

Pharma Practice

This practice project is an exploration of pharmaceutical data analysis techniques using sample datasets obtained from the Harvard Dataverse at <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/51B6NK>¹. The project focuses on applying Clinical Data Interchange Standards Consortium (CDISC) Study Data Tabulation Model (SDTM) standards to real-world clinical trial data, with the goal of creating Analysis Data Model (ADaM) datasets and generating statistical reports typical of those used in the pharmaceutical industry. This hands-on exercise serves as a foundation for developing practical skills in statistical programming within the pharma sector.

Load packages.

```
library(tidyverse)
library(metacore)
library(metatools)
library(pharmaversesdtm)
library(admiral)
library(xportr)
library(dplyr)
library(tidyr)
library(lubridate)
library(stringr)
```

Load meta data and SDTM data.

```
meta <-
  jsonlite::read_json("data/meta.json")

file_names <- list.files("data/", "*.csv", full.names = TRUE)

all_data <-
  file_names %>%
  map(read_csv) %>%
  set_names(str_remove(basename(file_names), ".csv"))
```

Based on the JSON metadata provided, here's a short description of each file:

- ae: This contains Adverse Events data, which likely includes details about any undesirable experiences associated with the use of a medical product in a patient.
- dm: This file contains Subject Demographics data, detailing information about the trial subjects such as age, sex, race, and other demographic details.
- eg: This file includes Electrocardiogram data, which are records of the electrical activity of the heart over a period of time.
- lb: This contains Laboratory Findings data, providing information from laboratory tests conducted during the study.
- mh: This file holds Medical History data, containing historical health information about the subjects enrolled in the clinical study.
- pe: This contains Physical Examination data, documenting the findings from physical exams conducted on study subjects.
- vs: This file contains Vital Signs data, which includes measurements like blood pressure, heart rate, body temperature, and respiratory rate.

Each file is in a comma-separated values format and was created for the purpose of demonstrating the use of CDISC-SDTM formatted datasets.

i Note

Does not seem like I need to create a metacore object as all data are in SDTM format... But if want to programatically manage and validate my data, maybe I should figure out how to create such a metacore object???

Create ADSL dataset

ADSL stands for Analysis Data Subject-Level Dataset. It's part of the ADaM (Analysis Data Model) standards provided by CDISC for use in statistical analysis related to clinical trials. The ADSL dataset contains one record per subject and includes key variables necessary for analysis, such as demographic information, treatment information, and other subject-level data. It's the foundational dataset used in many statistical analyses and is often required for regulatory submissions to agencies like the FDA.

i Note

No way around it. Gotta figure out how to create a metacore object. The meta data json file from the Harward site does not contain the required information to create a metacore object.

References

1. Emam, I. & CDISC. CDISC-SDTM sample study. (2019) doi:[10.7910/DVN/51B6NK](https://doi.org/10.7910/DVN/51B6NK).