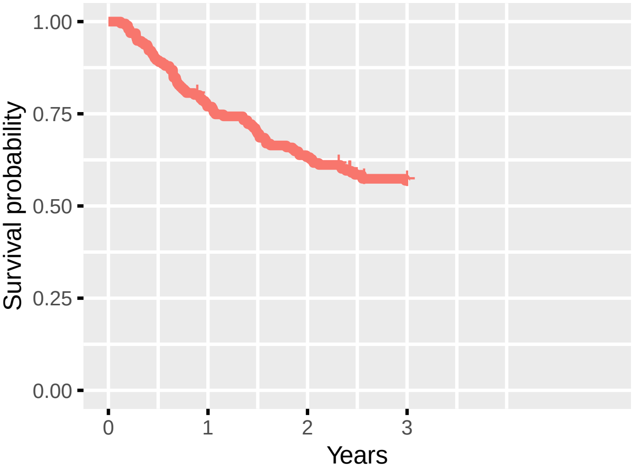
Survival extrapolation with external data: a new Bayesian model and R package

Christopher Jackson

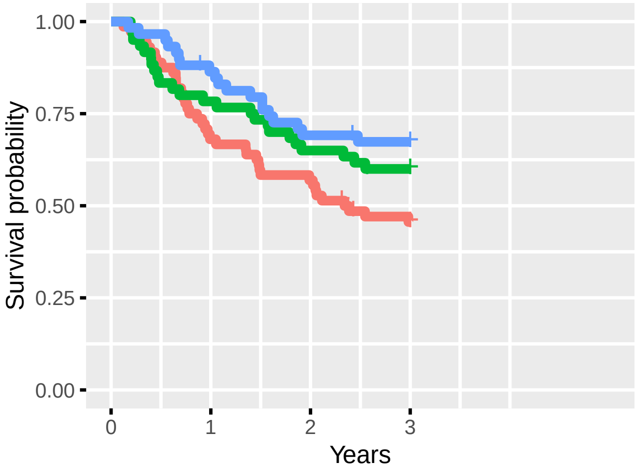
ESMI Seminar, Exeter, 18 October 2022

MRC Biostatistics Unit
UNIVERSITY OF CAMBRIDGE

Right-censored survival data



Right-censored survival data



Informing policy with censored time-to-event data

Should health service adopt a new treatment

- given 3 years of follow-up data from a clinical trial?

What is the predicted burden on hospitals in a epidemic?

- needs information on length of stay in hospital
- data from first 2 months of the epidemic, with many people still in hospital?

Consequences of policy decisions will last longer than the end of the data! Need longer-term judgements.

Informing policy with censored time-to-event data

Should health service adopt a new treatment

- given 3 years of follow-up data from a clinical trial?

What is the predicted burden on hospitals in a epidemic?

- needs information on length of stay in hospital
- data from first 2 months of the epidemic, with many people still in hospital?

Consequences of policy decisions will last longer than the end of the data! Need longer-term judgements.

Informing policy with censored time-to-event data

Should health service adopt a new treatment

- given 3 years of follow-up data from a clinical trial?

What is the predicted burden on hospitals in a epidemic?

- needs information on length of stay in hospital
- data from first 2 months of the epidemic, with many people still in hospital?

Consequences of policy decisions will last longer than the end of the data! Need longer-term judgements.

Long-term estimand

Policy-maker typically wants to know the expected time to event

- equivalent to knowing the total outcome (e.g. survival, hospital length of stay) over the population
- Not provided by most common survival analysis tools e.g. Kaplan-Meier estimators, Cox models.
- Provided by a fully-parametric distribution for the time T to the event.

  Many choices for how to specify these

Long-term estimand

Policy-maker typically wants to know the expected time to event

- equivalent to knowing the total outcome (e.g. survival, hospital length of stay) over the population
- Not provided by most common survival analysis tools e.g. Kaplan-Meier estimators, Cox models.
- Provided by a fully-parametric distribution for the time T to the event.

  Many choices for how to specify these

Long-term estimand

Policy-maker typically wants to know the expected time to event

- equivalent to knowing the total outcome (e.g. survival, hospital length of stay) over the population
- Not provided by most common survival analysis tools e.g. Kaplan-Meier estimators, Cox models.
- Provided by a fully-parametric distribution for the time T to the event.

  Many choices for how to specify these

Long-term estimand

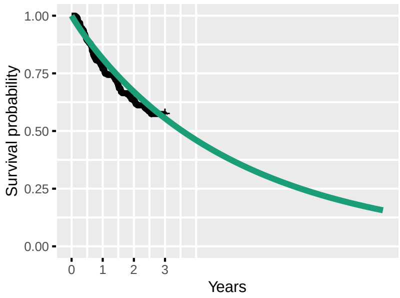
Policy-maker typically wants to know the expected time to event

- equivalent to knowing the total outcome (e.g. survival, hospital length of stay) over the population
- Not provided by most common survival analysis tools e.g. Kaplan-Meier estimators, Cox models.
- Provided by a fully-parametric distribution for the time T to the event.
  Many choices for how to specify these

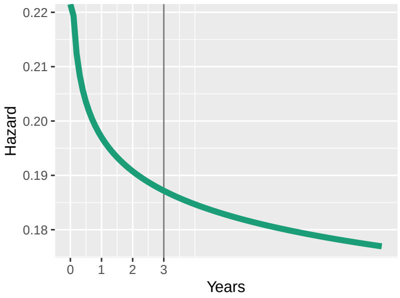
Parametric survival models

For example, Weibull distribution

Survivor function
$S(t|\lambda,\alpha) = \exp(-\lambda t^{\alpha})$



Hazard function
$h(t|\lambda,\alpha) = \lambda\alpha t^{\alpha-1}$

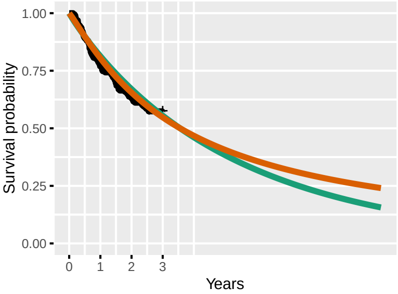


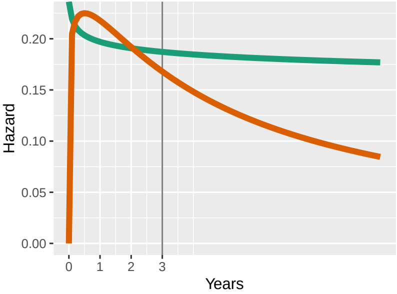
Parametric survival models

For example, Log-logistic distribution

Survivor function
$S(t|\lambda,\alpha) = 1/(1+(t/b)^a)$

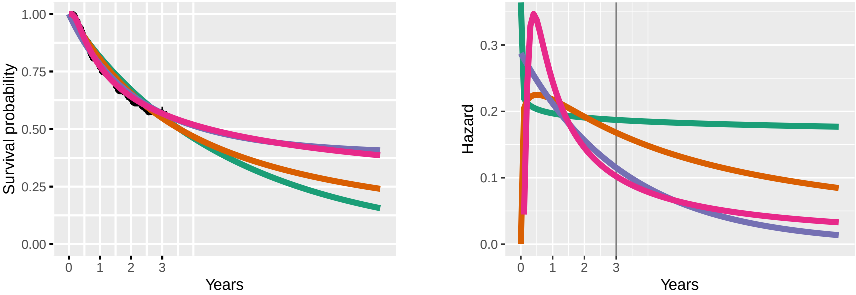
Hazard function $h(t|\lambda,\alpha) =$





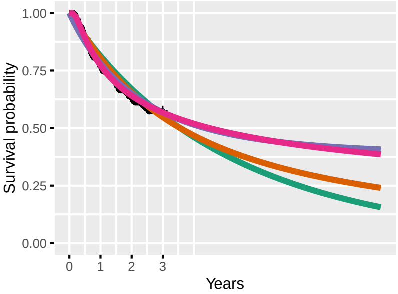
Parametric survival models

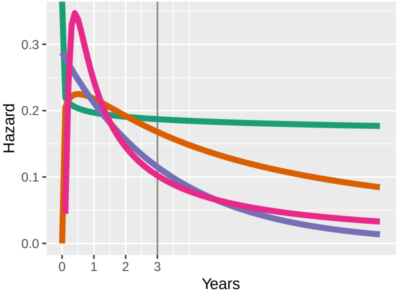
For example,
Typical set of models in standard software



Parametric survival models

For example,
Typical set of models in standard software





Fit a set of models and judge which :

- fits the data best in the short term? (easy, e.g. AIC)
- gives more plausible risk changes in the long term? harder

Where might information about the long term come from?

Clinical judgements about the mechanism

- e.g. some people get cured of the disease, so disease-specific hazard of death reduces to zero

Data on general population, or disease registry

- survival of people with a specific disease cannot be better than a comparable set of people in the general population?

Formally-elicited judgements about, e.g. 5-year, 10-year survival?

Want to be able to use this kind of information in a transparent and statistically-principled way

Recent methods reviews

- NICE DSU (2021) `https://www.sheffield.ac.uk/nice-dsu/tsds/flexible-methods-survival`

  NICE Decision Support Unit

  Training Appraisal specific Methods development TSDs Publications About

  Home › NICE Decision Support Unit › Technical support documents › **Flexible methods for survival analysis TSD**

- Jackson et al. (Medical Decision Making 2017)

  Extrapolating Survival from Randomized Trials Using External Data: A Review of Methods

  Christopher Jackson, PhD, John Stevens, PhD, Shijie Ren, MPhil, PhD, Nick Latimer, PhD, MSc, Laura Bojke, PhD, MSc, Andrea Manca, PhD, MSc, Linda Sharples, PhD Show less

  First Published July 10, 2016 | Research Article | Find in PubMed | Check for updates

  https://doi-org.ezp.lib.cam.ac.uk/10.1177/0272989X16639900

- piecewise models, spline models, cure, relative survival, proportional and additive hazards, converging hazards, diminishing treatment effects, Bayesian methods ...

What is missing is an easily usable tool in which to build different models with useful ranges of assumptions, in particular for Bayesian methods.

Ideal characteristics of a method / tool

1. Incorporate all available data
2. Fit the data as well as possible
3. Make any assumptions transparent
4. Quantify uncertainty
5. Be easy to use!

Ideal characteristics of a method / tool

1. Incorporate all available data
2. Fit the data as well as possible
3. Make any assumptions transparent
4. Quantify uncertainty
5. Be easy to use!

Ideal characteristics of a method / tool

1. Incorporate all available data
2. Fit the data as well as possible
3. Make any assumptions transparent
4. Quantify uncertainty
5. Be easy to use!

Ideal characteristics of a method / tool

1. Incorporate all available data
2. Fit the data as well as possible
3. Make any assumptions transparent
4. Quantify uncertainty
5. Be easy to use!

Ideal characteristics of a method / tool

1. Incorporate all available data
2. Fit the data as well as possible
3. Make any assumptions transparent
4. Quantify uncertainty
5. Be easy to use!

survextrap R package

https://chjackson.github.io/survextrap

Infrastructure based on

- multiparameter evidence synthesis, Bayesian inference
- probabilistic programming (MCMC and Stan)
- flexible / sensible default modelling choices

... hopefully allowing extension in future to wide range of survival and multi-state modelling situations

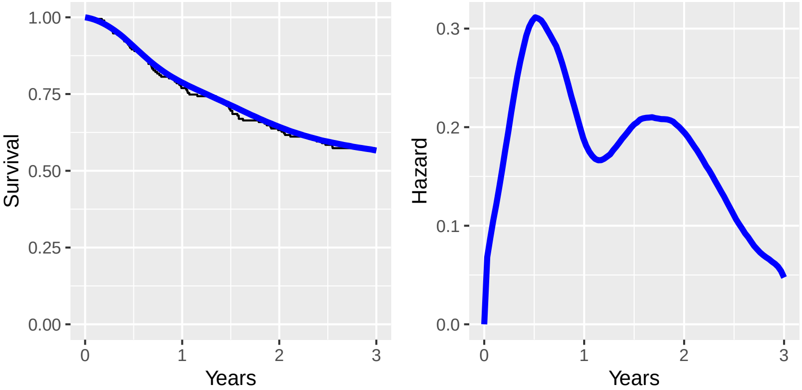
Ideal characteristics of a method / tool

1. Incorporate all available data
2. Fit the data as well as possible
3. Make any assumptions transparent
4. Quantify uncertainty
5. Be easy to use!

M-spline models to fit the individual data

Easiest of the challenges, for the individual data

- flexible parametric statistical models



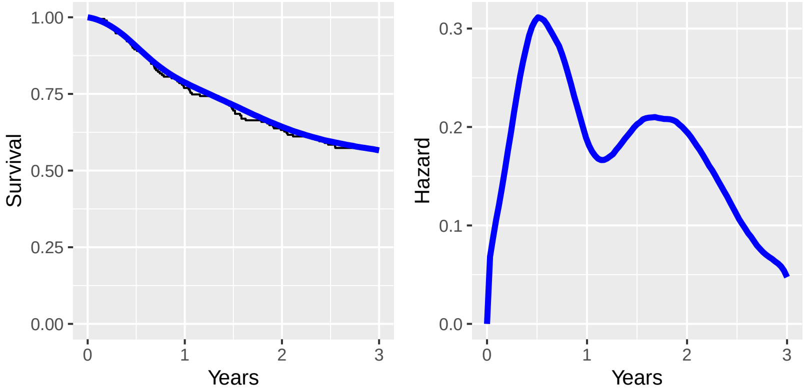
Hazard modelled with M-spline $h(t) = \eta \sum_{k=1}^{K} p_k b_k(t)$

inspired by the `https://mc-stan.org/rstanarm` package

M-spline models to fit the individual data

Easiest of the challenges, for the individual data

- flexible parametric statistical models



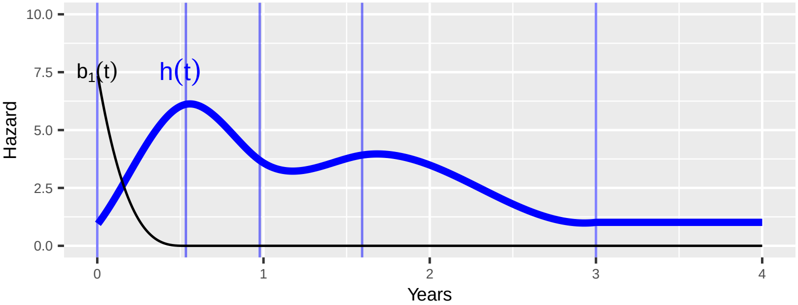
Hazard modelled with M-spline $h(t) = \eta \sum_{k=1}^{K} p_k b_k(t)$

inspired by the `https://mc-stan.org/rstanarm` package

M-spline models for hazard functions

M-spline $h(t) = \eta \sum_{k=1}^{K} p_k b_k(t)$

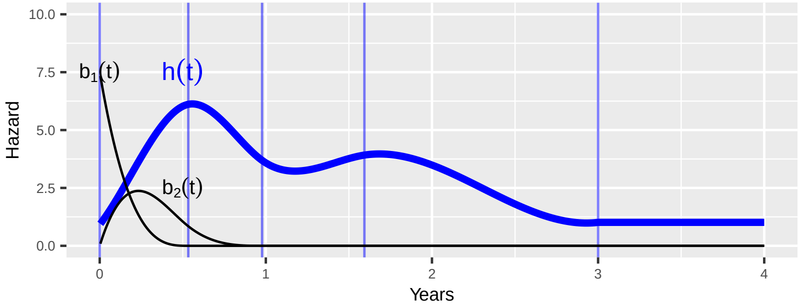
- Axis of time split into different periods defined by knots
- Weighted sum of basis functions that describe hazard in each period. Resulting $h(t)$ smooth.
- Use lots of knots, and Bayesian estimation with a prior that controls overfitting (shrink towards a constant hazard)



Covariates by proportional hazards (linear on $\log(\eta)$), or stratify, or flexible nonproportional hazards (put linear models on the p_k)

M-spline $h(t) = \eta \sum_{k=1}^{K} p_k b_k(t)$

- Axis of time split into different periods defined by knots
- Weighted sum of basis functions that describe hazard in each period. Resulting $h(t)$ smooth.
- Use lots of knots, and Bayesian estimation with a prior that controls overfitting (shrink towards a constant hazard)

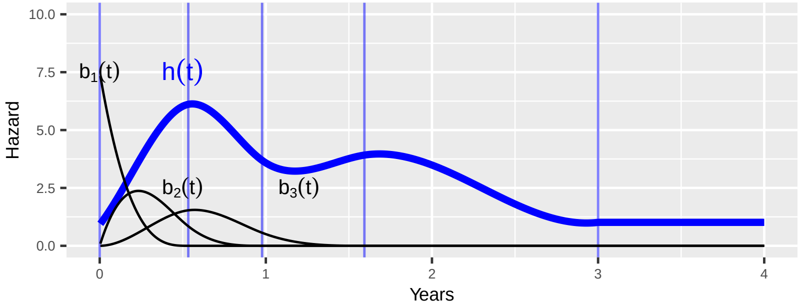


Covariates by proportional hazards (linear on $\log(\eta)$), or stratify, or flexible nonproportional hazards (put linear models on the p_k)

M-spline models for hazard functions

M-spline $h(t) = \eta \sum_{k=1}^{K} p_k b_k(t)$

- Axis of time split into different periods defined by knots
- Weighted sum of basis functions that describe hazard in each period. Resulting $h(t)$ smooth.
- Use lots of knots, and Bayesian estimation with a prior that controls overfitting (shrink towards a constant hazard)

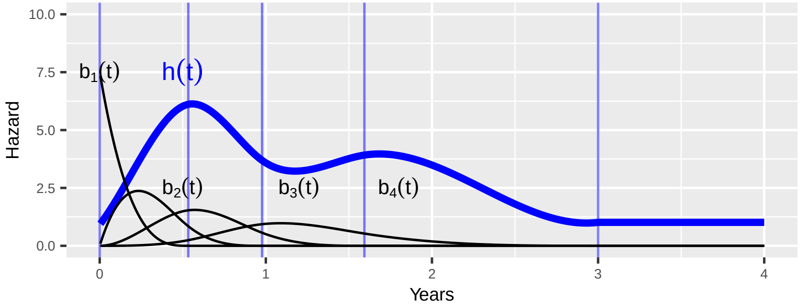


Covariates by proportional hazards (linear on $\log(\eta)$), or stratify, or flexible nonproportional hazards (put linear models on the p_k)

M-spline models for hazard functions

M-spline $h(t) = \eta \sum_{k=1}^{K} p_k b_k(t)$

- Axis of time split into different periods defined by knots
- Weighted sum of basis functions that describe hazard in each period. Resulting $h(t)$ smooth.
- Use lots of knots, and Bayesian estimation with a prior that controls overfitting (shrink towards a constant hazard)

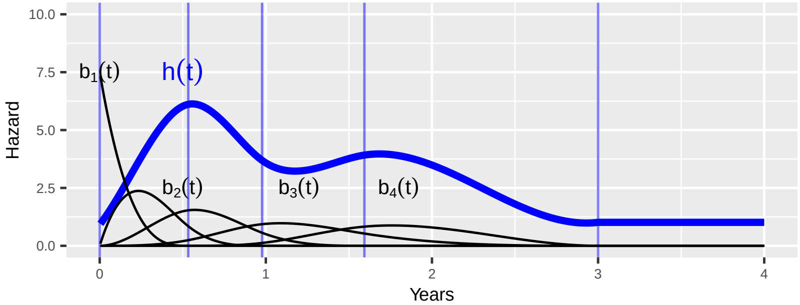


Covariates by proportional hazards (linear on $\log(\eta)$), or stratify, or flexible nonproportional hazards (put linear models on the p_k)

M-spline models for hazard functions

M-spline $h(t) = \eta \sum_{k=1}^{K} p_k b_k(t)$

- Axis of time split into different periods defined by knots
- Weighted sum of basis functions that describe hazard in each period. Resulting $h(t)$ smooth.
- Use lots of knots, and Bayesian estimation with a prior that controls overfitting (shrink towards a constant hazard)

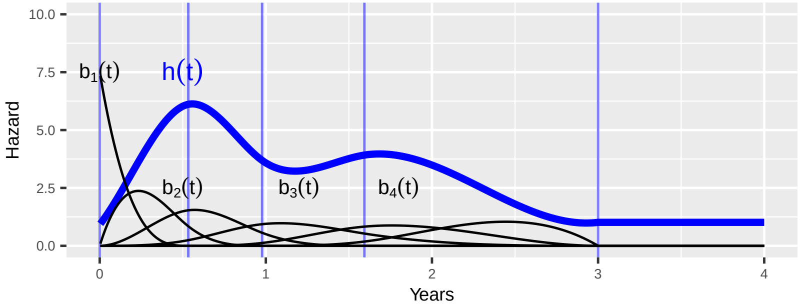


Covariates by proportional hazards (linear on $\log(\eta)$), or stratify, or flexible nonproportional hazards (put linear models on the p_k)

M-spline models for hazard functions

M-spline $h(t) = \eta \sum_{k=1}^{K} p_k b_k(t)$

- Axis of time split into different periods defined by knots
- Weighted sum of basis functions that describe hazard in each period. Resulting $h(t)$ smooth.
- Use lots of knots, and Bayesian estimation with a prior that controls overfitting (shrink towards a constant hazard)

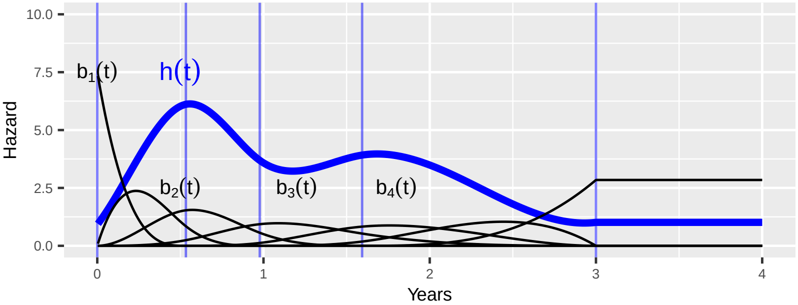


Covariates by proportional hazards (linear on $\log(\eta)$), or stratify, or flexible nonproportional hazards (put linear models on the p_k)

M-spline models for hazard functions

M-spline $h(t) = \eta \sum_{k=1}^{K} p_k b_k(t)$

- Axis of time split into different periods defined by knots
- Weighted sum of basis functions that describe hazard in each period. Resulting $h(t)$ smooth.
- Use lots of knots, and Bayesian estimation with a prior that controls overfitting (shrink towards a constant hazard)



Covariates by proportional hazards (linear on $\log(\eta)$), or stratify, or flexible nonproportional hazards (put linear models on the p_k)

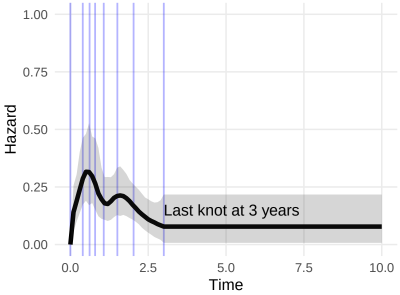
Ideal characteristics of a method / tool

1. Incorporate all available data
2. Fit the data as well as possible
3. Make any assumptions transparent
4. Quantify uncertainty
5. Be easy to use!

Uncertainty and extrapolation

Bayesian model gives posterior distribution of the hazard function

- Default: hazard assumed constant after last knot, here 3 years (end of follow up)
- Or: suppose we think hazard might change between 3 and 10 years. Place final knot at 10 years instead!
- If no data in long-term period: Bayesian model acknowledges uncertainty!

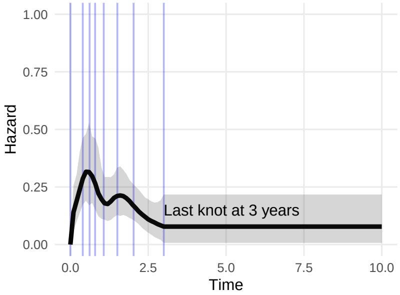


But there is rarely no data at all about the long term

Uncertainty and extrapolation

Bayesian model gives posterior distribution of the hazard function

- Default: hazard assumed constant after last knot, here 3 years (end of follow up)
- Or: suppose we think hazard might change between 3 and 10 years. Place final knot at 10 years instead!
- If no data in long-term period: Bayesian model acknowledges uncertainty!

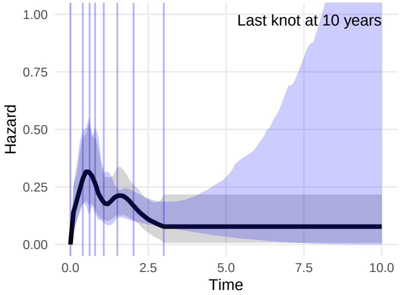


But there is rarely no data at all about the long term

Uncertainty and extrapolation

Bayesian model gives posterior distribution of the hazard function

- Default: hazard assumed constant after last knot, here 3 years (end of follow up)
- Or: suppose we think hazard might change between 3 and 10 years. Place final knot at 10 years instead!
- If no data in long-term period: Bayesian model acknowledges uncertainty!

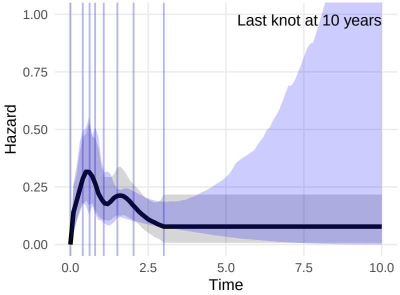


But there is rarely no data at all about the long term

Uncertainty and extrapolation

Bayesian model gives posterior distribution of the hazard function

- Default: hazard assumed constant after last knot, here 3 years (end of follow up)
- Or: suppose we think hazard might change between 3 and 10 years. Place final knot at 10 years instead!
- If no data in long-term period: Bayesian model acknowledges uncertainty!



But there is rarely no data at all about the long term

Ideal characteristics of a method / tool

1. Incorporate all available data
2. Fit the data as well as possible
3. Make any assumptions transparent
4. Quantify uncertainty
5. Be easy to use!

Jointly modelling all data: Bayesian evidence synthesis

External data, e.g. from matched population data, registry data, or elicited judgements?

Supply as aggregate counts of survival on any / multiple time periods

Follow-up period		Number		Covariates
Start time t	End time u	Alive at t	Still alive at u	
t_1	u_1	n_1	r_1	$\mathbf{x}_1$
t_2	u_2	n_2	r_2	$\mathbf{x}_2$
etc.				

- External and individual-level data jointly generated by the same spline survival distribution
- Any differences between data sources expressed through covariates, e.g. as proportional hazards

Jointly modelling all data: Bayesian evidence synthesis

External data, e.g. from matched population data, registry data, or elicited judgements?

Supply as aggregate counts of survival on any / multiple time periods

Follow-up period		Number		Covariates
Start time t	End time u	Alive at t	Still alive at u	
t_1	u_1	n_1	r_1	$\mathbf{x}_1$
t_2	u_2	n_2	r_2	$\mathbf{x}_2$
etc.				

- External and individual-level data jointly generated by the same spline survival distribution
- Any differences between data sources expressed through covariates, e.g. as proportional hazards

Jointly modelling all data: Bayesian evidence synthesis

External data, e.g. from matched population data, registry data, or elicited judgements?

Supply as aggregate counts of survival on any / multiple time periods

Follow-up period		Number		Covariates
Start time t	End time u	Alive at t	Still alive at u	
t_1	u_1	n_1	r_1	$\mathbf{x}_1$
t_2	u_2	n_2	r_2	$\mathbf{x}_2$
etc.				

- External and individual-level data jointly generated by the same spline survival distribution
- Any differences between data sources expressed through covariates, e.g. as proportional hazards

Ideal characteristics of a method / tool

1. Incorporate all available data
2. Fit the data as well as possible
3. Make any assumptions transparent
4. Quantify uncertainty
5. Be easy to use!

Are the assumptions transparent?

Foreground assumptions

- Data included: both short term and long term
- Range of knots: period in which we think the hazard might change
- Effects of covariates (e.g. proportional hazards)

Background "black box" assumptions

- Prior distributions
- Exact location of knots within that range
- Exact shape of spline curve

Needs work to develop software defaults that can be shown empirically to be unimportant

Despite this, we don't have to rely on opaque assumptions like e.g. "Weibull hazard trajectory is valid beyond the study data"

Are the assumptions transparent?

Foregrounded assumptions

- Data included: both short term and long term
- Range of knots: period in which we think the hazard might change
- Effects of covariates (e.g. proportional hazards)

Background "black box" assumptions

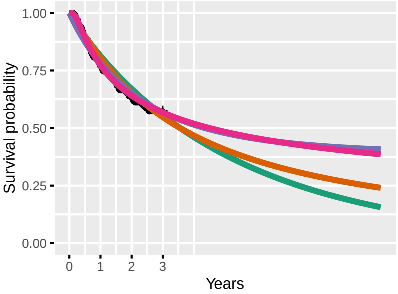
- Prior distributions
- Exact location of knots within that range
- Exact shape of spline curve

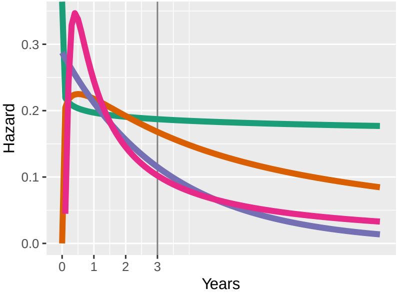
Needs work to develop software defaults that can be shown empirically to be unimportant

Despite this, we don't have to rely on opaque assumptions like e.g. "Weibull hazard trajectory is valid beyond the study data"

Parametric survival models

For example,
Typical set of models in standard software





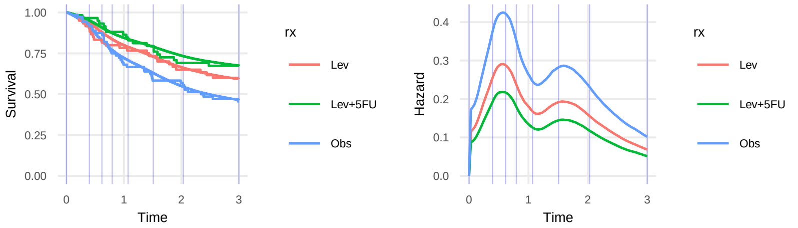
Ideal characteristics of a method / tool

1. Incorporate all available data
2. Fit the data as well as possible
3. Make any assumptions transparent
4. Quantify uncertainty
5. Be easy to use!

Using the survextrap R package

- Install from `http://chjackson.github.io/survextrap`
- Example: default spline model with proportional hazards between three treatment groups

```
library(survextrap)
mod <- survextrap(Surv(years, status) ~ rx, data=colons)
plot(mod)
```



Using the survextrap R package

Restricted mean survival time over 20 years

```r
rmst(mod, t=20)
```

```
##          rx variable  t median 2.5% 97.5%
## 1       Obs     rmst 20   5.66 3.35  9.52
## 2       Lev     rmst 20   8.19 4.55 13.12
## 3 Lev+5FU     rmst 20  10.16 5.61 14.49
```

Hazard ratios and other parameter estimates

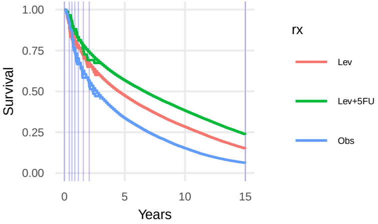
```r
summary(mod)
```

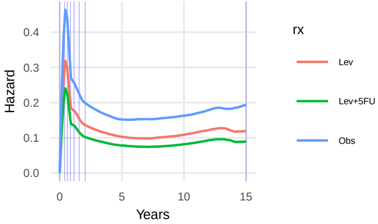
```
## # A tibble: 16 x 9
##    variable  basis_num term        median    lower   upper      sd  rhat ess_bulk
##    <chr>         <dbl> <chr>        <dbl>    <dbl>   <dbl>   <dbl> <dbl>    <dbl>
##  1 alpha            NA <NA>      -0.266   -0.592    0.0314 0.161    1.00    1630.
##  2 coefs             1 <NA>       0.0222   0.00586  0.0385 0.00877  1.00    1197.
##  3 coefs             2 <NA>       0.0314   0.00131  0.0976 0.0261   1.00    1337.
##  4 coefs             3 <NA>       0.0926   0.0119   0.235  0.0577   1.00    1619.
##  5 coefs             4 <NA>       0.160    0.0307   0.310  0.0707   1.00    1755.
## ...
## 11 coefs            10 <NA>       0.0328   0.00276  0.0921 0.0238   1.00    1395.
## 12 loghr            NA rxLev     -0.381   -0.898    0.0992 0.254    1.00    1952.
## 13 loghr            NA rxLev+5FU -0.666   -1.22    -0.140  0.281    1.00    1811.
## 14 hr               NA rxLev      0.683    0.407    1.10   0.178    1.00    1952.
## 15 hr               NA rxLev+5FU  0.514    0.295    0.870  0.151    1.00    1811.
## 16 smooth_sd        NA <NA>       0.592    0.174    1.86   0.417    1.00     562.
```

Using survextrap with external data

Example: we observe 40/100 survivors between 10 and 15 years in an external population like those in the control group (`rx="Obs"`)

```
extdat <- data.frame(start=10, stop=15,
                     n=100, r=40, rx = "Obs")
mod <- survextrap(Surv(years, status) ~ rx, data=colons,
                  external=extdat)
plot(mod, xlab="Years")
```



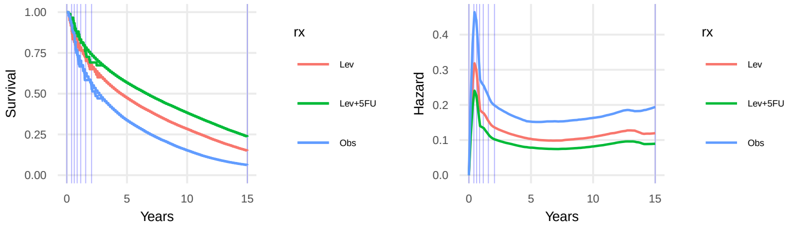


Hazard from 10-15 has been "pulled up" by inclusion of external data

Using survextrap with external data

Example: we observe 40/100 survivors between 10 and 15 years in an external population like those in the control group (`rx="Obs"`)

```
extdat <- data.frame(start=10, stop=15,
                     n=100, r=40, rx = "Obs")
mod <- survextrap(Surv(years, status) ~ rx, data=colons,
                  external=extdat)
plot(mod, xlab="Years")
```



Hazard from 10-15 has been "pulled up" by inclusion of external data

Other package features

- Mixture cure models
  - Latent proportion of people become "cured" of cause of interest
- Relative survival models
  - Overall hazard = background general population hazard (inputted as constant, from matched national statistics) + disease-specific hazard

Further examples at `http://chjackson.github.io/survextrap`

Future work

- Wide experience in real situations
- Worked examples of incorporating external long-term data
  - National statistics, registry, elicited ...
- Empirical assessment of how well the default settings do
  - Simulation studies, real data examples

Other package features

- Mixture cure models
  - Latent proportion of people become "cured" of cause of interest
- Relative survival models
  - Overall hazard = background general population hazard (inputted as constant, from matched national statistics) + disease-specific hazard

Further examples at `http://chjackson.github.io/survextrap`

Future work

- Wide experience in real situations
- Worked examples of incorporating external long-term data
  - National statistics, registry, elicited ...
- Empirical assessment of how well the default settings do
  - Simulation studies, real data examples