Package: PERSUADE (via r-universe)

November 2, 2026

Title Parametric Survival Model Selection for Decision-Analytic Models

Version 0.1.2

Description Provides a standardized framework to support the selection and evaluation of parametric survival models for time-to-event data. Includes tools for visualizing survival data, checking proportional hazards assumptions (Grambsch and Therneau, 1994, <doi:10.1093/biomet/81.3.515>), comparing parametric (Ishak and colleagues, 2013, <doi:10.1007/s40273-013-0064-3>), spline (Royston and Parmar, 2002, <doi:10.1002/sim.1203>) and cure models, examining hazard functions, and evaluating model extrapolation. Methods are consistent with recommendations in the NICE Decision Support Unit Technical Support Documents (14 and 21 <https://sheffield.ac.uk/nice-dsu/tsds/survival-analysis>). Results are structured to facilitate integration into decision-analytic models, and reports can be generated with 'rmarkdown'. The package builds on existing tools including 'flexsurv' (Jackson, 2016, <doi:10.18637/jss.v070.i08>)) and 'flexsurvcure' for estimating cure models.

License GPL (>= 3)

URL https://github.com/Bram-R/PERSUADE

BugReports https://github.com/Bram-R/PERSUADE/issues

Depends R (>= 4.1.0)

Imports data.table, flexsurv, flexsurvcure, ggplot2, muhaz, rmarkdown, rms, sft, stats, survival, survminer

Suggests kableExtra, knitr, testthat (>= 3.1.2), utils

VignetteBuilder knitr

Config/testthat/edition 3

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.2

Config/pak/sysreqs
cmake libgmp3-dev make libicu-dev libjpeg-dev libpng-dev libuv1-dev libxml2-dev libssl-dev

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

Date/Publication 2026-11-02 16:31:58 UTC

RemoteUrl https://github.com/bram-r/persuade

RemoteRef HEAD

RemoteSha 7be0e3668b7c5b6e2f05fef526147fcbd8f66667

Contents

f_cum_hazard . . . 3
f_generate_report . . . 4
f_get_excel_template . . . 6
f_hazard . . . 7
f_PERSUADE . . . 8
f_plot_cure_surv_extrap . . . 9
f_plot_cure_surv_model . . . 10
f_plot_diag_cure_surv_model . . . 11
f_plot_diag_param_surv_model . . . 12
f_plot_diag_spline_surv_model . . . 13
f_plot_hazard_cure_extrap . . . 14
f_plot_hazard_parametric_extrap . . . 15
f_plot_hazard_spline_extrap . . . 16
f_plot_hazard_with_models . . . 17
f_plot_km_survival . . . 18
f_plot_km_survival_base . . . 19
f_plot_log_cumhaz . . . 20
f_plot_param_surv_extrap . . . 21
f_plot_param_surv_model . . . 22
f_plot_schoenfeld_residuals . . . 23
f_plot_smoothed_hazard . . . 24
f_plot_spline_surv_extrap . . . 25
f_plot_spline_surv_model . . . 26
f_plot_tp_cure_surv_extrap . . . 27
f_plot_tp_cure_surv_model . . . 28
f_plot_tp_param_surv_extrap . . . 29
f_plot_tp_param_surv_model . . . 30
f_plot_tp_spline_surv_extrap . . . 31
f_plot_tp_spline_surv_model . . . 32
f_summary . . . 33
f_surv_model . . . 34
f_surv_model_excel . . . 35
f_surv_model_pred . . . 36
f_surv_model_pred_gr . . . 38
f_surv_model_pred_tp_gr . . . 39

f_tp 41
nsclc_ipd 42
nsclc_trials 56
plot.PERSUADE 59
print.PERSUADE 60
summary.PERSUADE 61

Index **62**

`f_cum_hazard` *Calculate Cumulative Hazard Estimates*

Description

Computes cumulative hazard estimates for up to three groups along with variance and confidence intervals, using the **estimateNAH** package.

Usage

```
f_cum_hazard(years, status, group, ngroups, time_pred, time_unit)
```

Arguments

`years`	Numeric vector of time-to-event data.
`status`	Numeric vector indicating event occurrence (1 = event, 0 = censoring).
`group`	Factor indicating group membership.
`ngroups`	Integer. Number of groups (1-3).
`time_pred`	Numeric vector of prediction times.
`time_unit`	Numeric. Time unit length for scaling.

Value

A data frame with columns:

- `group`: Group identifier.
- `time`: Prediction times.
- `H`: Cumulative hazard values.
- `var`: Variance estimates.
- `H_upper`, `H_lower`: 95% confidence interval bounds.
- `H_delta`, `H_upper_delta`, `H_lower_delta`: Differences between time steps.

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
f_cum_hazard(
  years = years,
  status = status,
  group = group,
  ngroups = nlevels(group),
  time_pred = seq(0, 5000, 365.25),
  time_unit = 30
)
```

`f_generate_report` *Generate PDF Report for a PERSUADE Analysis*

Description

Save the PERSUADE object and render a PDF report using the bundled `PERSUADE_output.Rmd` template, or a user-specified template.

Usage

```
f_generate_report(
  PERSUADE,
  output_dir = NULL,
  template_dir = NULL,
  open = FALSE
)
```

Arguments

`PERSUADE`	A PERSUADE object returned by `f_PERSUADE()`.
`output_dir`	Character string giving the directory to copy the function output to. If `NULL` (the default), the function uses: `file.path(tempdir(), paste0(PERSUADE$name, "_output"))`. Change `tempdir()` into `getwd()` for copying to working directory.
`template_dir`	Optional character string giving the full path to an Rmd template. If `NULL` (the default), the function looks for `PERSUADE_output.Rmd` within the package installation directory.
`open`	Logical. Whether to browse the generated file.

Details

The default R markdown file `PERSUADE_output.Rmd` is stored within the package under `inst/rmd/`. Figures are written to a subdirectory `Images/` inside the output folder, and the knit environment is initialised with the supplied `PERSUADE` object. Supplying a custom `template_dir` allows alternative report formats to be used, and simplifies testing. This function requires the following suggested packages: **knitr**, **kableExtra**, and **rmarkdown**. If not installed, the function will throw an error.

Value

A length-1 character string giving the absolute path to the generated PDF, returned invisibly.

See Also

`f_PERSUADE()`

Examples

```
## Not run:  # Requires LaTeX to be installed
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
PERSUADE <- f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = FALSE,
  cure_mod = FALSE,
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
# Copy output to temporary directory
# (change `tempdir()` into `getwd()`
# for copying to working directory)
f_generate_report(
  PERSUADE,
  output_dir = file.path(
  tempdir(), paste0(PERSUADE$name, "_output")
  ),
  template_dir = NULL
)

## End(Not run)
```

`f_get_excel_template` *Copy Excel Template for Model Parameters*

Description

Copy the bundled Excel template `PERSUADE_Excel_template.xltx` to a user-specified directory. This template provides a convenient structure for transferring survival model outputs from **PERSUADE** into health economic models.

Usage

```
f_get_excel_template(output_dir = NULL)
```

Arguments

`output_dir`	Character string giving the directory to copy the template to. If `NULL` (the default), the function uses: `tempdir()`. Change `tempdir()` into `getwd()` for copying to working directory.

Details

The default Excel file `PERSUADE_Excel_template.xltx` is stored within the package under `inst/excel_template/`. This function locates the installed file via `system.file()` and copies it into the requested directory. If a file with the same name already exists at the destination, it will be overwritten.

The Excel template provides a standardized format for entering parametric survival model parameters, making it easier to use PERSUADE outputs in downstream decision-analytic models. Users may adapt the template as needed for their specific workflows.

Value

A length-1 character string giving the absolute path to the copied template file, returned invisibly.

See Also

`f_generate_report()`, `system.file()`

Examples

```
# Copy output to temporary directory
# (change `tempdir()` into `getwd()` for
# copying to working directory)
f_get_excel_template(
output_dir = file.path(tempdir(), paste0("BC_OS", "_output"))
)
```

`f_hazard` *Calculate Smoothed Hazard Estimates*

Description

Computes smoothed hazard estimates for up to three groups using the **muhaz** package.

Usage

```
f_hazard(years, status, group, ngroups)
```

Arguments

`years`	Numeric vector of time-to-event data.
`status`	Numeric vector indicating event occurrence (1 = event, 0 = censoring).
`group`	Factor indicating group membership.
`ngroups`	Integer. Number of groups (1-3).

Value

A list with elements:

- `hazards`: List of hazard objects (one per group).
- `names`: Vector of group identifiers for hazard values.
- `max`: Data frame with maximum time and hazard values.

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
f_hazard(
  years = years,
  status = status,
  group = group,
  ngroups = nlevels(group)
)
```

f_PERSUADE *Main PERSUADE Function*

Description

Executes the PERSUADE workflow for parametric survival analysis, including Kaplan-Meier, parametric, spline, and cure models. Produces outputs for visualization, prediction, and Excel export.

Usage

```
f_PERSUADE(
  name = "no_name",
  years,
  status,
  group,
  strata = FALSE,
  spline_mod = FALSE,
  cure_mod = FALSE,
  cure_link = "logistic",
  time_unit,
  time_horizon,
  time_pred_surv_table
)
```

Arguments

name	Character. Name identifier for the analysis (default: "no_name").
years	Numeric vector of time-to-event data.
status	Numeric vector indicating event occurrence (1 = event, 0 = censoring).
group	Factor indicating group membership.
strata	Logical. Whether to stratify models by group.
spline_mod	Logical. Whether spline models should be fitted.
cure_mod	Logical. Whether cure models should be fitted.
cure_link	Character string specifying the link function for cure models ("logistic", "loglog", "identity", "probit"; default = "logistic").
time_unit	Numeric. The unit of time for annualization.
time_horizon	Numeric. The maximum prediction time horizon.
time_pred_surv_table	Numeric vector of time points for survival table predictions.

Details

The workflow proceeds in three main stages:

1. Observed data (Kaplan-Meier, hazards, Cox regression).
2. Parametric, spline, and cure model fitting.
3. Prediction and export of results.

Value

A list of class "PERSUADE" containing:

- input: Input arguments used in the analysis.
- surv_obs: Observed survival results (Kaplan-Meier, hazards, Cox model).
- surv_model: Fitted parametric/spline/cure models.
- surv_pred: Model predictions.
- surv_model_excel: Excel-ready parameter table.
- misc: Auxiliary results (labels, number of groups, etc.).

See Also

f_hazard(), f_cum_hazard(), f_surv_model()

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = FALSE,
  cure_mod = FALSE,
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
```

---

f_plot_cure_surv_extrap

Plot Extrapolated Cure Survival Models per Group

Description

Plot Kaplan-Meier curves per group with shaded confidence bands and overlay fitted cure survival models (Weibull, log-normal, log-logistic; mixture and non-mixture forms) extrapolated to the analysis time horizon. Runs only when PERSUADE$input$cure_mod is TRUE.

Usage

```
f_plot_cure_surv_extrap(PERSUADE)
```

Arguments

PERSUADE	A PERSUADE object returned by f_PERSUADE().

Value

Invisibly returns NULL. The function draws one or more base R plots as side effects.

See Also

f_PERSUADE()

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
PERSUADE <- f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = FALSE,
  cure_mod = TRUE,
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
f_plot_cure_surv_extrap(PERSUADE)
```

---

f_plot_cure_surv_model

Cure Survival Model Overlay

Description

Overlays a fitted cure survival model on KM curves, including shaded KM confidence bands per group.

Usage

```
f_plot_cure_surv_model(PERSUADE, model_index = 1)
```

Arguments

PERSUADE	A PERSUADE object created by f_PERSUADE().
model_index	Integer. Index of the cure model in PERSUADE$surv_pred$model$cure.

Value

A base R plot of KM curves with cure model overlays.

See Also

f_PERSUADE()

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
PERSUADE <- f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = FALSE,
  cure_mod = TRUE,
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
f_plot_cure_surv_model(PERSUADE, model_index = 1)
```

---

f_plot_diag_cure_surv_model

Diagnostic Plot for Cure Survival Models

Description

Produces diagnostic plots for mixture and non-mixture cure survival models, using transformations depending on the underlying distribution (Weibull, log-normal, log-logistic).

Usage

```
f_plot_diag_cure_surv_model(PERSUADE, model_index = 1)
```

Arguments

PERSUADE	A PERSUADE object created by f_PERSUADE().
model_index	Integer. Index of the cure model in PERSUADE$surv_pred$model$cure.

Value

A base R diagnostic plot for the selected cure survival model.

See Also

f_PERSUADE()

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
PERSUADE <- f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = FALSE,
  cure_mod = TRUE,
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
f_plot_diag_cure_surv_model(PERSUADE, model_index = 1)
```

---

f_plot_diag_param_surv_model

Diagnostic Plot for Parametric Survival Models

Description

Produces diagnostic plots for standard parametric survival models, using appropriate transformations depending on the model family (exponential, Weibull, Gompertz, log-normal, log-logistic, gamma, generalized gamma).

Usage

```
f_plot_diag_param_surv_model(PERSUADE, model_index = 1)
```

Arguments

PERSUADE	A PERSUADE object created by f_PERSUADE().
model_index	Integer. Index of the parametric model in PERSUADE$surv_pred$model.

Value

A base R diagnostic plot for the selected parametric survival model.

See Also

f_PERSUADE()

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
PERSUADE <- f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = FALSE,
  cure_mod = FALSE,
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
f_plot_diag_param_surv_model(PERSUADE, model_index = 1)
```

---

f_plot_diag_spline_surv_model

Diagnostic Plot for Spline Survival Models

Description

Produces diagnostic plots for spline-based survival models, using log-time transformations adapted to hazard, odds, or normal scales depending on the spline model type.

Usage

```
f_plot_diag_spline_surv_model(PERSUADE, model_index = 1)
```

Arguments

PERSUADE	A PERSUADE object created by f_PERSUADE().
model_index	Integer. Index of the spline model in PERSUADE$surv_pred$model$spline.

Value

A base R diagnostic plot for the selected spline-based survival model.

See Also

f_PERSUADE()

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
PERSUADE <- f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = TRUE,
  cure_mod = FALSE,
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
f_plot_diag_spline_surv_model(PERSUADE, model_index = 1)
```

---

f_plot_hazard_cure_extrap

Plot Extrapolated Hazard Functions (Cure Models)

Description

Plot observed smoothed hazard rates per group and overlay extrapolated hazard functions from all fitted cure survival models (mixture and non-mixture). Runs only when `PERSUADE$input$cure_mod` is `TRUE`.

Usage

```
f_plot_hazard_cure_extrap(PERSUADE)
```

Arguments

`PERSUADE`	A PERSUADE object returned by `f_PERSUADE()`.

Value

Invisibly returns `NULL`. The function draws one or more base R plots as side effects.

See Also

`f_PERSUADE()`

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
PERSUADE <- f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = FALSE,
  cure_mod = TRUE,
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
f_plot_hazard_cure_extrap(PERSUADE)
```

---

f_plot_hazard_parametric_extrap

Plot Extrapolated Hazard Functions (Parametric Models)

Description

Plot observed smoothed hazard rates per group and overlay extrapolated hazard functions from all fitted parametric survival models.

Usage

```
f_plot_hazard_parametric_extrap(PERSUADE)
```

Arguments

PERSUADE	A PERSUADE object returned by `f_PERSUADE()`.

Value

Invisibly returns `NULL`. The function draws one or more base R plots as side effects.

See Also

`f_PERSUADE()`

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
PERSUADE <- f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = FALSE,
  cure_mod = FALSE,
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
f_plot_hazard_parametric_extrap(PERSUADE)
```

---

f_plot_hazard_spline_extrap

Plot Extrapolated Hazard Functions (Spline Models)

Description

Plot observed smoothed hazard rates per group and overlay extrapolated hazard functions from all fitted spline survival models. Runs only when `PERSUADE$input$spline_mod` is `TRUE`.

Usage

```
f_plot_hazard_spline_extrap(PERSUADE)
```

Arguments

`PERSUADE`	A PERSUADE object returned by `f_PERSUADE()`.

Value

Invisibly returns `NULL`. The function draws one or more base R plots as side effects.

See Also

`f_PERSUADE()`

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
PERSUADE <- f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = TRUE,
  cure_mod = FALSE,
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
f_plot_hazard_spline_extrap(PERSUADE)
```

---

f_plot_hazard_with_models

Hazard Plot with Model Overlays

Description

Plots observed smoothed hazard estimates together with hazard predictions from parametric, spline, and cure survival models (if fitted).

Usage

```
f_plot_hazard_with_models(PERSUADE)
```

Arguments

PERSUADE	A PERSUADE object created by f_PERSUADE().

Value

A series of base R plots, one per group, with hazard overlays by model family.

See Also

f_PERSUADE()

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
PERSUADE <- f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = FALSE,
  cure_mod = FALSE,
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
f_plot_hazard_with_models(PERSUADE)
```

---

f_plot_km_survival *Plot Kaplan-Meier Survival Curves (ggsurvplot)*

Description

Generates Kaplan-Meier survival plots from a PERSUADE object using `survminer::ggsurvplot()`, automatically adapting to the number of groups.

Usage

```
f_plot_km_survival(PERSUADE)
```

Arguments

PERSUADE	A PERSUADE object created by `f_PERSUADE()`.

Value

A `ggsurvplot` object with KM curves, risk table, CI bands, and optional censor marks.

See Also

`f_PERSUADE()`

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
PERSUADE <- f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = FALSE,
  cure_mod = FALSE,
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
f_plot_km_survival(PERSUADE)
```

---

f_plot_km_survival_base

Plot Kaplan-Meier Survival Curves (Base R)

Description

Generates Kaplan-Meier survival plots from a PERSUADE object using base R graphics, with shaded confidence intervals and group-specific legends.

Usage

```
f_plot_km_survival_base(PERSUADE)
```

Arguments

PERSUADE	A PERSUADE object created by f_PERSUADE().

Value

A base R plot.

See Also

f_PERSUADE()

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
PERSUADE <- f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = FALSE,
  cure_mod = FALSE,
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
f_plot_km_survival_base(PERSUADE)
```

---

f_plot_log_cumhaz *Log-Log Survival Diagnostic Plot*

Description

Creates a log(-log(S(t))) vs log(time) plot to visually assess proportional hazards.

Usage

```
f_plot_log_cumhaz(PERSUADE)
```

Arguments

PERSUADE	A PERSUADE object created by f_PERSUADE().

Value

A base R plot showing ln(-ln(S(t))) against ln(time).

See Also

f_PERSUADE()

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
PERSUADE <- f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = FALSE,
  cure_mod = FALSE,
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
f_plot_log_cumhaz(PERSUADE)
```

---

f_plot_param_surv_extrap

Plot Extrapolated Parametric Survival Models per Group

Description

Plot Kaplan-Meier curves per group with shaded confidence bands and overlay fitted parametric survival models (Exponential, Weibull, Gompertz, log-normal, log-logistic, Gamma, generalized Gamma) extrapolated to the analysis time horizon.

Usage

```
f_plot_param_surv_extrap(PERSUADE)
```

Arguments

PERSUADE	A PERSUADE object returned by f_PERSUADE().

Value

Invisibly returns NULL. The function draws one or more base R plots as side effects.

See Also

f_PERSUADE()

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
PERSUADE <- f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = FALSE,
  cure_mod = FALSE,
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
f_plot_param_surv_extrap(PERSUADE)
```

---

f_plot_param_surv_model

Parametric Survival Model Overlay

Description

Overlays a fitted parametric survival model on top of KM curves, including shaded KM confidence bands per group.

Usage

```
f_plot_param_surv_model(PERSUADE, model_index = 1)
```

Arguments

PERSUADE	A PERSUADE object created by f_PERSUADE().
model_index	Integer. Index of the parametric model in PERSUADE$surv_pred$model.

Value

A base R plot of KM curves with parametric model overlays.

See Also

f_PERSUADE()

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
PERSUADE <- f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = FALSE,
  cure_mod = FALSE,
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
f_plot_param_surv_model(PERSUADE, model_index = 1)
```

---

f_plot_schoenfeld_residuals

Schoenfeld Residuals Plot

Description

Produces scaled Schoenfeld residual plots with fitted regression lines to evaluate Cox proportional hazards assumptions.

Usage

```
f_plot_schoenfeld_residuals(PERSUADE)
```

Arguments

PERSUADE	A PERSUADE object created by f_PERSUADE().

Value

One or more base R plots, one per group comparison.

See Also

f_PERSUADE()

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
PERSUADE <- f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = FALSE,
  cure_mod = FALSE,
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
f_plot_schoenfeld_residuals(PERSUADE)
```

---

f_plot_smoothed_hazard

Smoothed Hazard Function Plot

Description

Plots smoothed hazard estimates for each group in the PERSUADE object.

Usage

```
f_plot_smoothed_hazard(PERSUADE)
```

Arguments

PERSUADE	A PERSUADE object created by f_PERSUADE().

Value

A base R plot of smoothed hazards by group.

See Also

f_PERSUADE()

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
PERSUADE <- f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = FALSE,
  cure_mod = FALSE,
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
f_plot_smoothed_hazard(PERSUADE)
```

---

f_plot_spline_surv_extrap

Plot Extrapolated Spline Survival Models per Group

Description

Plot Kaplan-Meier curves per group with shaded confidence bands and overlay fitted spline survival models (hazard, odds, normal scales) extrapolated to the analysis time horizon. Runs only when `PERSUADE$input$spline_mod` is `TRUE`.

Usage

```
f_plot_spline_surv_extrap(PERSUADE)
```

Arguments

`PERSUADE`	A PERSUADE object returned by `f_PERSUADE()`.

Value

Invisibly returns `NULL`. The function draws one or more base R plots as side effects.

See Also

`f_PERSUADE()`

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
PERSUADE <- f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = TRUE,
  cure_mod = FALSE,
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
f_plot_spline_surv_extrap(PERSUADE)
```

---

f_plot_spline_surv_model

Spline Survival Model Overlay

Description

Overlays a spline-based survival model on KM curves, including shaded KM confidence bands and vertical lines for knot positions.

Usage

```
f_plot_spline_surv_model(PERSUADE, model_index = 1)
```

Arguments

PERSUADE	A PERSUADE object created by f_PERSUADE().
model_index	Integer. Index of the spline model in PERSUADE$surv_pred$model$spline.

Value

A base R plot of KM curves with spline model overlays and knots.

See Also

f_PERSUADE()

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
PERSUADE <- f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = TRUE,
  cure_mod = FALSE,
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
f_plot_spline_surv_model(PERSUADE, model_index = 1)
```

---

f_plot_tp_cure_surv_extrap

Plot Extrapolated Annual Transition Probabilities (Cure Models)

Description

Plot smoothed observed annual transition probabilities with shaded confidence intervals and overlay predictions from all fitted cure survival models. Runs only when `PERSUADE$input$cure_mod` is `TRUE`.

Usage

```
f_plot_tp_cure_surv_extrap(PERSUADE)
```

Arguments

`PERSUADE`	A PERSUADE object returned by `f_PERSUADE()`.

Value

Invisibly returns `NULL`. The function draws one or more base R plots as side effects.

See Also

`f_PERSUADE()`

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
PERSUADE <- f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = FALSE,
  cure_mod = TRUE,
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
f_plot_tp_cure_surv_extrap(PERSUADE)
```

---

f_plot_tp_cure_surv_model

Plot Annual Transition Probabilities for Cure Survival Models

---

Description

Plot smoothed observed annual transition probabilities with shaded confidence intervals, overlaid with predictions from a selected cure survival model (mixture or non-mixture; Weibull, log-normal, or log-logistic).

Usage

```
f_plot_tp_cure_surv_model(PERSUADE, model_index = 1)
```

Arguments

PERSUADE	A PERSUADE object returned by f_PERSUADE().
model_index	Integer index selecting the cure model within PERSUADE$surv_model$cure_models (1-based). Defaults to 1.

Value

Invisibly returns NULL. The function draws a base R plot as a side effect.

See Also

f_PERSUADE()

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
PERSUADE <- f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = FALSE,
  cure_mod = TRUE,
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
f_plot_tp_cure_surv_model(PERSUADE, model_index = 1)
```

---

f_plot_tp_param_surv_extrap

Plot Extrapolated Annual Transition Probabilities (Parametric Models)

Description

Plot smoothed observed annual transition probabilities with shaded confidence intervals and overlay predictions from all fitted parametric survival models.

Usage

```
f_plot_tp_param_surv_extrap(PERSUADE)
```

Arguments

PERSUADE	A PERSUADE object returned by f_PERSUADE().

Value

Invisibly returns NULL. The function draws one or more base R plots as side effects.

See Also

f_PERSUADE()

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
PERSUADE <- f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = FALSE,
  cure_mod = FALSE,
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
f_plot_tp_param_surv_extrap(PERSUADE)
```

---

f_plot_tp_param_surv_model

Plot Annual Transition Probabilities for Parametric Survival Models

---

Description

Plot smoothed observed annual transition probabilities alongside model-predicted probabilities for a selected parametric model, with shaded confidence intervals per group.

Usage

```
f_plot_tp_param_surv_model(PERSUADE, model_index = 1)
```

Arguments

PERSUADE	A PERSUADE object returned by f_PERSUADE().
model_index	Integer index selecting the parametric model within PERSUADE$surv_model$param_models (1-based). Defaults to 1.

Value

Invisibly returns NULL. The function draws a base R plot as a side effect.

See Also

f_PERSUADE()

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
PERSUADE <- f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = FALSE,
  cure_mod = FALSE,
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
f_plot_tp_param_surv_model(PERSUADE, model_index = 1)
```

---

f_plot_tp_spline_surv_extrap

Plot Extrapolated Annual Transition Probabilities (Spline Models)

Description

Plot smoothed observed annual transition probabilities with shaded confidence intervals and overlay predictions from all fitted spline survival models. Runs only when `PERSUADE$input$spline_mod` is `TRUE`.

Usage

```
f_plot_tp_spline_surv_extrap(PERSUADE)
```

Arguments

`PERSUADE`	A PERSUADE object returned by `f_PERSUADE()`.

Value

Invisibly returns `NULL`. The function draws one or more base R plots as side effects.

See Also

`f_PERSUADE()`

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
PERSUADE <- f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = TRUE,
  cure_mod = FALSE,
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
f_plot_tp_spline_surv_extrap(PERSUADE)
```

---

f_plot_tp_spline_surv_model

Plot Annual Transition Probabilities for Spline Survival Models

---

Description

Plot smoothed observed annual transition probabilities together with predictions from a selected spline survival model (hazard/odds/normal scale), including shaded confidence intervals and vertical lines for spline knots.

Usage

```
f_plot_tp_spline_surv_model(PERSUADE, model_index = 1)
```

Arguments

PERSUADE	A PERSUADE object returned by f_PERSUADE().
model_index	Integer index selecting the spline model within PERSUADE$surv_model$spline_models (1-based). Defaults to 1.

Value

Invisibly returns NULL. The function draws a base R plot as a side effect.

See Also

f_PERSUADE()

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
PERSUADE <- f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = TRUE,
  cure_mod = FALSE,
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
f_plot_tp_spline_surv_model(PERSUADE, model_index = 1)
```

---

f_summary *Compute Summary Statistics for Numeric Variables*

Description

Compute descriptive statistics for each numeric variable in a data frame: mean, standard deviation, minimum, first quartile (Q1), median, third quartile (Q3), maximum, and interquartile range (IQR). Results are rounded to three decimals.

Usage

```
f_summary(df)
```

Arguments

df	A data frame; numeric columns are summarized.

Value

A data frame (one row per variable) with columns: `Mean`, `Std.Dev`, `Min`, `Q1`, `Median`, `Q3`, `Max`, `IQR`.

Examples

```
f_summary(mtcars)
```

`f_surv_model` *Fit Parametric Survival Models*

Description

Fits standard parametric models, spline models, and cure models using the **flexsurv** package.

Usage

```
f_surv_model(
  years,
  status,
  group,
  strata,
  ngroups,
  form,
  spline_mod,
  cure_mod,
  cure_link,
  group_names
)
```

Arguments

`years`	Numeric vector of time-to-event data.
`status`	Numeric vector indicating event occurrence (1 = event, 0 = censoring).
`group`	Factor indicating group membership.
`strata`	Logical. Whether to stratify models by group.
`ngroups`	Integer. Number of groups.
`form`	A survival model formula (e.g., `Surv(years, status) ~ group`).
`spline_mod`	Logical. Whether spline models should be fitted.
`cure_mod`	Logical. Whether cure models should be fitted.
`cure_link`	Character string specifying the link function for cure models ("logistic", "loglog", "identity", "probit"; default = "logistic").
`group_names`	Character vector of group labels (for cure fractions).

Details

Models fitted include Exponential, Weibull, Gompertz, Log-normal, Log-logistic, Gamma, Generalised Gamma. Optional spline models (1-3 knots, scales: hazard, odds, normal) and cure models (Weibull, Log-normal, Log-logistic with logistic/probit/etc. link).

Value

A list containing:

- param_models, param_ic: Parametric models and information criteria.
- spline_models, spline_ic: Spline models and IC (if fitted).
- cure_models, cure_ic: Cure models and IC (if fitted).

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
form <- stats::as.formula(survival::Surv(years, status) ~ group)
f_surv_model(
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  ngroups = nlevels(group),
  form = form,
  spline_mod = FALSE,
  cure_mod = FALSE,
  cure_link = "logistic",
  group_names = levels(group)
)
```

---

f_surv_model_excel *Prepare Excel-Ready Survival Model Output*

Description

Formats model parameters (including spline knots) into a table suitable for export to Excel.

Usage

```
f_surv_model_excel(ngroups, strata, surv_model, spline_mod, cure_mod)
```

Arguments

ngroups	Integer. Number of groups.
strata	Logical. Whether stratified models were used.
surv_model	List of fitted models from f_surv_model().
spline_mod	Logical. Whether spline models were included.
cure_mod	Logical. Whether cure models were included.

Value

A transposed data frame containing:

- Distribution names
- Parameter names
- Estimates, SE, CI
- Knot values (if splines fitted)
- Covariance matrix

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
form <- stats::as.formula(survival::Surv(years, status) ~ group)
surv_model <- f_surv_model(
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  ngroups = nlevels(group),
  form = form,
  spline_mod = FALSE,
  cure_mod = FALSE,
  cure_link = "logistic",
  group_names = levels(group)
)
f_surv_model_excel(
  ngroups = nlevels(group),
  strata = FALSE,
  surv_model = surv_model,
  spline_mod = FALSE,
  cure_mod = FALSE
)
```

---

f_surv_model_pred *Predict from Survival Models*

Description

Generates predicted survival and hazard values from fitted parametric, spline, and cure models.

Usage

```
f_surv_model_pred(
  ngroups,
  time_pred,
  surv_model,
  spline_mod,
  cure_mod,
  group_names
)
```

Arguments

ngroups	Integer. Number of groups.
time_pred	Numeric vector of prediction times.
surv_model	List of fitted survival models from f_surv_model().
spline_mod	Logical. Whether spline models were fitted.
cure_mod	Logical. Whether cure models were fitted.
group_names	Character vector of group labels.

Value

A list of predictions containing:

- param_models: Survival & hazard predictions for standard models.
- spline: Predictions for spline models (if fitted).
- cure: Predictions for cure models (if fitted).

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
form <- stats::as.formula(survival::Surv(years, status) ~ group)
surv_model <- f_surv_model(
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  ngroups = nlevels(group),
  form = form,
  spline_mod = FALSE,
  cure_mod = FALSE,
  cure_link = "logistic",
  group_names = levels(group)
)
f_surv_model_pred(
  ngroups = nlevels(group),
  time_pred = seq(0, 5000, 365.25),
```

```
  surv_model = surv_model,
  spline_mod = FALSE,
  cure_mod = FALSE,
  group_names = levels(group)
)
```

`f_surv_model_pred_gr` *Group Predictions by Survival Model*

Description

Consolidates predictions from `f_surv_model_pred()` into group-specific data frames.

Usage

```
f_surv_model_pred_gr(
  ngroups,
  surv_model,
  surv_model_pred,
  spline_mod,
  cure_mod
)
```

Arguments

`ngroups`	Integer. Number of groups.
`surv_model`	List of survival models from `f_surv_model()`.
`surv_model_pred`	List of predictions from `f_surv_model_pred()`.
`spline_mod`	Logical. Whether spline models were fitted.
`cure_mod`	Logical. Whether cure models were fitted.

Value

A list of length `ngroups`, each a data frame with columns:

- `time`
- survival predictions for all models (parametric, spline, cure).

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
form <- stats::as.formula(survival::Surv(years, status) ~ group)
surv_model <- f_surv_model(
  years = years,
```

```
  status = status,
  group = group,
  strata = FALSE,
  ngroups = nlevels(group),
  form = form,
  spline_mod = FALSE,
  cure_mod = FALSE,
  cure_link = "logistic",
  group_names = levels(group)
)
surv_model_pred <- f_surv_model_pred(
  ngroups = nlevels(group),
  time_pred = seq(0, 5000, 365.25),
  surv_model = surv_model,
  spline_mod = FALSE,
  cure_mod = FALSE,
  group_names = levels(group)
)
f_surv_model_pred_gr(
  ngroups = nlevels(group),
  surv_model = surv_model,
  surv_model_pred = surv_model_pred,
  spline_mod = FALSE,
  cure_mod = FALSE
)
```

`f_surv_model_pred_tp_gr` *Compute Transition Probabilities for Survival Model Predictions*

Description

Compute Transition Probabilities for Survival Model Predictions

Usage

```
f_surv_model_pred_tp_gr(
  ngroups,
  time_pred,
  time_unit,
  surv_model_pred_gr,
  cols_tp
)
```

Arguments

`ngroups`	Integer, number of groups.
`time_pred`	Numeric vector of prediction times (currently unused).

time_unit	Numeric, time unit for transition probability calculation.
surv_model_pred_gr	List of group predictions. Each group's table should have a time column in column 1 and survival-related columns from 2:cols_tp.
cols_tp	Integer, index of the last survival-related column (i.e., use columns 2:cols_tp).

Value

Named list of data.frames with transition probabilities (truncated after threshold).

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
form <- stats::as.formula(survival::Surv(years, status) ~ group)
surv_model <- f_surv_model(
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  ngroups = nlevels(group),
  form = form,
  spline_mod = FALSE,
  cure_mod = FALSE,
  cure_link = "logistic",
  group_names = levels(group)
)
surv_model_pred <- f_surv_model_pred(
  ngroups = nlevels(group),
  time_pred = seq(0, 5000, 365.25),
  surv_model = surv_model,
  spline_mod = FALSE,
  cure_mod = FALSE,
  group_names = levels(group)
)
surv_model_pred_gr <- f_surv_model_pred_gr(
  ngroups = nlevels(group),
  surv_model = surv_model,
  surv_model_pred = surv_model_pred,
  spline_mod = FALSE,
  cure_mod = FALSE
)
f_surv_model_pred_tp_gr(
  ngroups = nlevels(group),
  time_pred = seq(0, 5000, 365.25),
  time_unit = 365.25/12,
  surv_model_pred_gr = surv_model_pred_gr,
  cols_tp = 8
)
```

f_tp *Calculate Transition Probabilities*

Description

Derives annualized transition probabilities (and confidence bounds) from cumulative hazard estimates, smoothed with LOESS.

Usage

```
f_tp(ngroups, cum_haz, time_unit)
```

Arguments

ngroups	Integer. Number of groups (1-3).
cum_haz	Data frame from f_cum_hazard() with columns group, time, H_delta, H_upper_delta, H_lower_delta.
time_unit	Numeric. Time unit for annualization.

Value

A list with:

- gr_1, gr_2, gr_3: Data frames of smoothed probabilities per group.
- max: Maximum upper bound across all groups.

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
cum_haz <- f_cum_hazard(
  years = years,
  status = status,
  group = group,
  ngroups = nlevels(group),
  time_pred = seq(0, 5000, 365.25),
  time_unit = 30
)
f_tp(ngroups = nlevels(group), cum_haz = cum_haz, time_unit = 30)
```

`nsclc_ipd` *Reconstructed Individual Patient Data for Overall Survival in NSCLC Trials*

Description

Reconstructed pseudo-individual patient data (pseudo-IPD) for overall survival (OS) from published Kaplan-Meier curves in trials of targeted therapies and immunotherapies for advanced non-small cell lung cancer (NSCLC). Reconstruction used digitized survival coordinates and, where available, published numbers at risk and total event counts, following the approach described by Guyot et al. (2012).

A review of all 33 trial-population datasets found that the reconstructed curves, survival medians, and treatment-effect estimates generally agree closely with their published sources. Known discrepancies and limitations are documented below; the records are not the original patient data.

Usage

```
nsclc_ipd
```

Format

A data frame with 11,572 reconstructed patient records from 33 trial-population datasets and 66 treatment arms, with 8 variables:

`dataset_id` Character. Identifier for a trial-population comparison. One trial can contribute more than one population or subgroup.

`trial_id` Character. Name or identifier of the source clinical trial.

`population` Character. Population or subgroup represented by the published OS curves.

`arm` Factor with levels `control` and `intervention`. Harmonized treatment assignment used in this data frame. These labels identify the comparison, not necessarily the reference category of a published HR.

`source_arm` Factor with levels `control` and `intervention`. Arm label used in the digitized source archive. It can differ from `arm` where source-arm labels have been reassigned.

`patient_id` Numeric. Artificial identifier within each dataset and arm. These identifiers do not identify real patients, including across overlapping subgroups reconstructed from the same trial.

`time_months` Numeric. Reconstructed time from randomization to death or right-censoring, expressed in months.

`event` Numeric. Event indicator: `1` for death and `0` for right-censoring.

Details

Origin and purpose.

These data build on the published survival curves and digitized source data from Simons et al. (2020). Kaplan-Meier coordinates were converted into pseudo-IPD using an inversion/reconstruction

procedure based on Guyot et al. (2012). Where reported, numbers at risk and total deaths informed the reconstruction. The resulting observations are estimates consistent with published aggregate survival information; they are *not* the original individual patient records.

The collection contains comparisons involving EGFR-targeted treatment, ALK-targeted treatment, and immunotherapy. Each `dataset_id` identifies one two-arm survival comparison. Population eligibility, line of treatment, treatment sequence, analysis cutoff, and survival maturity differ across trials. Some datasets are biomarker-defined subgroups rather than the whole randomized trial. In some trials, different PD-L1 or mutation-defined datasets overlap. Such datasets must not be analysed or pooled as independent patient cohorts.

Overall validation conclusion.

All 33 datasets were reviewed against the corresponding published OS Kaplan-Meier figures and reported survival outcomes. Across the collection, reconstructed curve shapes, median OS point estimates, and Cox hazard ratios generally correspond very closely to the published results and to the original recreated estimates in Simons et al. (2020). Several comparisons reproduce published arm-specific death totals exactly. For other comparisons, reconstructed death totals differ by a few events, and some original publications do not report arm-specific event totals. The most pronounced event-count discrepancy concerns CTONG-0901.

The validation is a comparison with *published aggregate information*, not patient-level validation against unavailable original IPD. Visual agreement, agreement in the HR, or a matching median does not establish correct individual event or censoring histories. The trial-specific notes below disclose important differences without implying that they must be resolved before the datasets can be used for methodological research.

PICO and reconstructed event proportions.

The trial-specific section reports the population, intervention, comparator, and outcome (PICO) for each comparison, together with the reconstructed sample sizes and event proportions for both arms. Because OS is the common outcome for all datasets, it is consistently defined as death from any cause. Event proportions describe the reconstructed observations rather than validated published event totals; they should not be treated as direct measures of follow-up maturity or survival-tail precision. They can be calculated directly from `nsclc_ipd`, as shown in the examples.

Interpreting reported quantities.

Unless stated otherwise, reconstructed HRs compare `intervention` with `control` in an unstratified Cox model, and reconstructed median OS values come from the reconstructed Kaplan-Meier estimator. Published HRs can be stratified, adjusted, or expressed in the opposite direction; published p-values can additionally reflect prespecified interim analyses and multiplicity adjustment. Therefore, differences in HRs, confidence intervals, and p-values need not be reconstruction errors. A confidence interval for median OS may be unreached even when its point estimate is available; estimates near the 50 percent KM threshold are especially sensitive to small differences in event and censoring allocation.

Scope of the data.

The pseudo-IPD contain survival times, censoring indicators, trial-population identifiers, and treatment assignments, but not individual mutation status, PD-L1 expression, other prognostic factors, subsequent therapy, crossover, or treatment-switching history. Where crossover or subsequent therapies occurred, the published OS curves generally describe survival under the randomized treatment strategies, not a counterfactual estimate adjusted for switching. Reconstructed subgroup records cannot be matched to the same individuals in a separately reconstructed whole-trial dataset.

Finding original KM curves.

The companion data documentation nsclc_trials (see 'nsclc_trials.Rd') describes the trial-level metadata, including dataset_id, treatment labels, publication details, DOI, and the trial_url column. A convenient way to open the original publication containing a survival curve is:

```
browseURL(nsclc_trials$trial_url[28])
```

With the supplied ordering, row 28 is the PD-L1 >=50 percent KEYNOTE-042 comparison. The URL resolves to a publication page; the KM figure may need to be opened within the article or its PDF, and publisher access conditions may apply. For reproducible code, selecting a row by dataset_id is safer than relying on its numerical row position.

Trial-specific validation findings

The descriptions below summarize a manual comparison of the reconstructed pseudo-IPD with the relevant published OS figure and statistics. Values headed as published or reported refer to the particular source analysis, not necessarily a later follow-up publication. Death-count discrepancies are stated where there is an identifiable published or independently sourced reference; unknown or unspecified counts are not treated as zero. Each entry provides a compact PICO description (population, intervention, comparator, outcome), the reconstructed sample size, number of deaths, and event proportion for each harmonized treatment arm, followed by the original validation findings. Arm-level figures are calculated from the current nsclc_ipd records; they are not necessarily the deaths reported in the corresponding publication. *Intervention* and *control* refer to the harmonized arm labels in this package, not necessarily the reference contrast in the original study. Treatment and population details refer to the particular reconstructed comparison, which may be a subgroup or two selected arms from a larger randomized trial.

The event proportion is the reconstructed number of deaths divided by the reconstructed number of patients in that arm (multiplied by 100). It is a simple descriptor of observed event information, sometimes informally used as a proxy for survival-data maturity, but *not* a direct measure of follow-up completeness, precision of tail estimates, or the percentage of patients who will eventually experience the event. The event proportion may differ between arms because of a treatment effect, different censoring patterns, or reconstruction discrepancies. In particular, the high reconstructed event proportions in CTONG-0901 should not be interpreted as verified published OS maturity. Comparisons with very small sample sizes, such as TITAN, remain highly uncertain irrespective of the percentage.

Abbreviations: HR, hazard ratio; NR, not reached; ITT, intention to treat; TPS, tumour proportion score; WT, wild type.

ALK_ALEX **Population (P):** Previously untreated advanced ALK-positive NSCLC in the ALEX trial.

Intervention (I): Alectinib.

Comparator (C): Crizotinib.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=152, deaths=36 (23.7 percent); control: N=151, deaths=40 (26.5 percent).

Validation and limitations: Alectinib versus crizotinib in ALK-positive NSCLC. Both OS medians are NR. The reconstructed HR is approximately 0.81 versus a published HR of approximately 0.76. The reconstruction has 36 alectinib deaths versus 35 specified in the source reconstruction information, while the 40 crizotinib deaths agree. OS was relatively immature.

ALK_ALTA-L1 **Population (P):** Previously untreated advanced ALK-positive NSCLC in ALTA-1L.

Intervention (I): Brigatinib.

Comparator (C): Crizotinib.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=137, deaths=17 (12.4 percent); control: N=138, deaths=17 (12.3 percent).

Validation and limitations: Brigatinib versus crizotinib, ALK-positive population. Both arms have 17 reconstructed deaths, matching reported counts, and neither median is reached. Reconstructed HR 1.01 is close to the published 0.98; only 34 deaths in 275 patients inform this early OS comparison.

ALK_ASCEND-4 **Population (P):** Previously untreated advanced ALK-positive NSCLC in ASCEND-4.

Intervention (I): Ceritinib.

Comparator (C): Platinum-based chemotherapy plus pemetrexed.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=189, deaths=47 (24.9 percent); control: N=187, deaths=59 (31.6 percent).

Validation and limitations: Ceritinib versus platinum/pemetrexed in ALK-positive NSCLC. The reconstructed chemotherapy median is 26.33 months versus 26.2 published; ceritinib median is NR. The reconstructed HR is 0.71 versus 0.73 published. There is one fewer reconstructed ceritinib death (47 versus 48); the chemotherapy event count (59) matches.

ALK_PROFILE1007 **Population (P):** Previously treated advanced ALK-positive NSCLC in PROFILE 1007.

Intervention (I): Crizotinib.

Comparator (C): Pemetrexed or docetaxel chemotherapy.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=173, deaths=48 (27.7 percent); control: N=174, deaths=47 (27.0 percent).

Validation and limitations: Crizotinib versus pemetrexed or docetaxel in previously treated ALK-positive NSCLC. The reconstructed HR is approximately 1.02, matching the published 1.02; medians are approximately 20.5 and 23.0 months versus 20.3 and 22.8 published. The reconstruction has 48 crizotinib deaths versus 49 specified, with 47 chemotherapy deaths reproduced. Subsequent crizotinib treatment in the chemotherapy arm complicates interpretation of the ITT OS comparison.

ALK_PROFILE1014 **Population (P):** Previously untreated advanced ALK-positive NSCLC in PROFILE 1014; mature OS analysis.

Intervention (I): Crizotinib.

Comparator (C): Platinum-based chemotherapy plus pemetrexed.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=172, deaths=72 (41.9 percent); control: N=171, deaths=81 (47.4 percent).

Validation and limitations: Crizotinib versus platinum/pemetrexed in ALK-positive NSCLC. The digitized curves correspond to the more mature 2018 OS report, not the earlier 2014 initial report. Reconstructed chemotherapy median 43.91 months differs from 47.5 months published, largely in a KM region near the 50 percent threshold. Crizotinib median is NR;

reconstructed HR 0.82 agrees closely with the Simons-recreated HR. There is one additional reconstructed crizotinib death (72 versus 71).

`EGFR_CTONG-0901` **Population (P):** Advanced EGFR-mutated NSCLC in the CTONG-0901 comparison.

Intervention (I): Erlotinib.

Comparator (C): Gefitinib.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=128, deaths=117 (91.4 percent); control: N=128, deaths=121 (94.5 percent).

Validation and limitations: Erlotinib versus gefitinib in EGFR-mutated NSCLC. The original digitized source-arm labels were reversed relative to these treatment names; `arm` gives the harmonized assignment. Median OS and the reconstructed HR (approximately 0.86) match the publication closely. *However*, the reconstruction contains 121 gefitinib and 117 erlotinib deaths versus 92 deaths reported for each arm. This major discrepancy indicates that the event/censoring histories, despite their similar KM shapes and HR, should not be treated as faithfully recovered.

`EGFR_DELTA` **Population (P):** Previously treated advanced NSCLC with an activating EGFR mutation; small DELTA subgroup.

Intervention (I): Erlotinib.

Comparator (C): Docetaxel.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=21, deaths=4 (19.0 percent); control: N=30, deaths=12 (40.0 percent).

Validation and limitations: EGFR-mutated subgroup of DELTA: erlotinib (21 patients) versus docetaxel (30). The reconstructed HR 0.45 resembles the published 0.43; the docetaxel median is approximately 27.8 months and the erlotinib median is NR. Only four and 12 deaths, respectively, are reconstructed. Original arm-specific total-event constraints were recorded as NA; the small sample and sparse tail limit precision.

`EGFR_ENSURE` **Population (P):** Previously untreated advanced EGFR-mutated NSCLC in ENSURE.

Intervention (I): Erlotinib.

Comparator (C): Cisplatin plus gemcitabine.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=110, deaths=57 (51.8 percent); control: N=107, deaths=56 (52.3 percent).

Validation and limitations: Erlotinib versus cisplatin/gemcitabine. Reconstructed medians are 26.42 and 25.66 months, closely matching published medians 26.3 and 25.5 months; reconstructed HR 0.94 versus 0.91 published. The chemotherapy reconstruction contains 56 rather than 54 specified deaths; the 57 erlotinib deaths agree.

`EGFR_EURTAC` **Population (P):** Previously untreated advanced EGFR mutation-positive NSCLC in EURTAC.

Intervention (I): Erlotinib.

Comparator (C): Platinum-based chemotherapy.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=86, deaths=38 (44.2 percent); control: N=87, deaths=31 (35.6 percent).

Validation and limitations: Erlotinib versus platinum-based chemotherapy in EGFR-mutated NSCLC. Reconstructed median OS (19.37 versus 19.56 months) and HR (1.05) are close to the published estimates (19.3 versus 19.5 months and HR 1.04). The specified death totals, 38 and 31, are reproduced; confidence limits for median survival are not identical.

EGFR_FLAURA **Population (P):** Previously untreated advanced EGFR-mutated NSCLC (exon 19 deletion or L858R); initial FLAURA OS analysis.

Intervention (I): Osimertinib.

Comparator (C): Gefitinib or erlotinib.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=279, deaths=58 (20.8 percent); control: N=277, deaths=83 (30.0 percent).

Validation and limitations: First-line osimertinib versus gefitinib or erlotinib. Both median OS estimates are NR at the early analysis cutoff. Published and reconstructed event counts match (58 osimertinib, 83 control), and reconstructed HR 0.62 closely matches the published 0.63. These pseudo-IPD do not represent the later mature OS analysis.

EGFR_IMPRESS **Population (P):** Advanced EGFR mutation-positive NSCLC following progression on first-line gefitinib.

Intervention (I): Continued gefitinib plus cisplatin and pemetrexed.

Comparator (C): Placebo plus cisplatin and pemetrexed.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=133, deaths=94 (70.7 percent); control: N=132, deaths=80 (60.6 percent).

Validation and limitations: Gefitinib plus cisplatin/pemetrexed versus placebo plus the same chemotherapy following progression on gefitinib. Reconstructed medians are 13.40 and 19.85 months versus published 13.4 and 19.5; reconstructed HR 1.46 agrees with 1.44 published. The chemotherapy-only arm has 80 versus 82 specified deaths; 94 deaths in the gefitinib-containing arm are reproduced.

EGFR_IPASS **Population (P):** EGFR mutation-positive subgroup of the first-line IPASS trial, not its full population.

Intervention (I): Gefitinib.

Comparator (C): Carboplatin plus paclitaxel.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=132, deaths=102 (77.3 percent); control: N=129, deaths=95 (73.6 percent).

Validation and limitations: EGFR mutation-positive subgroup, not the full IPASS randomized population. Gefitinib is compared with carboplatin/paclitaxel. Reconstructed medians are 21.02 and 22.19 months, close to published 21.6 and 21.9 months; reconstructed HR 1.05 versus published 1.00. The gefitinib reconstruction contains 102 versus 104 published deaths; all 95 chemotherapy deaths are reproduced.

EGFR_Lux-Lung3 **Population (P):** Previously untreated advanced EGFR-mutated NSCLC in the overall LUX-Lung 3 population.

Intervention (I): Afatinib.

Comparator (C): Cisplatin plus pemetrexed.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=230, deaths=139 (60.4 percent); control: N=115, deaths=72 (62.6 percent).

Validation and limitations: First-line afatinib versus cisplatin/pemetrexed in the overall EGFR- mutated trial population. Reconstructed medians are 29.44 and 28.31 months versus 28.2 months in both published arms. The reconstructed HR 0.883 closely matches the published 0.88 and the Simons reconstruction. One death is missing in each arm (139 versus 140 afatinib, 72 versus 73 chemotherapy). Treatment effects differ by EGFR mutation subtype; individual subtype information is unavailable here.

`EGFR_Lux-Lung6` **Population (P):** Previously untreated advanced EGFR-mutated NSCLC in LUX-Lung 6.

Intervention (I): Afatinib.

Comparator (C): Gemcitabine plus cisplatin.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=242, deaths=161 (66.5 percent); control: N=122, deaths=83 (68.0 percent).

Validation and limitations: First-line afatinib versus gemcitabine/cisplatin. Reconstructed medians are 23.21 and 23.80 months versus 23.1 and 23.5 published. Reconstructed HR 0.89 matches Simons' recreated 0.89, compared with a published HR of 0.93. One death is missing in each arm (161 versus 162 afatinib, 83 versus 84 chemotherapy). Published stratified and reconstructed unstratified p-values differ.

`EGFR_Lux-Lung7` **Population (P):** Previously untreated advanced NSCLC with EGFR exon 19 deletion or L858R mutation.

Intervention (I): Afatinib.

Comparator (C): Gefitinib.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=160, deaths=109 (68.1 percent); control: N=159, deaths=118 (74.2 percent).

Validation and limitations: First-line afatinib versus gefitinib among patients with exon 19 deletion or L858R EGFR mutations. Reconstructed medians (27.39 and 24.34 months) and HR (0.852) closely match published values (27.9, 24.5, and 0.86). The reconstruction has one additional gefitinib death (118 versus 117); all 109 afatinib deaths are reproduced. Late KM steps are based on small risk sets.

`EGFR_NCT02959749` **Population (P):** Advanced EGFR T790M-positive NSCLC previously treated with an EGFR-TKI and platinum-based chemotherapy.

Intervention (I): Osimertinib.

Comparator (C): Docetaxel plus bevacizumab.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=74, deaths=14 (18.9 percent); control: N=73, deaths=15 (20.5 percent).

Validation and limitations: Osimertinib versus docetaxel/bevacizumab after prior EGFR-directed and platinum-based therapy in EGFR T790M-positive NSCLC. Both KM medians are NR and reconstructed HR 0.80 agrees with published 0.79. The original report contains internally inconsistent median OS values: its text and curves imply NR, whereas a table reports numerical medians with incompatible confidence intervals. The reconstruction has 14 and 15 deaths; the disposition diagram uses a different cutoff and does not unambiguously verify final OS event totals.

EGFR_OPTIMAL **Population (P):** Previously untreated advanced EGFR-mutated NSCLC in OPTIMAL.

Intervention (I): Erlotinib.

Comparator (C): Gemcitabine plus carboplatin.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=82, deaths=68 (82.9 percent); control: N=72, deaths=53 (73.6 percent).

Validation and limitations: Erlotinib versus gemcitabine/carboplatin. Reconstructed medians (22.86 versus 27.22 months) and HR (1.20) closely reproduce published values (22.8 versus 27.2 months; HR 1.19). The reconstruction contains one extra chemotherapy death (53 versus 52), while all 68 erlotinib deaths are reproduced. The terminal numbers-at-risk interval was omitted during reconstruction because it was not adequately supported by available digitized KM coordinates; retain this provenance caveat.

EGFR_TITAN **Population (P):** Activating EGFR mutation-positive subgroup of previously treated advanced NSCLC in TITAN (11 patients).

Intervention (I): Erlotinib.

Comparator (C): Docetaxel or pemetrexed.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=7, deaths=3 (42.9 percent); control: N=4, deaths=1 (25.0 percent).

Validation and limitations: EGFR mutation-positive subgroup of TITAN, with only seven erlotinib and four chemotherapy patients. The reconstruction contains three and one deaths, respectively; published subgroup death totals are not explicitly reported. The published chemotherapy KM plateau is about 67 percent versus 75 percent reconstructed. Reconstructed HR 1.39 versus 1.19 published has an extremely wide confidence interval. This subgroup is unsuitable for reliable comparative-effect estimation and particularly sensitive to individual reconstructed observations.

EGFR_TORCH **Population (P):** EGFR mutation-positive subgroup of TORCH; randomized comparison of treatment sequences.

Intervention (I): Erlotinib first, followed by cisplatin plus gemcitabine at progression.

Comparator (C): Cisplatin plus gemcitabine first, followed by erlotinib at progression.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=19, deaths=14 (73.7 percent); control: N=20, deaths=10 (50.0 percent).

Validation and limitations: EGFR-mutated subgroup comparing *treatment sequences*: erlotinib followed by cisplatin/gemcitabine versus the reverse sequence. All published deaths are reproduced (14 versus 10), and reconstructed HR 1.51 closely matches Simons' 1.51 and published subgroup HR 1.58. The chemotherapy-first median is 26.0 versus 32.5 months published, although both curves lie very near the 50 percent KM threshold; the discrepancy remains documented rather than corrected. Subgroup numbers at risk are not provided in the figure.

EGFR_WJOG-5108L **Population (P):** Previously treated EGFR mutation-positive lung adenocarcinoma in WJOG-5108L, including uncommon mutations.

Intervention (I): Erlotinib.

Comparator (C): Gefitinib.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=198, deaths=96 (48.5 percent); control: N=203, deaths=107 (52.7 percent).

Validation and limitations: EGFR mutation-positive subgroup comparing erlotinib (`intervention`) with gefitinib (`control`). The published HR is expressed in the opposite direction (gefitinib versus erlotinib): published 1.189 versus inverted reconstructed approximately 1.181. Reconstructed medians are 30.36 months for erlotinib and 26.60 for gefitinib, versus published 31.4 and 26.5. Subgroup-specific published death totals are unavailable; counts for the larger trial population are not interchangeable.

`Immunotherapy_CheckMate-026_PDL1_5` **Population (P):** Previously untreated advanced NSCLC with PD-L1 expression of at least 5 percent.

Intervention (I): Nivolumab.

Comparator (C): Platinum-doublet chemotherapy.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=211, deaths=130 (61.6 percent); control: N=212, deaths=122 (57.5 percent).

Validation and limitations: First-line nivolumab versus platinum-doublet chemotherapy among patients with PD-L1 expression >=5 percent. Reconstructed medians are 14.57 and 13.40 months versus 14.4 and 13.2 published; reconstructed HR 1.04 versus published 1.02. Reconstructed deaths are 130 and 122; the figure alone does not establish arm-specific published OS totals. The small late risk sets affect the KM tails.

`Immunotherapy_CheckMate_017_PDL1_unselected` **Population (P):** Previously treated advanced squamous NSCLC, without a PD-L1 expression eligibility threshold.

Intervention (I): Nivolumab.

Comparator (C): Docetaxel.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=135, deaths=86 (63.7 percent); control: N=137, deaths=113 (82.5 percent).

Validation and limitations: Previously treated *squamous* NSCLC, nivolumab versus docetaxel; PD-L1 expression was not an entry requirement. All published deaths are reproduced (86 nivolumab; 113 docetaxel). Reconstructed medians are 9.27 and 6.04 months versus 9.2 and 6.0 published; HR 0.591 versus 0.59 published. The reconstructed upper confidence limit for nivolumab median OS is 15.39 versus 13.3 months published.

`Immunotherapy_CheckMate_057_PDL1_unselected` **Population (P):** Previously treated advanced nonsquamous NSCLC, without a PD-L1 expression eligibility threshold.

Intervention (I): Nivolumab.

Comparator (C): Docetaxel.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=292, deaths=192 (65.8 percent); control: N=290, deaths=221 (76.2 percent).

Validation and limitations: Previously treated *nonsquamous* NSCLC, nivolumab versus docetaxel; PD-L1 unselected. The total 413 published deaths are reproduced, but assigned differently: 192 versus 190 published nivolumab deaths and 221 versus 223 docetaxel deaths. Reconstructed medians (12.16 and 9.45 months) and HR 0.75 agree closely with published values. The published primary HR has a 96 percent confidence interval, whereas the reconstruction reports 95 percent; curves show early crossing and a later nivolumab advantage.

`Immunotherapy_IMpower150_PDL1_unselected` **Population (P):** Previously untreated metastatic nonsquamous NSCLC, EGFR/ALK wild-type ITT population, irrespective of PD-L1; two of three randomized arms.

Intervention (I): Atezolizumab plus bevacizumab, carboplatin, and paclitaxel (ABCP).

Comparator (C): Bevacizumab plus carboplatin and paclitaxel (BCP).

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=359, deaths=179 (49.9 percent); control: N=337, deaths=196 (58.2 percent).

Validation and limitations: Atezolizumab/bevacizumab/carboplatin/paclitaxel (ABCP) versus bevacizumab/carboplatin/paclitaxel (BCP). This dataset includes only these two arms of the *EGFR/ALK wild-type* ITT population, not the third randomized arm or all trial patients. Reconstructed medians (19.46 versus 14.72 months) and HR 0.784 closely match published estimates. ABCP deaths match (179); BCP has 196 versus 197 published. Data are from an interim OS analysis.

`Immunotherapy_KEYNOTE-010_PDL1_1` **Population (P):** Previously treated advanced NSCLC with PD-L1 TPS of at least 1 percent; includes the at least 50 percent subgroup.

Intervention (I): Pembrolizumab 2 mg/kg every three weeks; the 10 mg/kg arm is excluded.

Comparator (C): Docetaxel.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=344, deaths=174 (50.6 percent); control: N=343, deaths=197 (57.4 percent).

Validation and limitations: PD-L1 TPS >=1 percent, pembrolizumab *2 mg/kg* versus docetaxel. The pembrolizumab 10 mg/kg randomized arm is not included. Medians (10.42 and 8.57 months) almost exactly match published values; reconstructed HR 0.729 versus published 0.71. There are two extra pembrolizumab deaths (174 versus 172) and four extra docetaxel deaths (197 versus 193). This dataset overlaps the >=50 percent KEYNOTE-010 subgroup below.

`Immunotherapy_KEYNOTE-010_PDL1_50` **Population (P):** Previously treated advanced NSCLC with PD-L1 TPS of at least 50 percent; nested in the at least 1 percent population.

Intervention (I): Pembrolizumab 2 mg/kg every three weeks; the 10 mg/kg arm is excluded.

Comparator (C): Docetaxel.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=139, deaths=59 (42.4 percent); control: N=152, deaths=88 (57.9 percent).

Validation and limitations: PD-L1 TPS >=50 percent subset of KEYNOTE-010, again comparing pembrolizumab *2 mg/kg* with docetaxel. Published and reconstructed medians agree almost exactly (14.9 and 8.2 months); reconstructed HR 0.600 versus published 0.54. There is one extra pembrolizumab death (59 versus 58) and two extra docetaxel deaths (88 versus 86). This subset is not independent of the >=1 percent dataset.

`Immunotherapy_KEYNOTE-024_PDL1_50` **Population (P):** Previously untreated advanced NSCLC with PD-L1 TPS of at least 50 percent, without sensitizing EGFR/ALK alterations.

Intervention (I): Pembrolizumab 200 mg every three weeks.

Comparator (C): Platinum-doublet chemotherapy.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=154, deaths=44 (28.6 percent); control: N=151, deaths=64 (42.4 percent).

Validation and limitations: First-line pembrolizumab versus platinum-doublet chemotherapy, PD-L1 TPS >=50 percent without sensitizing EGFR/ALK alterations. Both median OS values were *not reached* in the original 2016 publication and in the current reconstruction. All 44 pembrolizumab and 64 chemotherapy deaths are reproduced; reconstructed HR 0.615 versus published 0.60. Simons et al. (2020) instead report a 12.6-month chemotherapy median in both comparison columns, inconsistent with the original analysis. The chemotherapy KM curve is near 50 percent, making the median classification sensitive to reconstruction details.

`Immunotherapy_KEYNOTE-042_PDL1_50` **Population (P):** Previously untreated advanced NSCLC with PD-L1 TPS of at least 50 percent, without sensitizing EGFR/ALK alterations.

Intervention (I): Pembrolizumab 200 mg every three weeks.

Comparator (C): Platinum-doublet chemotherapy.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=299, deaths=159 (53.2 percent); control: N=300, deaths=201 (67.0 percent).

Validation and limitations: First-line pembrolizumab versus chemotherapy, PD-L1 TPS >=50 percent. Reconstructed medians are 20.02 and 12.18 months versus 20.0 and 12.2 published; reconstructed HR 0.703 versus published 0.69. Published deaths inferred from the final cumulative censoring table are 157 pembrolizumab and 199 chemotherapy; reconstructed totals are 159 and 201, respectively. Published curves cross in early follow-up. The corresponding event references must not be reversed between arms.

`Immunotherapy_KEYNOTE-189_PDL1_50` **Population (P):** Previously untreated metastatic nonsquamous NSCLC without sensitizing EGFR/ALK alterations, PD-L1 TPS at least 50 percent.

Intervention (I): Pembrolizumab plus platinum/pemetrexed, followed by pembrolizumab/pemetrexed maintenance.

Comparator (C): Placebo plus platinum/pemetrexed, followed by placebo/pemetrexed maintenance.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=132, deaths=33 (25.0 percent); control: N=70, deaths=35 (50.0 percent).

Validation and limitations: PD-L1 TPS >=50 percent subgroup, pembrolizumab plus platinum/pemetrexed versus placebo plus the same chemotherapy (132 versus 70 patients). The published Figure 2C reports an HR of 0.42 versus reconstructed 0.408; the survival curves agree closely. Reconstructed deaths total 68 versus 70 reported for the subgroup. An ancillary source reports 34 intervention and 36 control deaths, versus 33 and 35 reconstructed; these arm-specific numbers are not in the main KM figure. The numerical subgroup medians are not reported in that figure. The subgroup overlaps the whole-trial dataset below.

`Immunotherapy_KEYNOTE-189_PDL1_unselected` **Population (P):** Previously untreated metastatic nonsquamous NSCLC without sensitizing EGFR/ALK alterations, irrespective of PD-L1.

Intervention (I): Pembrolizumab plus platinum/pemetrexed, followed by pembrolizumab/pemetrexed maintenance.

Comparator (C): Placebo plus platinum/pemetrexed, followed by placebo/pemetrexed maintenance.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=410, deaths=125 (30.5 percent); control: N=206, deaths=107 (51.9 percent).

Validation and limitations: Full ITT population with metastatic *nonsquamous* NSCLC without sensitizing EGFR/ALK alterations, receiving pembrolizumab plus platinum/pemetrexed or placebo plus the same regimen, including maintenance treatment. Reconstructed median OS is NR versus 11.25 months (published NR versus 11.3), and reconstructed HR 0.492 matches published 0.49. Reconstructed events are 125 and 107, compared with 127 and 108 in regulatory efficacy information. The reconstructed chemotherapy median upper confidence limit is NR versus published 15.1 months. The PD-L1 >=50 percent subgroup is nested in this cohort.

`Immunotherapy_KEYNOTE-407_PDL1_unselected` **Population (P):** Previously untreated metastatic squamous NSCLC, irrespective of PD-L1 expression.

Intervention (I): Pembrolizumab with carboplatin and paclitaxel/nab-paclitaxel, followed by pembrolizumab maintenance.

Comparator (C): Placebo with carboplatin and paclitaxel/nab-paclitaxel, followed by placebo maintenance.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=278, deaths=87 (31.3 percent); control: N=281, deaths=114 (40.6 percent).

Validation and limitations: Full first-line *squamous* NSCLC population, pembrolizumab plus carboplatin and paclitaxel/nab-paclitaxel versus placebo plus the same chemotherapy, with subsequent study-treatment maintenance. Reconstructed medians (15.89 versus 11.44 months) agree with published 15.9 versus 11.3; reconstructed HR 0.677 is close to Simons' 0.68 and published 0.64. Death counts differ: 87 versus 85 intervention and 114 versus 120 control. The source is an immature interim OS analysis.

`Immunotherapy_OAK_PDL1_unselected` **Population (P):** Previously treated advanced squamous or nonsquamous NSCLC, irrespective of PD-L1; first 850 randomized patients in the primary ITT analysis.

Intervention (I): Atezolizumab.

Comparator (C): Docetaxel.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=425, deaths=271 (63.8 percent); control: N=425, deaths=298 (70.1 percent).

Validation and limitations: Atezolizumab versus docetaxel in previously treated, PD-L1-unselected squamous or nonsquamous NSCLC. The reconstruction covers the first *850 randomized patients* used in the primary efficacy analysis, not all 1,225 ultimately randomized. Published event totals agree exactly (271 atezolizumab, 298 docetaxel). Medians (13.68 and 9.63 months) and reconstructed HR 0.731 closely match published values. The trial-flow diagram omits one subsequently ascertained death in each arm; the reconstruction correctly reflects the efficacy totals.

`Immunotherapy_POPLAR_PDL1_unselected` **Population (P):** Previously treated advanced squamous or nonsquamous NSCLC, irrespective of PD-L1; full randomized ITT population.

Intervention (I): Atezolizumab.

Comparator (C): Docetaxel.

Outcome (O): Overall survival (death from any cause).

Reconstructed arm-level information: Intervention: N=144, deaths=77 (53.5 percent); control: N=143, deaths=95 (66.4 percent).

Validation and limitations: Atezolizumab versus docetaxel in previously treated, PD-L1-unselected squamous or nonsquamous NSCLC. Reconstructed medians (12.63 versus 9.72 months) and HR 0.724 are very close to published 12.6 versus 9.7 months and HR 0.73. The reconstruction contains 77 versus 78 published atezolizumab deaths and correctly reproduces all 95 docetaxel deaths. The trial-flow diagram lists 93 docetaxel deaths, but its footnote adds two deaths identified from public records. The upper reconstructed atezolizumab median confidence limit is NR versus published 16.4 months.

Appropriate and inappropriate uses

These data are intended for education, reproducible methodological illustration, Kaplan-Meier estimation, proportional hazards diagnostics, parametric and flexible survival modelling, and survival extrapolation. They may also be used to study the sensitivity of survival analyses to imperfect reconstruction, incomplete follow-up, small risk sets, and alternative censoring allocations.

They should not be presented as verified original patient-level trial data. Patient-specific covariates or causal effects of subsequent therapy cannot be inferred from reconstructed records. Comparisons from nested or overlapping subgroups should not be treated as statistically independent; differences in analysis cutoffs should be respected. The particularly large event-count mismatch in CTONG-0901 and the extremely small TITAN subgroup warrant caution in analyses that depend on accurate event or censoring histories. More broadly, late-tail predictions should account for reconstruction uncertainty as well as sampling and extrapolation uncertainty.

Survival times are expressed in months. For applications using years, including `f_PERSUADE()`, divide `time_months` by 12.

Source

The digitized Kaplan-Meier coordinates and associated numbers-at-risk source files originate from:

Simons, M., et al. (2020). Digitized patient level time to event data of overall survival. Zenodo. doi:10.5281/zenodo.4288605

The original digitized files were released under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence. The accompanying reconstructed pseudo-IPD in this package were assembled from these sources. The companion `nsclc_trials` dataset supplies the individual source publications and their URLs.

References

Simons, M., Ramaekers, B., Peeters, A., Mankor, J., Paats, M., Aerts, J., van Harten, W., Retel, V., and Joore, M. (2020). Observed versus modelled lifetime overall survival of targeted therapies and immunotherapies for advanced non-small cell lung cancer patients - A systematic review. Critical Reviews in Oncology/Hematology, 153, 103035. doi:10.1016/j.critrevonc.2020.103035

Guyot, P., Ades, A. E., Ouwens, M. J. N. M., and Welton, N. J. (2012). Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Medical Research Methodology, 12, 9. doi:10.1186/1471-2288-12-9

Trial-specific primary publications are identified by DOI and URL in `nsclc_trials`.

See Also

nsclc_trials

Examples

```
data("nsclc_ipd")
data("nsclc_trials")

# Inspect reconstructed patient histories and available comparisons
head(nsclc_ipd)
unique(nsclc_ipd$dataset_id)

# Arm-level reconstructed sample sizes, deaths, and event percentages
arm_summary <- aggregate(
  data.frame(
    N = rep(1L, nrow(nsclc_ipd)),
    deaths = nsclc_ipd$event
  ),
  by = list(
    dataset_id = nsclc_ipd$dataset_id,
    arm = as.character(nsclc_ipd$arm)
  ),
  FUN = sum
)
arm_summary$event_pct <- round(
  100 * arm_summary$deaths / arm_summary$N, 1
)

# Compare arms in a specific trial population
subset(
  arm_summary,
  dataset_id == "Immunotherapy_KEYNOTE-042_PDL1_50"
)

# Optionally attach the population and treatment descriptions
ix <- match(arm_summary$dataset_id, nsclc_trials$dataset_id)
arm_summary$population <- nsclc_trials$population[ix]
arm_summary$control_treatment <- nsclc_trials$control_treatment[ix]
arm_summary$intervention_treatment <- nsclc_trials$intervention_treatment[ix]

# Select LUX-Lung 3 and inspect deaths and censoring by treatment
luxlung3 <- subset(nsclc_ipd, dataset_id == "EGFR_Lux-Lung3")
with(luxlung3, table(arm, event))

# Recreate Kaplan-Meier estimates from the pseudo-IPD
km <- survival::survfit(
  survival::Surv(time_months, event) ~ arm,
  data = luxlung3
)
print(km)

# Trial metadata, citation, and original publication URL
```

```
subset(nsclc_trials, dataset_id == "EGFR_Lux-Lung3")

# Open the publication containing the original KM curves
# Row 28 is KEYNOTE-042 (PD-L1 >=50 percent) in the supplied order.
if (interactive()) {
  browseURL(nsclc_trials$trial_url[28])
}

# For reproducibility, locate the DOI URL using dataset_id instead
idx <- match("EGFR_Lux-Lung3", nsclc_trials$dataset_id)
if (interactive() && !is.na(idx)) {
  browseURL(nsclc_trials$trial_url[idx])
}
```

`nsclc_trials` *Trial Characteristics and References for Reconstructed NSCLC Survival Data*

Description

Trial-level metadata accompanying the reconstructed pseudo-individual patient data (pseudo-IPD) in `nsclc_ipd`. The dataset provides information on the original clinical trials, patient populations, treatment comparisons, publication references, and data reconstruction provenance.

Usage

```
nsclc_trials
```

Format

A data frame containing 33 trial-population datasets with the following 13 variables:

`dataset_id` Character. Unique identifier linking each trial-population dataset to the corresponding reconstructed observations in `nsclc_ipd`.

`trial_id` Character. Name or identifier of the original clinical trial.

`population` Character. Patient population or subgroup represented by the reconstructed survival data. Different populations from the same clinical trial may have separate dataset identifiers.

`source_group` Character. Broad clinical or treatment category used to organize the source data, including EGFR-targeted therapies, ALK-targeted therapies, and immunotherapies.

`control_treatment` Character. Treatment or treatment combination assigned to the control arm.

`intervention_treatment` Character. Treatment or treatment combination assigned to the intervention arm.

`included_in_paper_analysis` Logical. Indicates whether the dataset was included in the quantitative survival analysis reported by Simons et al. (2020). Additional datasets available in the source archive are also represented.

`trial_publication` Character. Short reference identifying the original clinical trial publication or publications.

`trial_doi` Character. Digital Object Identifier (DOI) of the original clinical trial publication, where available. Missing or empty entries do not necessarily indicate that a publication has no DOI.

`trial_bibliography` Character. Bibliographic information for the original clinical trial publication or publications associated with the reconstructed survival data.

`source_notes` Character. Additional information concerning the source data, population definitions, treatment assignments, or data reconstruction.

`swap_source_arms` Numeric. Indicator specifying whether the original source-file arm labels were reversed when assembling the reconstructed dataset: 0 = original labels retained; 1 = source labels reversed.

`trial_url` Character. Direct URL to the original clinical trial publication, constructed from `trial_doi`. The URL resolves through https://doi.org/.

Details

The metadata accompany reconstructed overall survival data from clinical trials investigating targeted therapies and immunotherapies in patients with locally advanced or metastatic non-small cell lung cancer (NSCLC).

The original source data were collected and digitized as part of the systematic review by Simons et al. (2020). Published Kaplan-Meier curves and numbers at risk were used to reconstruct pseudo-individual patient time-to-event data using the methodology described by Guyot et al. (2012).

Each row represents a unique trial-population dataset containing a control and intervention arm. Some clinical trials contribute multiple datasets representing distinct or potentially overlapping patient subgroups.

The variable `dataset_id` provides the link between this metadata and the reconstructed observations in `nsclc_ipd`.

The treatment variables describe the control and intervention assignments in the assembled dataset. The variable `swap_source_arms` documents whether these assignments required reversing the original source-file labels. The corresponding `arm` and `source_arm` variables are available in `nsclc_ipd`.

The variable `included_in_paper_analysis` distinguishes datasets included in the quantitative analysis by Simons et al. (2020) from additional datasets available in the digitized source archive. This indicator should not be interpreted as an assessment of data quality or reconstruction accuracy.

Bibliographic information is provided to enable users to identify and cite the original clinical trial publications alongside the Simons dataset and reconstruction methodology.

The reconstructed observations are not original patient-level trial data. They are intended primarily for educational and methodological research, including survival analysis, model selection, and survival extrapolation.

Source

The reconstructed pseudo-individual patient data were generated from digitized Kaplan-Meier survival curves and numbers-at-risk tables made publicly available by Simons et al. (2020):

Simons, M., et al. (2020). Digitized patient level time to event data of overall survival. Zenodo. doi:10.5281/zenodo.4288605

The original digitized data are distributed under the Creative Commons Attribution 4.0 International (CC BY 4.0) license.

References

Simons, M., Ramaekers, B., Peeters, A., Mankor, J., Paats, M., Aerts, J., van Harten, W., Retel, V., and Joore, M. (2020). Observed versus modelled lifetime overall survival of targeted therapies and immunotherapies for advanced non-small cell lung cancer patients - A systematic review. Critical Reviews in Oncology/Hematology, 153, 103035. doi:10.1016/j.critrevonc.2020.103035

Guyot, P., Ades, A. E., Ouwens, M. J. N. M., and Welton, N. J. (2012). Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Medical Research Methodology, 12, 9. doi:10.1186/1471-2288-12-9

References to the original clinical trial publications are included in the dataset.

See Also

nsclc_ipd

Examples

```
data("nsclc_trials")

# Inspect trial metadata
head(nsclc_trials)

# List available clinical trials
unique(nsclc_trials$trial_id)

# Select LUX-Lung 3 metadata
luxlung3_info <- subset(
  nsclc_trials,
  dataset_id == "EGFR_Lux-Lung3"
)

# View treatment comparison and references
luxlung3_info[, c(
  "trial_id",
  "control_treatment",
  "intervention_treatment",
  "trial_bibliography"
)]

# Identify datasets included in the published analysis
published_trials <- subset(
  nsclc_trials,
  included_in_paper_analysis
)

nrow(published_trials)
```

plot.PERSUADE *Plot Method for PERSUADE Objects*

Description

Generates diagnostic and model fit plots for PERSUADE survival analysis objects. The type argument controls which plot(s) are produced:

- "km": Kaplan-Meier survival curves.
- "ph": Proportional hazards diagnostics.
- "hr": Hazard function with fitted models.
- "param_models": Fitted parametric survival models with diagnostics and transition probability plots.
- "spline_models": Fitted spline-based survival models with diagnostics and transition probability plots.
- "cure_models": Fitted cure survival models with diagnostics and transition probability plots.

Usage

```
## S3 method for class 'PERSUADE'
plot(x, type = "km", ...)
```

Arguments

x	A PERSUADE object from f_PERSUADE().
type	Character. The type of plot to produce.
...	Additional arguments (currently unused).

Value

Invisibly returns a list of results from the plotting functions. Also produces base R plots as side effects.

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
PERSUADE <- f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = FALSE,
  cure_mod = FALSE,
```

```
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
plot(PERSUADE, "km")
```

`print.PERSUADE` *Print Method for PERSUADE Objects*

Description

Displays a brief summary of the PERSUADE object in the console.

Usage

```
## S3 method for class 'PERSUADE'
print(x, ...)
```

Arguments

`x`	A PERSUADE object from `f_PERSUADE()`.
`...`	Additional arguments (currently unused).

Value

Invisibly returns the PERSUADE object.

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
PERSUADE <- f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = FALSE,
  cure_mod = FALSE,
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
print(PERSUADE)
```

summary.PERSUADE *Summary Method for PERSUADE Objects*

Description

The type argument controls which summary is produced:

- "km": Kaplan-Meier estimates (default).
- "surv_probs": Survival probabilities at specified prediction times for each group.
- "gof": Goodness-of-fit statistics for standard parametric models.
- "gof_spline": Goodness-of-fit statistics for spline models.
- "gof_cure": Goodness-of-fit statistics for cure models (including cure fraction).

Usage

```
## S3 method for class 'PERSUADE'
summary(object, ..., type = "km")
```

Arguments

object	A PERSUADE object from f_PERSUADE().
...	Additional arguments. Currently only type is used.
type	Character string, one of "km", "surv_probs", "gof", "gof_spline", "gof_cure". Controls the type of summary output.

Value

A data frame or list of data frames depending on type.

Examples

```
years <- survival::lung$time
status <-  survival::lung$status
group <- factor(survival::lung$sex)
PERSUADE <- f_PERSUADE(
  name = "Example",
  years = years,
  status = status,
  group = group,
  strata = FALSE,
  spline_mod = FALSE,
  cure_mod = FALSE,
  time_unit = 365.25/12,
  time_horizon = 2000,
  time_pred_surv_table = seq(0, 2000, 365.25)
)
summary(PERSUADE, type = "surv_probs")
```

Index

∗ **datasets**
    nsclc_ipd, 42
    nsclc_trials, 56

f_cum_hazard, 3
f_cum_hazard(), *9*, *41*
f_generate_report, 4
f_generate_report(), *6*
f_get_excel_template, 6
f_hazard, 7
f_hazard(), *9*
f_PERSUADE, 8
f_PERSUADE(), *4, 5*, *10–32*
f_plot_cure_surv_extrap, 9
f_plot_cure_surv_model, 10
f_plot_diag_cure_surv_model, 11
f_plot_diag_param_surv_model, 12
f_plot_diag_spline_surv_model, 13
f_plot_hazard_cure_extrap, 14
f_plot_hazard_parametric_extrap, 15
f_plot_hazard_spline_extrap, 16
f_plot_hazard_with_models, 17
f_plot_km_survival, 18
f_plot_km_survival_base, 19
f_plot_log_cumhaz, 20
f_plot_param_surv_extrap, 21
f_plot_param_surv_model, 22
f_plot_schoenfeld_residuals, 23
f_plot_smoothed_hazard, 24
f_plot_spline_surv_extrap, 25
f_plot_spline_surv_model, 26
f_plot_tp_cure_surv_extrap, 27
f_plot_tp_cure_surv_model, 28
f_plot_tp_param_surv_extrap, 29
f_plot_tp_param_surv_model, 30
f_plot_tp_spline_surv_extrap, 31
f_plot_tp_spline_surv_model, 32
f_summary, 33
f_surv_model, 34
f_surv_model(), *9*, *35*, *37, 38*
f_surv_model_excel, 35
f_surv_model_pred, 36
f_surv_model_pred(), *38*
f_surv_model_pred_gr, 38
f_surv_model_pred_tp_gr, 39
f_tp, 41

nsclc_ipd, 42, *58*
nsclc_trials, *44*, *54, 55*, 56

plot.PERSUADE, 59
print.PERSUADE, 60

summary.PERSUADE, 61
survminer::ggsurvplot(), *18*
system.file(), *6*