Package 'hybridEHR'

December 19, 2025

Type Package

Title Synthetic Hybrid Electronic Health Records Dataset Generator with COVID/CT Research Views

Version 0.1.0

Maintainer Dennis Boadu <doboadu@st.ug.edu.gh>

Description Tools to generate synthetic electronic health records including patients, encounters, vitals, labs, medications, procedures, and allergies, with optional COVID-19-focused and computed tomography (CT)-research views, and export them to comma separated values ('CSV'), 'SQLite', and 'Excel' formats for researchers and developers.

License MIT + file LICENSE

Encoding UTF-8

Depends R (>= 4.1.0)

Imports dplyr, tidyr, tibble, lubridate, jsonlite, openxlsx, DBI, RSQLite, magrittr

Suggests knitr, rmarkdown

RoxygenNote 7.3.3

VignetteBuilder knitr

NeedsCompilation no

Author Dennis Boadu [aut, cre], Isaac Osei [aut], Justice Appati [aut]

Repository CRAN

Date/Publication 2025-12-19 21:00:09 UTC

Contents

`export_hybrid_ehr_dataset` *Export a hybrid EHR dataset to disk*

Description

Export a hybrid EHR dataset to disk

Usage

```
export_hybrid_ehr_dataset(dataset, output_dir, verbose = TRUE)
```

Arguments

`dataset`	A list as returned by `generate_hybrid_ehr_dataset()`.
`output_dir`	Directory to write files into.
`verbose`	Logical; if TRUE, print messages.

Value

The output directory (invisibly).

`generate_hybrid_ehr` *High-level wrapper to generate and export a hybrid EHR dataset*

Description

High-level wrapper to generate and export a hybrid EHR dataset

Usage

```
generate_hybrid_ehr(
  n_patients = 500,
  n_sites = 3,
  covid_focused = TRUE,
  include_ct_links = FALSE,
  output_dir,
  seed = NULL,
  verbose = TRUE
)
```

Arguments

n_patients	Number of unique patients.
n_sites	Number of sites/hospitals to simulate.
covid_focused	Logical; if TRUE, use COVID-era encounter and lab patterns.
include_ct_links	Logical; if TRUE, add CT timing variables and a CT severity score in the CT research view.
output_dir	Directory for exported files.
seed	Optional integer used to set the random seed for reproducibility.
verbose	Logical; if TRUE, print progress messages to the console.

Value

A list with:

dataset The in-memory dataset list (as from generate_hybrid_ehr_dataset).

output_dir The output directory path where files were written.

A list containing:

dataset Generated dataset object

output_dir Path to exported files

Examples

```
ehr <- generate_hybrid_ehr_dataset(
  n_patients = 10,
  seed = 123,
  verbose = FALSE
)

export_hybrid_ehr_dataset(
  ehr,
  output_dir = tempdir(),
  verbose = FALSE
)
```

`generate_hybrid_ehr_dataset` *Generate synthetic hybrid EHR tables*

Description

Generate synthetic hybrid EHR tables

Usage

```
generate_hybrid_ehr_dataset(
  n_patients = 500,
  n_sites = 3,
  covid_focused = TRUE,
  include_ct_links = FALSE,
  seed = NULL,
  verbose = TRUE
)
```

Arguments

`n_patients`	Number of unique patients.
`n_sites`	Number of sites/hospitals to simulate.
`covid_focused`	Logical; if TRUE, use COVID-era encounter and lab patterns.
`include_ct_links`	Logical; if TRUE, add CT timing variables and a CT severity score in the CT research view.
`seed`	Optional integer used to set the random seed for reproducibility.
`verbose`	Logical; if TRUE, print progress messages to the console.

Value

A list with elements:

tables Named list of core EHR tables (patients, encounters, vitals, labs, medications, procedures, allergies).

research Named list with `ct_research_view` (if covid_focused) and `ml_flat_view` (aggregated ML-ready table).

metadata List of high-level generation settings and table metadata.

Examples

```
ehr <- generate_hybrid_ehr_dataset(
  n_patients = 10,
  n_sites = 2,
  covid_focused = TRUE,
  include_ct_links = FALSE,
  seed = 123,
  verbose = FALSE
)

names(ehr$tables)
head(ehr$tables$patients)
```

Index