Package: ofhsyn (via r-universe)

June 9, 2026

Type Package

Title Synthetic Our Future Health Data Generator

Version 0.1.1

Author Hannah Nicholls [aut, cre]

Maintainer Hannah Nicholls <h.l.nicholls@qmul.ac.uk>

Description Generates synthetic Our Future Health cohort datasets for method development, including participant, questionnaire, clinic measurements, outpatient, inpatient, emergency, mortality, primary care medication, and geography outputs. Supports reproducible generation with configurable cohort size and user-defined International Classification of Diseases, Tenth Revision (ICD-10), Office of Population Censuses and Surveys Classification of Interventions and Procedures, version 4 (OPCS-4), and British National Formulary (BNF) code pools.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.3.2

Depends R (>= 4.2.0)

Imports methods, utils, stats

Suggests testthat (>= 3.0.0), knitr, rmarkdown

VignetteBuilder knitr

Config/testthat/edition 3

NeedsCompilation no

Repository https://cran.r-universe.dev

Date/Publication 2026-06-09 15:40:08 UTC

RemoteUrl https://github.com/cran/ofhsyn

RemoteRef HEAD

RemoteSha fae282c36ef62cee18abf7e833a0e997ce9cfc34

Contents

`generate_ofh_cohort` *Generate Synthetic OFH Cohort Datasets*

Description

Generate linked synthetic health datasets for a configurable cohort.

Usage

```
generate_ofh_cohort(
  n = 5000,
  seed = 42,
  icd10 = NULL,
  icd10_file = NULL,
  opcs4 = NULL,
  opcs4_file = NULL,
  bnf_codes = NULL,
  bnf_codes_file = NULL,
  proportions = NULL,
  record_multipliers = NULL,
  code_config = NULL,
  save_csv = TRUE,
  return_objects = TRUE,
  output_dir = NULL
)
```

Arguments

`n`	Total synthetic cohort size.
`seed`	Random seed.
`icd10`	Optional named character vector of ICD-10 descriptions.
`icd10_file`	Optional path to a TXT/CSV file containing ICD-10 code and description pairs. TXT format should be tab-separated with code and description columns. CSV format should provide code and description columns.
`opcs4`	Optional named character vector of OPCS-4 descriptions.
`opcs4_file`	Optional path to a TXT/CSV file containing OPCS-4 code and description pairs. TXT format should be tab-separated with code and description columns. CSV format should provide code and description columns.

bnf_codes	Optional BNF input for primary care meds. Can be either a character vector of BNF codes or a data frame with columns for code, name, and formulation (optional strength).
bnf_codes_file	Optional path to a TXT/CSV file for BNF input. TXT supports one BNF code per line. CSV supports either code-only or structured medication rows containing code, name, and formulation (optional strength).
proportions	Optional named list of dataset-level coverage proportions. Names should match `names(ofh_default_proportions())`.
record_multipliers	Optional named list of multipliers for multi-record datasets. Names should match `names(ofh_default_record_multipliers())`.
code_config	Optional nested list overriding field-level code generation probabilities and pools. Structure should follow `ofh_default_code_config()`.
save_csv	Whether to write CSV outputs to disk.
return_objects	Whether to return generated data frames as an R object.
output_dir	Output directory when `save_csv = TRUE`.

Value

Named list of generated data frames when `return_objects = TRUE`; otherwise invisible `NULL`.

Acknowledgement

We extend our thanks to GitHub user **@icallumwebb** for contributing a bug fix that improved custom code handling.

Examples

```
out <- generate_ofh_cohort(n = 200, seed = 123, save_csv = FALSE, return_objects = TRUE)
names(out)
```

`generation_primitives` *Standalone Synthetic Generation Primitives*

Description

Utility functions for generating participant populations and event-level synthetic records.

Usage

```
generate_ofh_population(n = 1000, seed = 123)

add_inpatient_events(
  data,
  events_per_person = 5,
  icd10_codes = c("I210", "I500", "I639", "E110", "J440"),
```

```
  opcs4_codes = c("K401", "K451", "K561", "M011", "E033"),
  seed = 123
)

synthesize_drug_exposure(
  data,
  drug_list = c("0212000B0", "0601023A0"),
  seed = 123,
  mean_items_per_person = 2
)
```

Arguments

`data`	Input data frame containing a `pid` column.
`n`	Number of participants.
`seed`	Random seed.
`events_per_person`	Mean events per participant.
`icd10_codes`	ICD-10 code pool.
`opcs4_codes`	OPCS-4 code pool.
`drug_list`	Medication code pool.
`mean_items_per_person`	Mean prescription items per participant.

Value

Return value depends on the function called:

`generate_ofh_population()` Data frame with one row per participant and columns including `pid`, `sex`, and `birth_year`.

`add_inpatient_events()` Data frame of synthetic inpatient events with columns `pid`, `admidate`, `icd10`, and `opcs4`.

`synthesize_drug_exposure()` Data frame of synthetic primary-care medication records with participant IDs and prescribing/dispensing fields (for example `prescribedbnfcode`, `paidbnfcode`).

---

ofh_build_config *Configuration Helpers for OFH Generation*

Description

Helper functions that return default settings and compose full generation configuration lists.

Usage

```
ofh_default_proportions()
ofh_default_record_multipliers()
ofh_default_code_config()
ofh_build_config(
  n = 5000,
  proportions = ofh_default_proportions(),
  record_multipliers = ofh_default_record_multipliers(),
  code_config = list()
)
```

Arguments

`n`	Total cohort size.
`proportions`	Dataset proportions list.
`record_multipliers`	Record multiplier list for event datasets.
`code_config`	Optional code configuration overrides.

Value

Return value depends on the function called:

`ofh_default_proportions()` Named numeric list of dataset proportions in [0, 1].

`ofh_default_record_multipliers()` Named numeric list of multipliers for multi-record datasets.

`ofh_default_code_config()` Nested named list containing default code pools, weights, and generation controls by dataset.

`ofh_build_config()` Named list with `total_pid_count` (integer), `datasets` (nested list of dataset sizing settings), and `code_config` (merged code configuration list).

`OFHCohortSynthesizer` *Reference Class for OFH Cohort Generation*

Description

Reference class API for configuring and running synthetic cohort generation.

Usage

```
OFHCohortSynthesizer
```

Details

Create an instance with `OFHCohortSynthesizer$new(...)` and run generation via `$run_all(n = ...)`.

Value

A `ReferenceClass` generator object. Use `OFHCohortSynthesizer$new(...)` to create an instance. Instance methods return the instance invisibly for chaining where applicable, and `$run_all()` returns a named list of data frames when `return_objects = TRUE` (otherwise invisible `NULL`).

Examples

```
syn <- OFHCohortSynthesizer$new(project_root = ".", seed = 123)
out <- syn$run_all(n = 100, save_csv = FALSE, return_objects = TRUE)
```

Index