

PERSUADE /tests/testthat/grid_test/check_9

2025-09-04

Contents

1. Kaplan-Meier	2
2. Proportional hazards assumption	5
3. Hazard function	6
3.1 Shape of the observed smoothed hazard function	6
3.2 Shape of the predicted hazard function	8
4 Parametric survival models	10
4.1 Standard parametric models	10
4.2 Parametric natural cubic spline models	19
4.3 Parametric (non-)mixture cure models	20
5 Extrapolation	22
5.1 Extrapolated survival	23
5.2 Extrapolated transition probabilities	24
5.3 Extrapolated hazard function	25
5.4 Tabulated results	26
6. PERSUADE object information	27
7. Session information	29

[Link to PERSUADE GitHub page](#)

1. Kaplan-Meier

The Kaplan-Meier (KM) curves provide a non-parametric estimate of the survival function, based solely on the observed data, without assuming any particular hazard shape or distribution. These plots allow visual comparison of survival experience between groups, highlighting differences in survival patterns and potential crossing of curves, which may later inform the choice of stratified vs. non-stratified models (see Section 2). The accompanying table summarises key KM estimates (e.g., median survival, survival probabilities at selected time points). These observed data serve as the benchmark against which all parametric, spline-based, and cure model fits will be assessed.

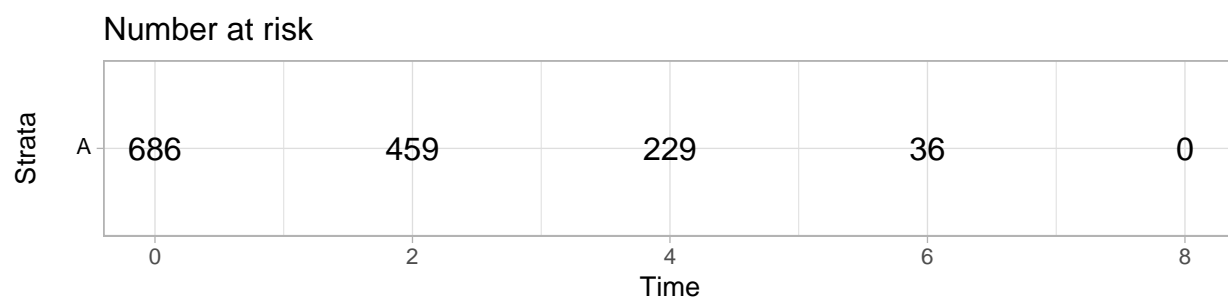


Table 1: Observed survival data

	x
records	686.0000
n.max	686.0000
n.start	686.0000
events	299.0000
rmean	4.5514
se(rmean)	0.1129
median	4.9507
0.95LCL	4.3479
0.95UCL	5.5616

2. Proportional hazards assumption

Should stratified parametric survival models be used?

To decide whether stratified or non-stratified parametric survival models should be applied, it is important to assess whether the **proportional hazards (PH) assumption** holds. This assumption means that the hazard ratio between two (or more) groups remains constant over time.

- **If the PH assumption holds:** Non-stratified parametric models can be used, with group effects included as covariates in a single model.
- **If the PH assumption does not hold:** It is advised to fit separate (stratified) parametric survival models for each group.

This section presents plots for assessing the PH assumption **over the observed data period only**. Even if the PH assumption appears valid here, it might be violated beyond the observed follow-up. For further details, see [NICE DSU TSD 14](#).

How to interpret the figures:

- **Figure A** plots the log of time against the log of the cumulative hazard for each group. Parallel lines suggest that the PH assumption holds.
- **Figures B and C** show scaled Schoenfeld residuals over time, with a smoothed spline (cyan) overlaid. If the spline systematically deviates from a horizontal line, this suggests a time-varying covariate effect and a potential PH violation ([Grambsch & Therneau, 1994](#)).
- **Crossing survival curves** in the Kaplan-Meier plot (previous page) may also indicate that the PH assumption is not satisfied.

Caution: These diagnostics only apply to the observed data period. Towards the end of follow-up, results may be unreliable due to sparse data, so the tail of curves should be interpreted with care.

3. Hazard function

3.1 Shape of the observed smoothed hazard function

The **hazard function** represents the instantaneous rate at which an event occurs at a specific time, conditional on survival until that time. Its shape varies depending on the assumed distribution (parametric survival model). When selecting an appropriate model, the plausibility of different hazard function shapes should be considered in light of:

- The shape of the *observed* hazard function (shown below).
- Known hazard shapes from standard parametric models (see Table).
- Relevant *external data* (not shown here).

Below is an overview of the potential hazard shapes for standard parametric survival models:

No	Distribution	Hazard function shape
1	Exponential	Constant hazard
2	Weibull	Monotonically increasing or decreasing hazards
3	Gompertz	Monotonically increasing or decreasing hazards
4	Log-normal	Arc-shaped or monotonically decreasing hazards
5	Log-logistic	Arc-shaped or monotonically decreasing hazards
6	Gamma	Monotonically increasing or decreasing hazards
7	Generalised Gamma	Arc-shaped, bathtub-shaped, monotonically increasing, or monotonically decreasing

Because **spline-based** and **(non-)mixture cure models** are derived from these standard parametric forms, their hazard functions can be interpreted in relation to the table above.

- For **spline-based models** using the *hazard scale*, the baseline hazard is Weibull, so hazards are monotonic within each spline segment (between knots).
- For **spline-based models** using the *odds scale* (log-logistic) or *normal scale* (log-normal), the baseline hazard follows the corresponding parametric form.
- For **(non-)mixture cure models**, the hazard shape applies separately to the cure and non-cure fractions.

For more details and visualisations of hazard function shapes across parametric survival models, see [Incerti \(2019\)](#).

Caution: These plots apply only to the *observed* data period and cannot directly inform the hazard shape beyond this range. Additionally, hazard estimates near the end of follow-up may be unstable due to sparse data. Interpret the tails of the curves with care.



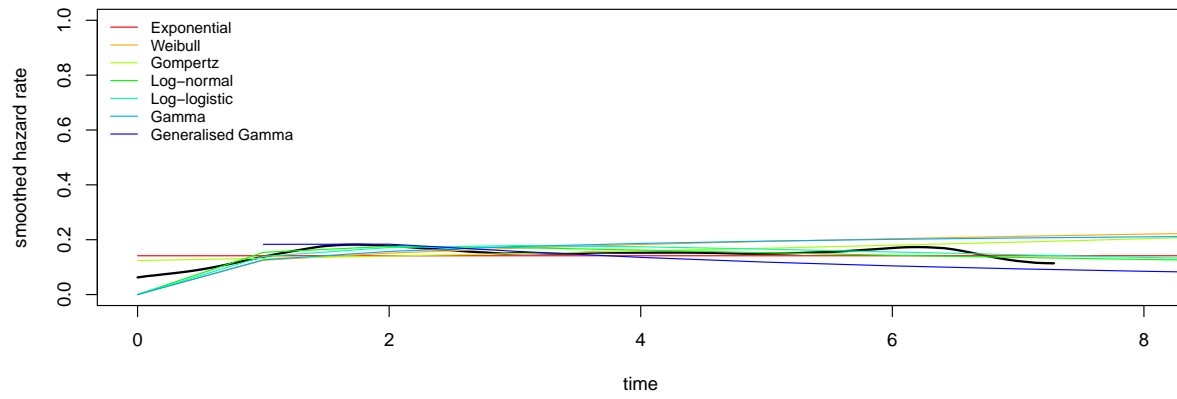
3.2 Shape of the predicted hazard function

The following plots show how the hazard rate is estimated under each fitted parametric survival model. These curves reflect the model-based *predicted* hazard functions over time, which may extend beyond the observed data period (see *Extrapolation Section*). By comparing these predicted hazard shapes, together with the known hazard function form for each parametric model, to:

- The **observed smoothed hazard functions**, and
- Prior knowledge or external evidence on plausible hazard behaviour,

you can assess whether each model's functional form provides a reasonable representation and extrapolation of the underlying event process.

Standard parametric models, Group: A



4 Parametric survival models

4.1 Standard parametric models

The Table below displays the goodness-of-fit statistics for each parametric survival model, ordered from ‘best’ fitting to less well fitting based on the Akaike Information Criterion (AIC). The AIC and Bayesian Information Criterion (BIC) provide a measure of the relative fit of each model to the observed data, while penalising for the number of parameters included in the fitted models. The lower the AIC or BIC, the better the relative fit of a model compared to other fitted models.

Rules of thumb for model selection:

- **Burnham and Anderson** (1998) suggest that models with an AIC within 4 points of the lowest AIC can be considered as having substantial support from the data [Burnham and Anderson, 1998](#).
- **Raftery** (1995) suggests that models with a BIC within 2 points of the lowest BIC can be considered the best-fitting models [Raftery et al., 1995](#).

In the following pages, three plots per fitted parametric survival model are displayed to support visual inspection of the fit of the models to the observed data:

- **Figure A:** Kaplan-Meier curves (black and grey) versus fitted parametric survival models (colour).
- **Figure B:** Diagnostic plot specific to each fitted parametric survival model (see graphical test in Table 1 of [Ishak et al., 2013](#)).
- **Figure C:** Comparison of smoothed hazard rates based on empirical data (black and grey) versus estimated transition probabilities (colour).

In Figures A and C, the Kaplan-Meier curves and smoothed hazard rates are identical across models; only the coloured fitted curves change.

The guiding rule is: the closer the coloured lines are to the black and grey lines, the better the fit.

Notably:

- Weibull (shape = 1), Gompertz (shape = 0) and gamma (shape = 1) distributions simplify to the exponential distribution.
- The generalised gamma distribution simplifies to:
 - Log-normal ($Q = 0$)
 - Weibull ($Q = 1$)
 - Exponential ($Q = 1$ and scale = 1)
 - Gamma ($Q = \text{scale}$)

Information regarding parameterisation of parametric survival models can be found [here](#).

CAUTION:

These goodness-of-fit statistics apply only to the *observed* data period. They do not indicate the suitability of extrapolated survival beyond this period. Additionally, the tail of the observed data period may be affected by a low number of observations, which should be considered when interpreting model performance.

Table 3: Goodness of fit statistics

Model	AIC	BIC
Generalised Gamma	1706	1720
Log-normal	1714	1723
Log-logistic	1732	1741
Gamma	1743	1752
Weibull	1750	1759
Gompertz	1769	1778
Exponential	1769	1774

Exponential

A: Kaplan–Meier (Exponential)



B: Diagnostic plot (Exponential)

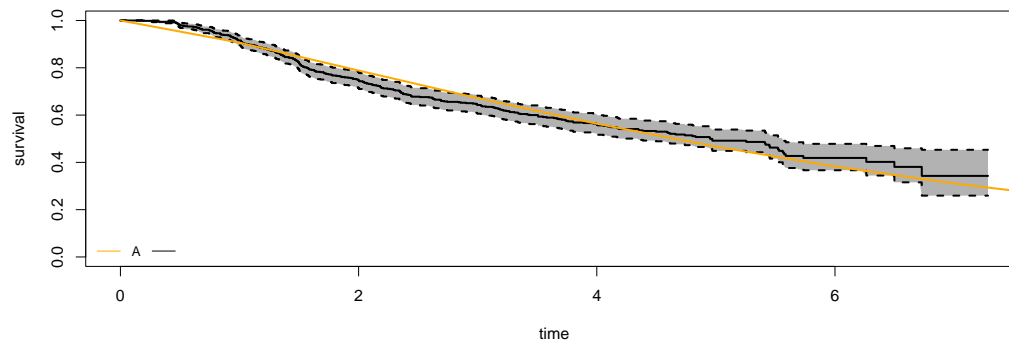


C: Annual transition probability (Exponential)

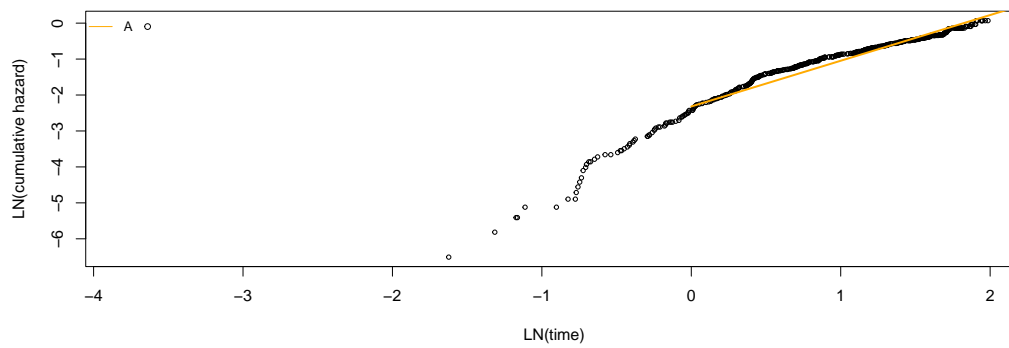


Weibull

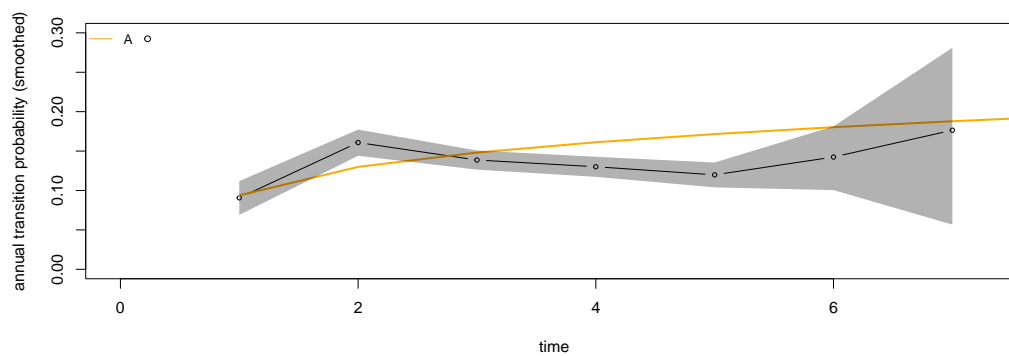
A: Kaplan–Meier (Weibull)



B: Diagnostic plot (Weibull)

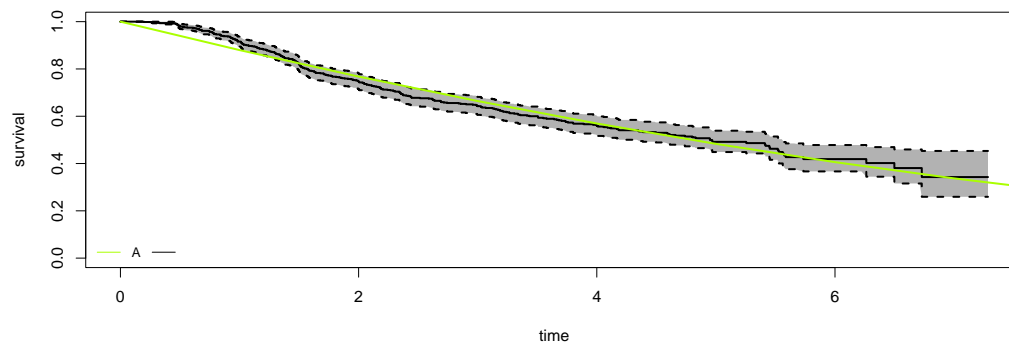


C: Annual transition probability (Weibull)

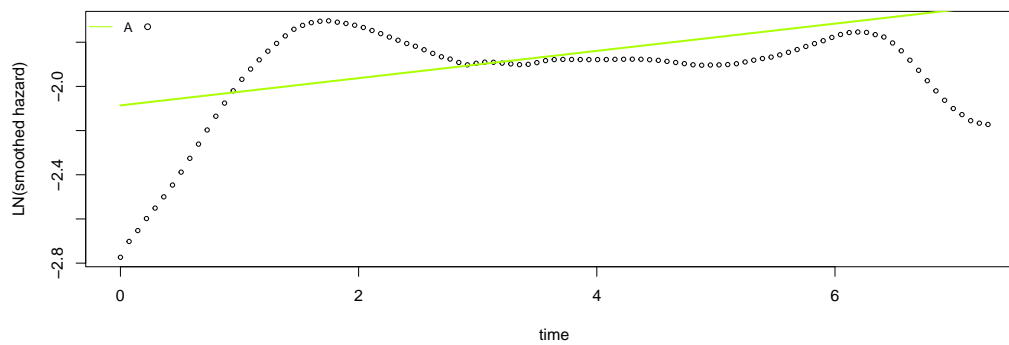


Gompertz

A: Kaplan-Meier (Gompertz)



B: Diagnostic plot (Gompertz)

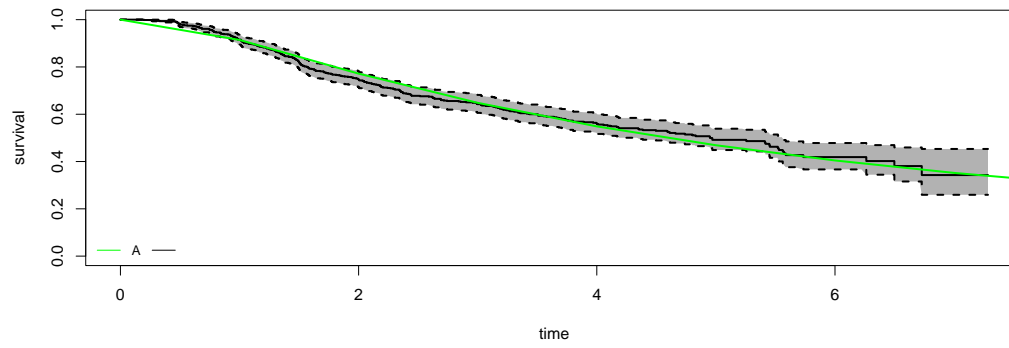


C: Annual transition probability (Gompertz)

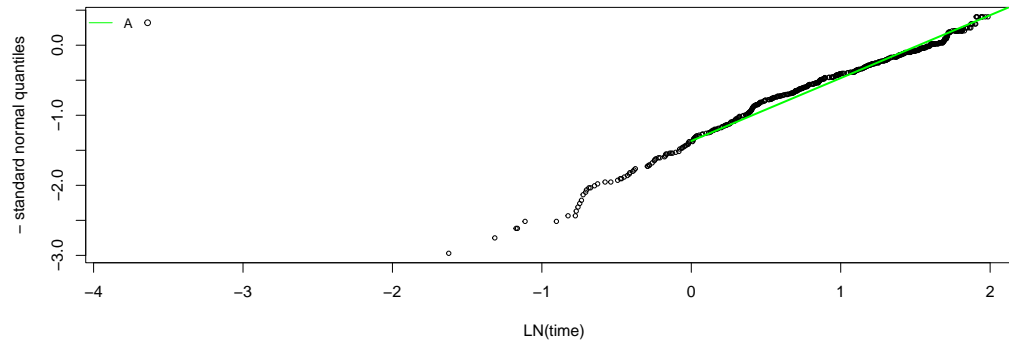


Log-normal

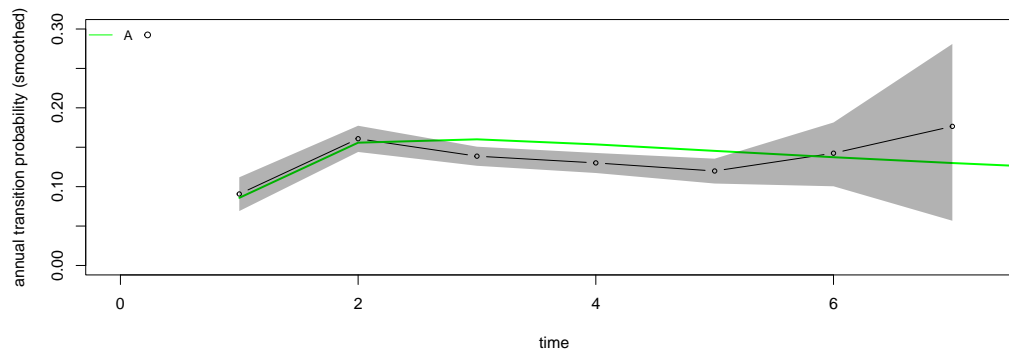
A: Kaplan-Meier (Log-normal)



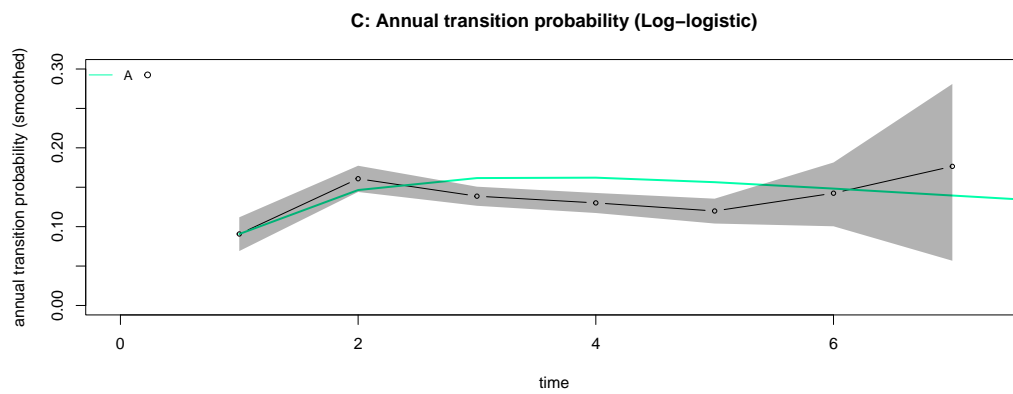
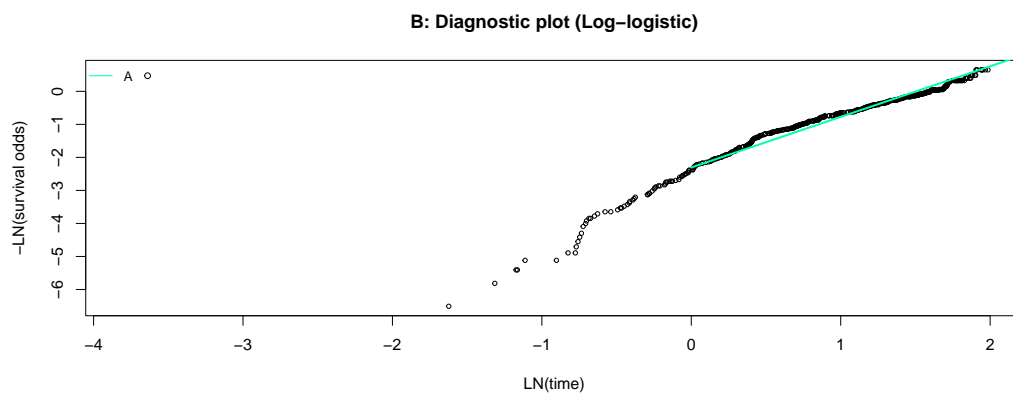
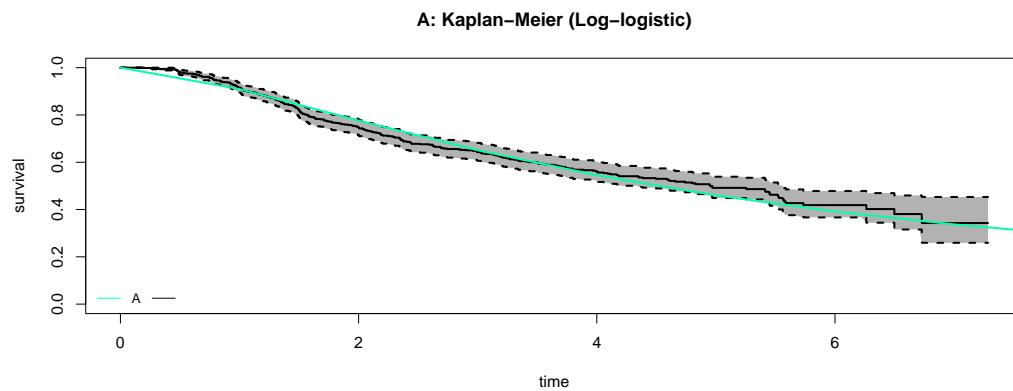
B: Diagnostic plot (Log-normal)



C: Annual transition probability (Log-normal)



Log-logistic



Gamma

A: Kaplan–Meier (Gamma)



B: Diagnostic plot (Gamma)



C: Annual transition probability (Gamma)



Generalised Gamma

A: Kaplan–Meier (Generalised Gamma)



B: Diagnostic plot (Generalised Gamma)



C: Annual transition probability (Generalised Gamma)



4.2 Parametric natural cubic spline models

If standard parametric models do not provide a satisfactory fit to the data (based on the visual assessments and goodness-of-fit statistics presented in the previous section or extrapolations presented in the next section), spline-based models may provide a more flexible alternative. An explanation of these natural cubic spline models, henceforth referred to as spline-based models, is provided in [Royston and Parmar \(2002\)](#).

Model complexity and parsimony

Following the principle of *Occam's Razor* and the recommendations in [NICE TSD 21](#), the preferred approach is to select the simplest model (e.g., a standard parametric model) that provides an adequate fit to the data and clinically plausible extrapolations. However, when simpler models do not provide a satisfactory fit, particularly in terms of long-term extrapolations and consistency with external evidence, more flexible models, such as spline-based models, can be considered.

Relationship to standard parametric models

Spline-based models extend standard parametric survival models by allowing the hazard, odds, or normal scale to vary flexibly over time:

- **Hazard scale** spline models extend the Weibull distribution.
- **Odds scale** spline models extend the log-logistic distribution.
- **Normal scale** spline models extend the log-normal distribution.

Model assessment and rules of thumb

The same goodness-of-fit statistics (AIC and BIC) used in Section 4.1 are presented here, and the same *rules of thumb* apply:

- **Burnham & Anderson (1998)**: Models within 4 AIC points of the lowest AIC are considered to have substantial support from the data [Burnham and Anderson, 1998](#).
- **Raftery (1995)**: Models within 2 BIC points of the lowest BIC are considered among the best-fitting models [Raftery et al., 1995](#).

Plots

As in Section 4.1, the following pages display three plots for each fitted spline-based survival model:

- **Figure A**: Kaplan-Meier curves (black and grey) versus the fitted spline model (colour).
- **Figure B**: Model-specific diagnostic plots (see [Ishak et al., 2013](#)).
- **Figure C**: Smoothed hazard rates from empirical data (black and grey) versus model-based estimates (colour).

CAUTION:

These statistics and plots apply only to the *observed* data period and do not directly inform the appropriateness of extrapolations beyond this period.

Additionally, the tail of the observed data may be affected by a low number of events, which should be considered when interpreting model performance.

4.3 Parametric (non-)mixture cure models

In some disease areas, a proportion of patients may be considered “cured,” meaning they have the same mortality risk as the general population (or a background hazard). Cure models explicitly allow for this possibility and estimate both the cure fraction and the survival of the uncured fraction.

- *Mixture cure models*: Assume the population is composed of a cured fraction and an uncured fraction, each with its own survival curve. The overall survival curve is a weighted mixture of the two. This means the survival curve approaches a horizontal asymptote at the estimated cure fraction.
- *Non-mixture cure models*: Directly model the hazard function so that it approaches zero over time for the cured fraction, without explicitly separating the two subpopulations in the survival function.

Both approaches can be fitted using standard parametric distributions (i.e., Weibull, log-normal or log-logistic).

The *link function* relates the linear predictor to the estimated cure fraction.

The most commonly used link function is the *logistic*, but others are available:

- "logistic": Ensures the cure fraction is between 0 and 1 via the logistic transformation.
- "loglog": Uses the complementary log-log link, which can be more appropriate if cure proportions are very close to 0 or 1.
- "identity": Fits the cure fraction directly on the probability scale (must remain in $[0,1]$).
- "probit": Uses the cumulative normal distribution as the link.

Model complexity and parsimony

Following *Occam's Razor* and [NICE TSD 21](#), standard parametric models should be preferred when they provide an adequate and clinically plausible fit. Cure models should be considered only when there is strong clinical or empirical evidence supporting the presence of a long-term cured fraction.

Model assessment and rules of thumb

The same goodness-of-fit statistics (AIC and BIC) used in Section 4.1 are presented here, and the same *rules of thumb* apply:

- **Burnham & Anderson (1998)**: Models within 4 AIC points of the lowest AIC are considered to have substantial support from the data [Burnham and Anderson, 1998](#).
- **Raftery (1995)**: Models within 2 BIC points of the lowest BIC are considered among the best-fitting models [Raftery et al., 1995](#).

Plots

As in Section 4.1, the following pages display three plots for each fitted spline-based survival model:

- **Figure A**: Kaplan-Meier curves (black and grey) versus the fitted spline model (colour).
- **Figure B**: Model-specific diagnostic plots (see [Ishak et al., 2013](#)).
- **Figure C**: Smoothed hazard rates from empirical data (black and grey) versus model-based estimates (colour).

CAUTION:

These statistics and plots apply only to the *observed* data period and do not directly inform the appropriateness of extrapolations beyond this period.

Additionally, the tail of the observed data may be affected by a low number of events, which should be considered when interpreting model performance.

5 Extrapolation

The plausibility of estimated survival beyond the observed data period is often a critical driver in health economic models. As highlighted in [NICE TSD 14](#) and [NICE TSD 21](#), model choice for extrapolation should consider whether the estimated long-term survival is consistent with **external data** (e.g., observational studies, registry data, or general population mortality).

In this section, extrapolation is examined by:

1. *Plotting survival curves* (Figures A) over the observed and extrapolated periods.
2. *Plotting annual transition probabilities* (Figures B) over the observed and extrapolated periods.
3. *Plotting hazard functions* (Figures C) over the observed and extrapolated periods.
4. *Tabulating survival probabilities* at multiple user-specified time points (first table).
5. *Summarising annual transition probabilities* (second table).

All outputs are presented *per group*.

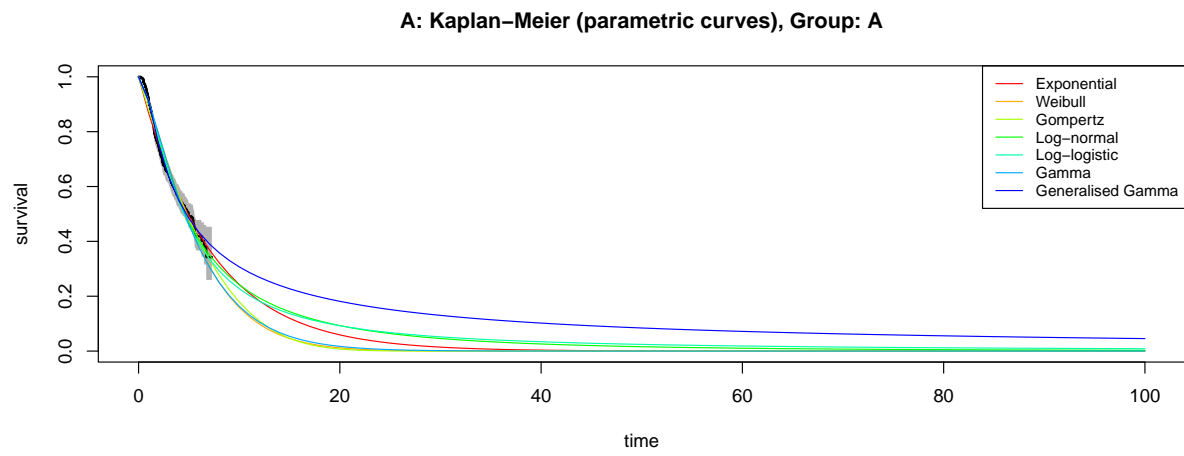
Assessing plausibility

- Compare survival probabilities at specific time points with *external data* (e.g., registry data, literature, expert opinion).
- Compare annual transition probabilities with *general population mortality* (e.g., from national life tables).
- If a model predicts conditional transition probabilities more favourable than the general population, this may be implausible unless clinically justified.
- In such cases, alternative models should be considered, or adjustments made (e.g., incorporating background mortality).

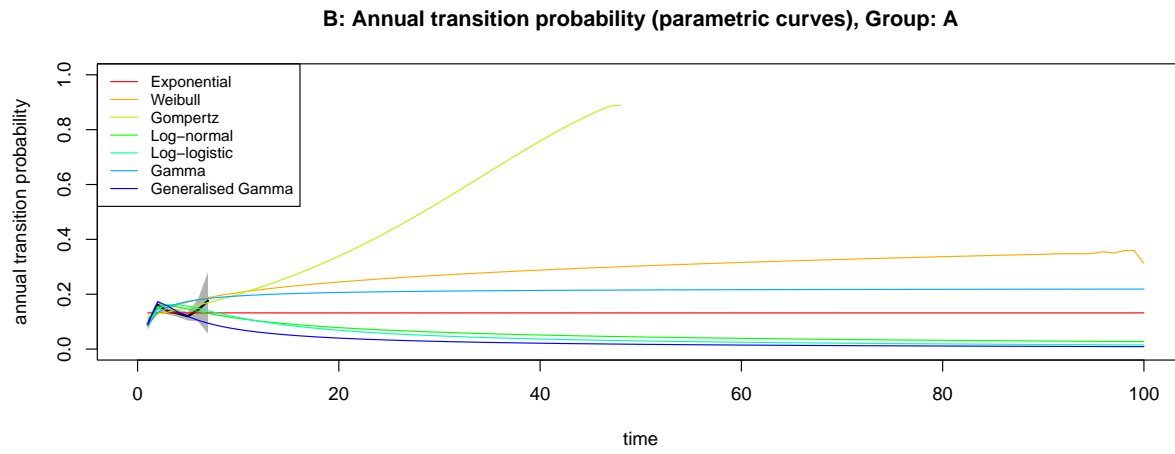
Note: These checks are consistent with the guidance in NICE TSD 14 and TSD 21, which recommend:

- Using *external evidence* to validate extrapolations.
- Keeping models as simple as possible (Occam's Razor), but adopting more flexible approaches when standard parametric models do not align with external evidence or long-term expectations.

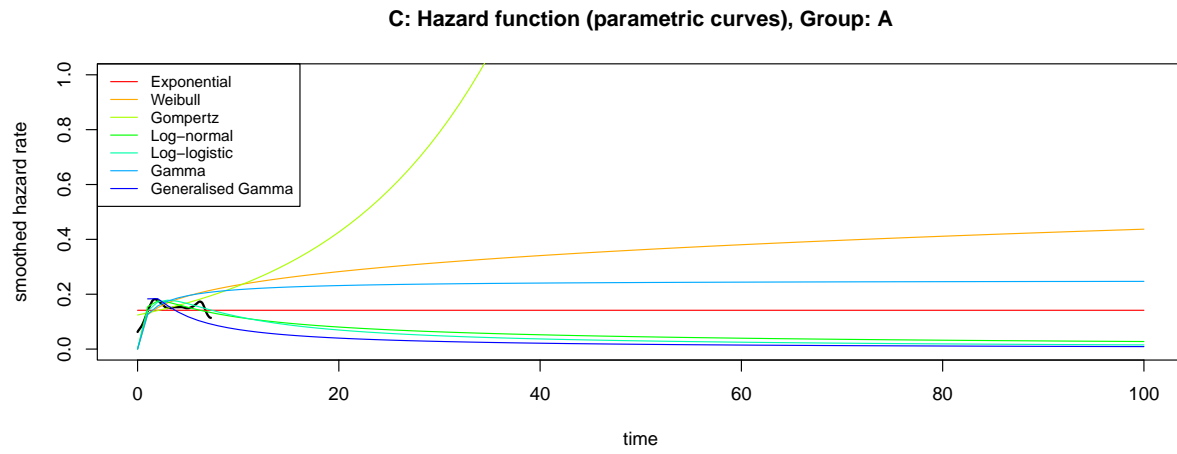
5.1 Extrapolated survival



5.2 Extrapolated transition probabilities



5.3 Extrapolated hazard function



5.4 Tabulated results

Group A

Table 4: Survival probability at different time points

	T= 0	T= 10	T= 20	T= 30	T= 40	T= 50	T= 60	T= 70	T= 80	T= 90	T= 100
Exponential	1	0.243	0.059	0.014	0.003	0.001	0.000	0.000	0.000	0.000	0.000
Weibull	1	0.159	0.012	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Gompertz	1	0.180	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Log-normal	1	0.242	0.093	0.046	0.026	0.016	0.010	0.007	0.005	0.004	0.003
Log-logistic	1	0.227	0.092	0.052	0.034	0.024	0.019	0.015	0.012	0.010	0.009
Gamma	1	0.163	0.017	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Generalised Gamma	1	0.307	0.181	0.130	0.102	0.084	0.072	0.063	0.056	0.050	0.046

Table 5: Summary statistics of annual transition probabilities

	Mean	Std.Dev	Min	Q1	Median	Q3	Max	IQR
Exponential	0.132	0.000	0.132	0.132	0.132	0.132	0.132	0.000
Weibull	0.289	0.056	0.094	0.260	0.304	0.331	0.360	0.072
Gompertz	0.459	0.250	0.120	0.232	0.421	0.676	0.889	0.444
Log-normal	0.056	0.032	0.027	0.033	0.044	0.068	0.160	0.034
Log-logistic	0.046	0.038	0.015	0.020	0.029	0.055	0.162	0.035
Gamma	0.209	0.018	0.087	0.210	0.215	0.218	0.219	0.008
Generalised Gamma	0.030	0.032	0.009	0.012	0.017	0.032	0.173	0.020

6. PERSUADE object information

```
## List of 7
## $ name          : chr "/tests/testthat/grid_test/check_9"
## $ input          :List of 11
## ..$ years       : num [1:686] 3.68 4.32 4.82 3.16 2.65 ...
## ..$ status      : num [1:686] 0 0 0 0 0 0 0 1 0 0 ...
## ..$ group       : Factor w/ 1 level "A": 1 1 1 1 1 1 1 1 1 1 ...
## ..$ strata      : logi FALSE
## ..$ spline_mod  : logi FALSE
## ..$ cure_mod    : logi FALSE
## ..$ cure_link   : logi NA
## ..$ time_unit   : num 1
## ..$ time_horizon : num 100
## ..$ time_pred_surv_table: num [1:11] 0 10 20 30 40 50 60 70 80 90 ...
## ..$ time_pred   : num [1:101] 0 1 2 3 4 5 6 7 8 9 ...
## $ surv_obs       :List of 6
## ..$ km          :List of 23
## .. ..- attr(*, "class")= chr [1:2] "npsurv" "survfit"
## ..$ km_names: num [1:574] 1 1 1 1 1 1 1 1 1 1 ...
## ..$ cum_haz : 'data.frame': 8 obs. of 9 variables:
## ..$ haz      :List of 3
## ..$ tp       :List of 2
## ..$ cox_reg  :List of 13
## .. ..- attr(*, "class")= chr [1:2] "coxph.null" "coxph"
## $ surv_model   :List of 2
## ..$ param_models:List of 7
## ..$ param_ic    : 'data.frame': 7 obs. of 3 variables:
## $ surv_pred     :List of 3
## ..$ model:List of 7
## ..$ gr         :List of 1
## ..$ tp_gr:List of 1
## $ surv_model_excel: 'data.frame': 12 obs. of 14 variables:
## ..$ rate      : chr [1:12] "expo" "rate" "-1.95562140689451" "-2.06896905124845" ...
## ..$ shape     : chr [1:12] "weib" "shape" "0.240212591277917" "0.142689503677317" ...
## ..$ scale     : chr [1:12] "weib" "scale" "1.82315748068365" "1.72351677506703" ...
## ..$ shape     : chr [1:12] "gom" "shape" "0.0616918163542825" "-0.0117272170838677" ...
## ..$ rate      : chr [1:12] "gom" "rate" "-2.08602448550314" "-2.28249227304772" ...
## ..$ meanlog   : chr [1:12] "lnorm" "meanlog" "1.52256312523025" "1.41193348418067" ...
## ..$ sdlog     : chr [1:12] "lnorm" "sdlog" "0.10779065830448" "0.0199566340454587" ...
## ..$ shape     : chr [1:12] "llog" "shape" "0.426887335363394" "0.331473259645795" ...
## ..$ scale     : chr [1:12] "llog" "scale" "1.50463029487844" "1.40035262576969" ...
## ..$ shape     : chr [1:12] "gam" "shape" "0.384497657236819" "0.248153288503327" ...
## ..$ rate      : chr [1:12] "gam" "rate" "-1.38079206032023" "-1.59361247119485" ...
## ..$ mu        : chr [1:12] "ggam" "mu" "1.18024455999323" "0.920871638954969" ...
## ..$ sigma     : chr [1:12] "ggam" "sigma" "0.221931058018601" "0.132033489771379" ...
## ..$ Q         : chr [1:12] "ggam" "Q" "-0.836327519386724" "-1.35461791241774" ...
## $ misc          :List of 5
## ..$ form       :Class 'formula' language survival::Surv(years, status) ~ 1
## .. ..- attr(*, ".Environment")=<environment: 0x00000239b2c89f58>
## ..$ group_names: chr "A"
## ..$ ngroups    : int 1
## ..$ lbls       : chr [1:7] "Exponential" "Weibull" "Gompertz" "Log-normal" ...
## ..$ cols_tp    : num 8
```

```
## - attr(*, "class")= chr "PERSUADE"  
## NULL
```

7. Session information

```
## R version 4.5.1 (2025-06-13 ucrt)
## Platform: x86_64-w64-mingw32/x64
## Running under: Windows 11 x64 (build 26100)
##
## Matrix products: default
## LAPACK version 3.12.1
##
## Random number generation:
## RNG:      Mersenne-Twister
## Normal:   Inversion
## Sample:   Rejection
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods    base
##
## other attached packages:
## [1] PERSUADE_0.0.0.9000 testthat_3.2.3
##
## loaded via a namespace (and not attached):
## [1] splines_4.5.1      later_1.4.4         polspline_1.1.25    bitops_1.0-9        tibble_3.3.0
## [7] deSolve_1.40       lifecycle_1.0.4     rstatix_0.7.2       rprojroot_2.1.1     processx_3.8.6
## [13] MASS_7.3-65       backports_1.5.0     magrittr_2.0.3      sass_0.4.10         Hmisc_5.2-3
## [19] jquerylib_0.1.4    yaml_2.3.10         remotes_2.5.0       httpuv_1.6.16       flexsurvcure_1
## [25] pkgbuild_1.4.8     RColorBrewer_1.1-3  multcomp_1.4-28     abind_1.4-8         pkgload_1.4.0
## [31] purrr_1.1.0        RCurl_1.98-1.17     nnet_7.3-20         pracma_2.4.4        TH.data_1.1-4
## [37] KMsurv_0.1-6       MatrixModels_0.5-4  svglite_2.2.1       commonmark_2.0.0    codetools_0.2-7
## [43] xml2_1.4.0         tidyselect_1.2.1    SuppDists_1.1-9.9   farver_2.1.2        base64enc_0.1-3
## [49] jsonlite_2.0.0     ks_1.15.1           ellipsis_0.3.2      Formula_1.2-5       survival_3.8-3
## [55] tools_4.5.1        Rcpp_1.1.0          glue_1.8.0          gridExtra_2.3        xfun_0.52
## [61] dplyr_1.1.4        withr_3.0.2         numDeriv_2016.8-1.1 muhaz_1.2.6.4       fastmap_1.2.0
## [67] xopen_1.0.1        mstate_0.3.3        SparseM_1.84-2      litedown_0.7        callr_3.7.6
## [73] rcmdcheck_1.4.0    R6_2.6.1            mime_0.13           textshaping_1.0.2   colorspace_2.1
## [79] tidyr_1.3.1        generics_0.1.4      data.table_1.17.8   prettyunits_1.2.0   htmlwidgets_1.
## [85] gtable_0.3.6       pcaPP_2.0-5         survMisc_0.5.6      brio_1.1.5          htmltools_0.5.
## [91] profvis_0.4.0      scales_1.4.0        kableExtra_1.4.0    knitr_1.50          km.ci_0.5-6
## [97] checkmate_2.3.3    nlme_3.1-168        curl_7.0.0          cachem_1.1.0        zoo_1.8-14
## [103] stringr_1.5.1      KernSmooth_2.23-26  miniUI_0.1.2        fda_6.3.0           foreign_0.8-90
## [109] pillar_1.11.0      grid_4.5.1          vctr_0.6.5          urlchecker_1.0.1    promises_1.3.3
## [115] car_3.1-3          xtable_1.8-4        cluster_2.1.8.1     waldo_0.6.2         htmlTable_2.4.
## [121] tinytex_0.57       mvtnorm_1.3-3       cli_3.6.5           compiler_4.5.1      rlang_1.1.6
## [127] ggsignif_0.6.4     labeling_0.4.3      survminer_0.5.1     mclust_6.1.1        fds_1.8
## [133] fs_1.6.6           stringi_1.8.7       rainbow_3.8         viridisLite_0.4.2   hdrdce_3.4
## [139] devtools_2.4.5     quantreg_6.1        Matrix_1.7-3        ggplot2_3.5.2       statmod_1.5.0
## [145] flexsurv_2.3.2     gridtext_0.1.5      broom_1.0.9         memoise_2.0.1       bslib_0.9.0
```