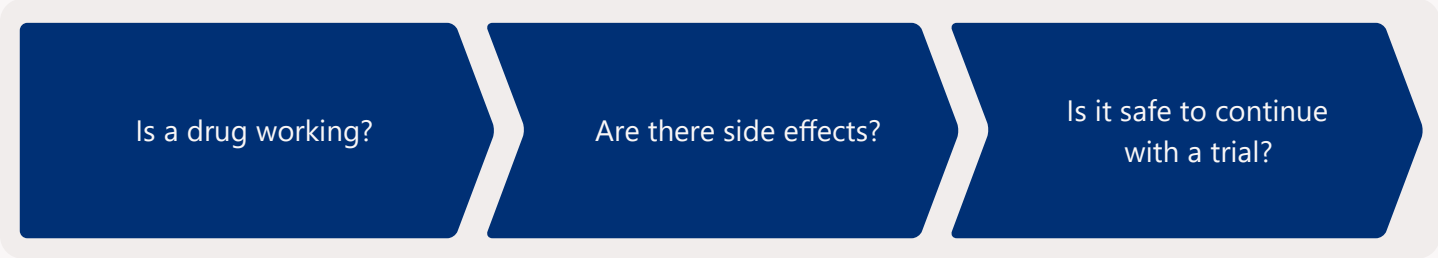
Pinnacle 21® - SDTM Specification Management

A best practice guide to CDISC SDTM mapping

Introduction

When it comes to clinical trials, data is everything.

Timely, clear and actionable data insights allow important clinical decisions to be made in real time. And earlier insights mean life-changing or life-enhancing medicines can be brought to those in need faster.



These questions can be answered by clear, comprehensive data analytics. But to get this information, you must first produce quality SDTM datasets. This can mean lots of manual spreadsheet work and programming to get your data into the required format. The process of mapping raw data to the SDTM format is known as SDTM mapping.

Whether you're mapping to SDTM before data collection or after, it's no simple task. That's why we've created this guide to help you understand the process and what tools are out there to help you map to SDTM more easily.

The CDISC SDTM structure

CDISC standards were put in place to allow the US Food and Drug Administration (FDA) to quickly and easily analyze clinical trial data to understand the safety and efficacy of new drugs.

For clinical trial data to be accepted by regulatory reviewers, the submission must be in the correct format. If not, it may mean lots of time, money, and hard work down the drain. At worst, it could mean the entire trial is jeopardized.

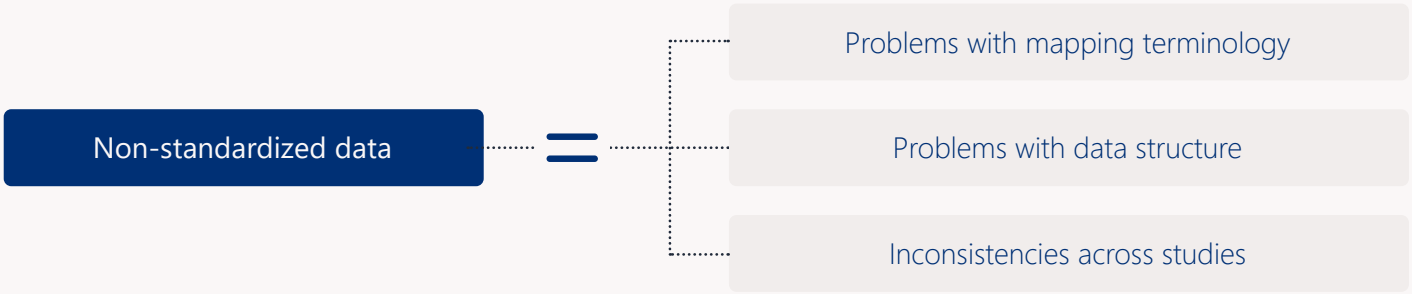
Implementing CDISC standards has a lot of benefits for the submitter:

- Better data quality
- Greater efficiency in the trial process
- Reduced trial timeline and costs
- Processes are streamlined
- Allows easier sharing of data
- Full traceability in the clinical research process from start to end
- Allows for greater innovation

The challenges of retrospective SDTM mapping

Typically, data is aligned with SDTM at the end of the trial after all patient data has been collected. But working with non-standardized data can pose some challenges, including terminology and structure inconsistencies. The manual work required to map non-standardized data to SDTM can be time-consuming and often leads to submission delays.

Later in this guide, we talk about the best practice approach you can adopt to avoid the headache of retrospective SDTM mapping.



The SDTM mapping specification document

SDTM mapping can be a complicated task, so it's important to plan everything out in advance. To help you do this, you'll need an SDTM mapping specification document.

Why do you need an SDTM mapping specification document?

This document specifies how the raw data is to be converted and is used by the SDTM programmer and testing team.

During the data mapping process, there's a risk that data can be lost or distorted. Consider a scenario where your source data is in text format, and your target data should be enumerated. Unless you provide a logical specification for how the text values should be mapped to the required values, you might get errors in the resulting data.

The specification document helps you identify and resolve these potential errors before you start data mapping.

The specification document can be created manually as follows:

1

Determine the SDTM domains needed

Examine the CRFs and raw data and identify which SDTM domains you need.

2

Note the relevant raw dataset for each

Against each SDTM domain, note which raw dataset will provide the input data.

3

List and describe programming for all variables

Against each SDTM domain, list all variables and describe how they are to be programmed.

The mapping process

To be CDISC compliant, raw datasets must be mapped from the structure used in your clinical data management system (or another database) to the CDISC SDTM structure.

This can be a long, complex, and daunting process. Not least because there isn't always a perfect match. Here, we describe the typical SDTM mapping process.

1. Identify the datasets you want to map
2. Identify the SDTM datasets that correspond to those datasets
3. Gather the metadata of the datasets and the corresponding SDTM metadata
4. Map variables in the datasets from Step 1 to SDTM domain variables
5. Create custom domains for other datasets that don't have corresponding SDTM datasets

Typical mapping scenarios

There are nine likely scenarios in a typical SDTM mapping process. Get to grips with these, and SDTM mapping becomes much more achievable.

The direct carryforward

These are variables that are already SDTM compliant. These can be directly carried forward to the SDTM datasets, and they don't need to be modified.

The variable attribute change

As well as variable names, variable attributes must be mapped. Attributes such as label, type, length, and format must comply with the SDTM attributes.

The reformat

The value that is represented doesn't change, but the format it's stored in does. For example, converting a SAS date to an ISO 8601 format character string.

The combine

In some cases, multiple variables must be combined to form a single SDTM variable.

The split

A non-SDTM variable might need to be split into two or more SDTM variables to comply with SDTM standards.

The derivation

Some SDTM variables are obtained by deriving a conclusion from data in the non-SDTM dataset. For example, using date of birth and study start date to derive a patient's age, instead of manually entering the age upfront.

The variable value map and new code list application

Some variable values need to be recoded or mapped to match the values of a corresponding SDTM variable.

This mapping is recommended for variables with a code list attached that has non-extensible controlled terminology.

It's also advisable to map all values in the controlled terminology, rather than just for the values present in the dataset. This would cover values that are not in the dataset currently but may come in during future dataset updates.

The horizontal to vertical data structure transpose

If the structure of the non-CDISC dataset is completely different from its corresponding SDTM dataset, you may need to transform it to one that is SDTM compliant.

The Vital Signs dataset is a good example. When data is collected in wide form, every test and recorded value is stored in separate variables. As SDTM requires data to be stored in a vertical form, the dataset must be transposed to have the tests, values, and unit under three variables. If there are variables that cannot be mapped to an SDTM variable, they would go into supplemental qualifiers.

Remember, you might need to use more than one type of mapping to create an SDTM variable.

SDTM mapping best practice

If the above process sounds daunting, don't worry! Our best practice approach explains some things you can do to help make the mapping process quicker and easier, accelerating your time to submission.

1. Read the SDTM implementation guide

Before you begin mapping, it's important to have a basic understanding of how SDTM works. Read through the whole SDTM Implementation Guide (SDTMIG) at least once. It's an essential resource to help you with data mapping. Before programming on a specific domain, read the part in the guide about that domain.

2. Create trial design domains

Trial Design domains are datasets that describe the planned structure and organization of a clinical trial, as defined by the protocol. These domains are crucial for understanding how the study is intended to be conducted and are essential for data validation and submission readiness.

3. Implement SDTM from the start of your trial

When data is aligned with SDTM at the end of the trial, trying to make it fit the SDTM structure retrospectively can be a headache. It takes a lot of time and manual work to retrospectively map data, and this can lead to submission delays.

So, best practice is always to align with CDISC standards before you collect any patient data. In fact, we recommend you consider the SDTM format when designing your case report forms (CRFs). Doing it this way means you'll save a lot of time and effort manually mapping the data down the line. And it's much easier and less time consuming to pull your submission deliverables together. This means you can submit your studies much faster!

4. Map your data upfront

EDC build can be a lengthy process, not to mention the time it takes to gather first patient data. This means it can be many weeks before you can see how data and metadata look and start the SDTM mappings.

So, what if you could see what your EDC datasets will look like before you get them? There are solutions to help you create simulated datasets before you've collected any patient data. That way, you get SDTM conversion done early on and can check that your datasets will meet regulatory requirements. You also get a head start on mapping your ADaM datasets back to SDTM.

What if I'm collecting non-EDC data?

The use of non-CRF data is increasing in modern clinical trials. In fact, over 70% of clinical data now originates from non-CRF sources such as biomarkers, labs, and imaging.

Data Transfer Specifications, also known as specs or DTS, detail how non-CRF data should be collected to ensure seamless information exchange between vendors and the research organizations. The specifications are created by the data management team with input from biostatistics and programming teams, CROs, and other external vendors. This should be done before data is collected to ensure it is collected exactly how you need it.

The specification itself should be compliant with CDISC and regulatory requirements, as well as in-house standards.

Data Transfer Specifications not only define your datasets upfront and ease data collection but also accelerate time to submission.

5. Do SDTM dataset validation

Another important step is CDISC SDTM validation: this means validating SDTM dataset designs against CDISC standards (SDTM-IG, NCI controlled terms etc). The FDA and PMDA publish compliance rules for clinical data submissions, including SDTM. The output must ultimately be compared against the rules, and any deviations must be resolved or fully documented. Validation is a crucial way to increase data quality while reducing time spent in development and review.

By building in SDTM compliance at the start of your trial, you maximize compliance throughout the process. Key factors such as correct use of controlled terms must be built into the EDC CRFs and SDTM conversion process. This includes EDC edit check programming and data management checks programmed to highlight data issues early.

Best practices

- Do SDTM dataset validation
- Read the SDTM implementation guide
- Implement SDTM from the start of your trial
- Make your data upfront

Certara has you covered

Our Pinnacle 21 (P21) clinical data management and automation suite can help you align with these best practice methods, save time, and avoid delays.

- Design CRFs to capture all the required data, in a standardized format
- Ditch the spreadsheets! P21 helps with SDTM mappings, and does conversions for you
- Create human-readable or machine-readable mappings
- No need for expertise in SAS programming
- Reuse standardized mappings over and over again for faster study designs, and better-quality data
- Get reports that give you full traceability of all mappings to ensure data is accurately mapped
- Easily transfer non-CRF and external vendor data to your specification
- Flexible mapping specification that supports downstream processes and can work with a variety of platforms
- In-stream validation and structural checks ensure high quality deliverables from the outset

What you've learned

- The purpose and importance of the CDISC SDTM structure
- That it's best practice to consider SDTM up front, before you begin collecting data
- The process for mapping raw datasets, and how using an SDTM mapping specification document can help identify problems before they happen
- The most common mapping scenarios you're likely to encounter, and…
- How the Pinnacle 21 platform can help you navigate the mapping process while saving time and effort!

By standardizing the mapping process, you'll maximize your chances of success and save time and effort on future trials.

Book a free no-obligation demo

About Certara

Certara accelerates medicines using proprietary biosimulation software, technology and services to transform traditional drug discovery and development. Its clients include more than 2,400 biopharmaceutical companies, academic institutions and regulatory agencies across 70 countries.

certara.com

CERTARA