**HIV Descriptive study protocol**

**Version history**

|  |  |  |
| --- | --- | --- |
| Date | Protocol version | Changes |
| Nov 11, 2018 | v0.5 | initial draft |
| July10, 2019 | v1.1 | basic covariates collection for HIV cohort |

# Study background

The study is part of a larger project titled Identification of Research Common Data Elements in HIV/AIDS using data science methods described at <https://github.com/lhncbc/CDE/blob/master/hiv/README.md>

Study is implemented as an R package.

# Study steps

The study consist of

1. Extracting data on HIV cohorts from multiple databases (Site Data Extraction)
2. Comparison of the data across sites (Centralized Processing)

# Site Data Extraction

After execution of the R study package, a set of files (in a .zip file) is transmitted to centralized processing (content of export folder). The extracted data is limited to aggregated data. No patient level information is included in the extract.

Study sites are encouraged to inspect the content of the export folder.

## Implementations

FeatureExtraction package is used to compute characteristics of the cohorts.

## Cohorts

cohorts.csv file contains a list of cohorts used by the package

# Centralized processing

Centralized processing consist of comparing datasets (from sites). The goal is to identify which data elements are common across datasets.