Package: XSRecencyX (via r-universe)

November 18, 2026

Type Package

Title HIV Incidence Estimation using Recency Testing Data with Population Adjustment

Version 0.1.0

Description Tools for estimating HIV incidence using cross-sectional recency testing data, adjusting for internal and external target populations and supporting subtype-specific parameters. The statistical methodology implemented builds on the framework described in Wang, Duerr, and Gao(2025) <doi:10.1002/sim.70216>.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.3.2

Depends R (>= 4.0)

Imports data.table, dplyr, geepack, purrr, magrittr, methods

NeedsCompilation no

Author Sirong Li [aut], Fei Gao [aut, cre] (ORCID: <https://orcid.org/0000-0001-6797-5468>), Marlena Bannick [aut] (ORCID: <https://orcid.org/0000-0001-6797-5978>)

Maintainer Fei Gao <fgao@fredhutch.org>

Config/pak/sysreqs libicu-dev

Repository https://feigao1.r-universe.dev

Date/Publication 2026-04-21 21:09:24 UTC

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

RemoteRef HEAD

RemoteSha 531ffb1fefa9701d60a190311dd17d0a3a0f3caf

Contents

cephia 2
estimate_incidence 4
estimate_weights_external 7
estimate_weights_internal 9
test.cross 10
test.target 11

Index **12**

`cephia` *CEPHIA Public-Use Dataset*

Description

A public-use dataset from the Consortium for the Evaluation and Performance of HIV Incidence Assays (CEPHIA).

Usage

```
cephia
```

Format

A data frame with 212831 rows and 38 variables:

assay Assay name.

cephia_panel CEPHIA panel identifier.

testing_laboratory Laboratory where testing was performed.

test_date Date of assay testing.

assay_result_field Field corresponding to assay result.

assay_result_value Numeric assay result value.

assay_result_method Method used to obtain assay result.

specific_result_identifier Specific result identifier.

generic_result_identifier Generic result identifier.

participant_identifier Unique participant identifier.

visit_identifier Visit identifier.

specimen_type Type of biological specimen.

hiv_status_at_visit HIV status at visit.

cohort_entry_hiv_status HIV status at cohort entry.

days_since_cohort_entry Days since cohort entry.

hiv_subtype HIV subtype classification.

hiv_subtype_confirmed Indicator whether subtype was confirmed.

country Country of participant.

sex Biological sex of participant.

age_in_years Age in years at visit.

eddi_interval_size Interval size for estimated date of detectable infection (EDDI).

days_since_eddi Days since estimated date of detectable infection.

days_since_ep_ddi Days since earliest possible date of detectable infection.

days_since_lp_ddi Days since latest possible date of detectable infection.

designated_as_elite_controller_at_visit Indicator for elite controller status at visit.

ever_designated_as_elite_controller Indicator whether participant was ever designated as elite controller.

treatment_naive_at_visit Indicator whether participant was treatment naive at visit.

on_treatment_at_visit Indicator whether participant was on treatment at visit.

first_treatment_episode Indicator for first treatment episode.

days_since_first_art Days since first antiretroviral therapy (ART).

days_since_current_art Days since current ART episode.

days_from_eddi_to_first_art Days from EDDI to first ART.

days_from_eddi_to_current_art Days from EDDI to current ART.

viral_load_closest_to_visit Viral load measurement closest to visit.

viral_load_date_offset_from_visit_date Offset between viral load date and visit date.

viral_load_type Type of viral load measurement.

viral_load_detectable Indicator whether viral load was detectable.

cd4_count_at_visit CD4 count at visit.

Details

The dataset was obtained from Zenodo (2025 release, version 2) and is redistributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) license.

The data are used internally by XSRecencyX for estimation of mean duration of recent infection (MDRI) and false recent rate (FRR) when these parameters are not supplied by the user.

Source

Grebe, E., et al. (2025). CEPHIA public use data. Zenodo. doi:10.5281/zenodo.17439895.

Distributed under the Creative Commons Attribution 4.0 International License (CC BY 4.0).

References

Facente, S.N., et al. (2020). Estimated Dates of Detectable Infection (EDDIs) as an improvement upon Fiebig staging for HIV infection dating. *Epidemiology and Infection*, 148:e53.

`estimate_incidence` *Estimate HIV Incidence from Recency Test Data with Population Adjustment*

Description

The function returns the estimated HIV incidence rate (cases per person-year) with optional adjustment to internal or external target populations using inverse probability weights. Supports subtype-specific parameters for recency test performance.

Usage

```
estimate_incidence(
  data,
  status_col,
  recency_col,
  covariates = NULL,
  target_data = NULL,
  target_col = NULL,
  subtype_col = NULL,
  recency.params = NULL,
  n_boot = NULL,
  seed = NULL,
  return_weights = FALSE,
  cephia_information_message = FALSE,
  assays = NULL,
  algorithm = NULL
)
```

Arguments

`data`	A data frame containing cross-sectional recency testing data. It must include HIV status and recency test results, and may optionally include covariates, a target group indicator and subtype label.
`status_col`	Character. Column name in `data` indicating HIV status (1 = positive, 0 = negative).
`recency_col`	Character. Column name in `data` indicating recency test result (1 = recent, 0 = non-recent, NA = missing).
`covariates`	Character vector. Vector of column names in `data` (and `target_data`), indicating covariates that may be heterogeneously distributed across populations and are used to compute adjustment weights.
`target_data`	Data frame (optional). A data frame containing covariates (and subtype if applicable) from the external target population. Required only when adjusting to an external population.

`target_col` Character (optional). Column name in data indicating inclusion in the internal target population data (1 = target, 0 = not target). Required only when estimating for an internal population.

`subtype_col` Character (optional). Column name in `data` (and `target_data`) indicating HIV subtype.

`recency.params` A named list with the following elements (element names must match those below, case-insensitive):

- `MDRI`: Numeric or numeric vector. Mean duration of recent infection (in days).
- `MDRI_CI`: Numeric or vector. Confidence interval(s) of MDRI (in days).
- `FRR`: Numeric or numeric vector. False recency rate.
- `FRR_CI`: Numeric or numeric vector. Confidence interval(s) of `FRR`.
- `T`: Numeric. Cut-off time (in years) used in the definition of `MDRI` and `FRR`.

Notes:

- MDRI and MDRI_CI must be specified in days. T must be specified in years. Internally, MDRI is converted to years before computing incidence.
- If `MDRI` is `NULL`, both `MDRI` and its confidence interval (`MDRI_CI`) will be estimated using the CEPHIA public-use dataset (Grebe et al., 2025).
- If `MDRI` is provided but `MDRI_CI` is `NULL`, `MDRI_CI` will be set equal to `MDRI` (i.e., assuming no variability). If `FRR` is `NULL`, `FRR` and its confidence interval (`FRR_CI`) will be set to zero. If `subtype_col` is provided, subtype-specific values for `MDRI` and `FRR` must be supplied, with each vector ordered to match the factor levels of `subtype_col`. The function does not automatically reorder subtype-specific parameters; users must ensure correct ordering.

`n_boot` Integer (optional). Number of bootstrap replicates for confidence intervals and variances.

`seed` Integer (optional). Seed for reproducibility.

`return_weights` Logical (optional). If `TRUE`, returns the weights used for estimation along with the incidence estimate.

`cephia_information_message` Logical (optional). If `TRUE`, prints informational messages related to `MDRI` and `FRR` estimation using the CEPHIA dataset.

`assays` Character vector (optional). Names of assays used in the recency testing algorithm. Default is `c("LAg-Sedia", "viral_load")`. Used only if `MDRI` is estimated internally. Suggested assays include:

- `"ArchitectAvidity"`
- `"Asante Visual"`
- `"Asante Electronic"`
- `"BED"`
- `"BioRadAvidity-Glasgow"`
- `"BioRadAvidity-CDC"`
- `"IDE-V3"`

- "Geenius"
- "LAg-Sedia"
- "LAg-Maxim"
- "LSVitros-Diluent"
- "VitrosAvidity"
- "viral load"
- "cd4"

algorithm Function(optional). Defines the recency indicator with arguments in the same order as the assays vector. E.g., if assays = c("LAg-Sedia", "viral_load"), you can have algorithm = function(l, v) ifelse(l < 1.5 & v > 1000, 1, 0). **Notes:** assays and algorithm must be provided together; otherwise, the function will stop with an error.

Details

This function estimates HIV incidence using cross-sectional recency testing data, optionally adjusting for differences between the observed sample and a specified target population. The target population can be:

- the same as the observed cross-sectional population by specifying (target_data = NULL and target_col = NULL),
- an internal subset of the cross-sectional population by specifying target_col,
- or a separate external population (e.g., for transportability applications) by specifying target_data.

Incidence is estimated using a weighted version of the adjusted cross-sectional incidence estimator as in Wang et al. (2025). Weights are derived via logistic regression to adjust for population heterogeneity in covariates. Subtype-specific MDRI and FRR parameters can be incorporated to improve estimation accuracy when recency test performance varies by HIV subtype. Specifically, the incidence is estimated by

$$\hat{\lambda}_{sub} = \sum_{j=1}^{J} \hat{\pi}_j \frac{\sum_{i=1}^{N} I(U_i = j) D_i (R_i - \omega_{T^*,j})}{\sum_{i=1}^{N} I(U_i = j)(1 - D_i)(\Omega_{T^*,j} - \omega_{T^*,j} T)}$$

where $\hat{\Omega}_{T^*,j}$ and $\hat{\omega}_{T^*,j}$ are the estimated mean duration of recent infection (MDRI) and false recent rate (FRR) for HIV subtype j,respectively. Bootstrapping is used to construct confidence interval for the incidence estimate. Uncertainty in MDRI/FRR is incorporated via their confidence intervals assuming lognormal distributions.

Value

A named list with the following elements:

- incidence: Point estimate of HIV incidence in the specified target population.
- se_incidence: Standard error of the incidence estimate based on bootstrap.
- ci_incidence: 95% confidence interval(s) of the incidence estimate.
- recency.params: a named list of recency test parameters, with specification in Arguments.
- weights: (Optional) A numeric vector of weights used in the point estimation, returned if return_weights = TRUE.

References

Wang, X., Duerr, A., & Gao, F. (2025). Addressing population heterogeneity for HIV incidence estimation based on recency test. *Statistics in Medicine*. doi:10.1002/sim.70216

Examples

```
## Example 1: Incidence estimation with full recency parameters

# Define covariates used in the model
covariates <- c("rInfection_pos", "Receptive",
                "Anal_nocondom", "College")

# Full recency parameters:
# MDRI, its 95% CI, FRR, its 95% CI, and time cutoff T
recency.params <- list(
  MDRI    = c(182, 186),                # MDRI (days)
  MDRI_CI = list(c(174, 189),
                 c(170, 198)),          # 95% CI for MDRI
  FRR     = c(0, 0.02),                 # False recent rate
  FRR_CI  = list(c(0, 0),
                 c(0.015, 0.03)),       # 95% CI for FRR
  T       = 2                           # Time cutoff (years)
)

# Run the estimator using observed recency status
estimate_incidence(
  data           = test.cross,
  target_data    = test.target,
  status_col     = "pos",
  recency_col    = "rpos",
  covariates     = covariates,
  recency.params = recency.params,
  subtype_col    = "Subtype",
  n_boot         = 3
)
```

`estimate_weights_external` *Estimate Weights for External Target Population*

Description

Computes weights for individuals in the cross-sectional HIV population when estimating HIV incidence for an external population.

Usage

```
estimate_weights_external(
  data,
  status_col,
  covariates,
  target_data,
  subtype_col = NULL
)
```

Arguments

data	A data frame containing cross-sectional recency testing data. It must include HIV status, recency test results, and covariates, and optionally subtype label.
status_col	Character. Column name in data indicating HIV status (1 = positive, 0 = negative).
covariates	Character vector. Vector of column names in data (and target_data) indicating covariates used to compute adjustment weights via logistic regression.
target_data	A data frame containing covariates (and subtype if applicable) from the external target population. Required only when estimating for a external population.
subtype_col	Character (optional). Column name in data (and target data) indicating HIV subtype. Required when using subtype-specific MDRI and FRR values.

Value

A numeric vector of estimated weights for each individual in the cross-sectional dataset.

Examples

```
## Example: external target population weighting

## Define covariates used for weighting
covariates <- c("rInfection_pos", "Receptive", "Anal_nocondom", "College")
## Estimate external weights
weights_ext <- estimate_weights_external(
  data        = test.cross,
  status_col  = "pos",
  covariates  = covariates,
  target_data = test.target,
  subtype_col = "Subtype"
)

## Inspect weights for different subtypes
unique(weights_ext)
```

`estimate_weights_internal` *Estimate Weights for Internal Target Population*

Description

Computes inverse probability weights for individuals in the cross-sectional HIV population when estimating HIV incidence for an internal population.

Usage

```
estimate_weights_internal(
  data,
  status_col,
  covariates,
  target_col,
  subtype_col = NULL
)
```

Arguments

`data`	A data frame containing cross-sectional recency testing data. It must include HIV status, recency test results, covariates, target group indicator, and optionally subtype label.
`status_col`	Character. Column name in `data` indicating HIV status (1 = positive, 0 = negative).
`covariates`	Character vector. Vector of column names in `data` indicating covariates used to compute adjustment weights via logistic regression.
`target_col`	Character. Column name in `data` indicating inclusion in the internal target population data (1 = target, 0 = not target).
`subtype_col`	Character (optional). Column name in `data` (and `target data`) indicating HIV subtype. Required when using subtype-specific MDRI and FRR values.

Value

A numeric vector of estimated weights for each individual in the cross-sectional dataset.

Examples

```
## Example: internal population weighting

## Define covariates used for weighting
covariates <- c("rInfection_pos", "Receptive", "Anal_nocondom", "College")
## Estimate external weights
weights_int <- estimate_weights_internal(
  data        = test.cross,
  status_col  = "pos",
```

```
  covariates  = covariates,
  target_col  = "intrial",
  subtype_col = "Subtype"
)

## Inspect the weights for different subtypes
unique(weights_int)
```

`test.cross` *Cross-sectional recency testing example dataset*

Description

A simulated dataset generated to illustrate cross-sectional HIV incidence estimation with subtype-specific recency parameters and population adjustment.

Usage

```
test.cross
```

Format

A data frame with 5000 rows and 9 variables:

pos Binary indicator of HIV infection status (1 = positive, 0 = negative).

rpos Binary indicator of recent infection among HIV-positive individuals (1 = recent, 0 = non-recent).

sim Binary simulation indicator used for internal data generation.

rInfection_pos Binary indicator of rectal infection.

Receptive Binary indicator of receptive anal intercourse.

Anal_nocondom Binary indicator of anal intercourse without condom use.

College Binary indicator of postsecondary education.

Subtype Factor indicating HIV subtype classification.

intrial Binary indicator of enrollment in the target (trial) population.

Details

The dataset is intended solely for demonstration and testing purposes.

Source

Simulated data generated under assumptions described in Wang, Duerr, and Gao (2025) for methodological illustration.

`test.target` *Target example dataset*

Description

A simulated cohort dataset representing an external target population used for evaluating transportability and population-adjusted HIV incidence estimation.

Usage

```
test.target
```

Format

A data frame with 2500 rows and 8 variables:

itime_trial Observed follow-up time (in years) in the target cohort.

event Binary indicator of HIV seroconversion during follow-up (1 = event, 0 = censored).

sim Binary simulation indicator used for internal data generation.

rInfection_pos Binary indicator of rectal infection.

Receptive Binary indicator of receptive anal intercourse.

Anal_nocondom Binary indicator of anal intercourse without condom use.

College Binary indicator of postsecondary education.

Subtype Factor indicating HIV subtype classification.

Details

The dataset includes follow-up time and event indicators, along with baseline covariates and subtype information. It is intended solely for methodological illustration and testing purposes.

This dataset can be used as an external target population when estimating inverse probability weights to transport cross-sectional incidence estimates.

Source

Simulated data generated under assumptions described in Wang, Duerr, and Gao (2025) for methodological illustration.

Index

∗ **datasets**
cephia, 2
test.cross, 10
test.target, 11

cephia, 2

estimate_incidence, 4
estimate_weights_external, 7
estimate_weights_internal, 9

test.cross, 10
test.target, 11