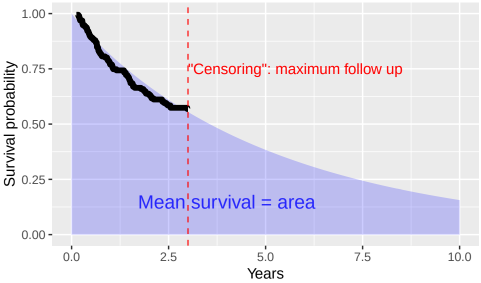
survextrap: a new tool for flexible and transparent extrapolation of survival data to inform health policy

Christopher Jackson

Royal Statistical Society, Belfast, 6 December 2023

Survival extrapolation: long-term decisions from short-term, time-to-event data



Examples:

Should health service adopt a new treatment, given 3 years follow up data from a trial?

Predicted burden on hospitals in an epidemic, given data on hospital stays, where many are people still in hospital?

Consequences of policy decisions will last longer than the end of the data

Estimating expected survival over the long term

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 this

Estimating expected survival over the long term

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 this

Estimating expected survival over the long term

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 this

Estimating expected survival over the long term

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.

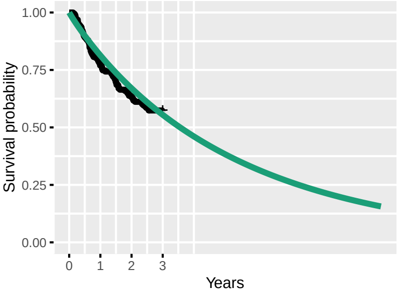
- Many choices for how to specify this

Parametric survival models: examples

Weibull distribution

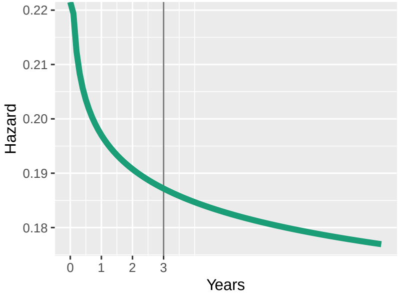
Survivor function

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



Hazard function

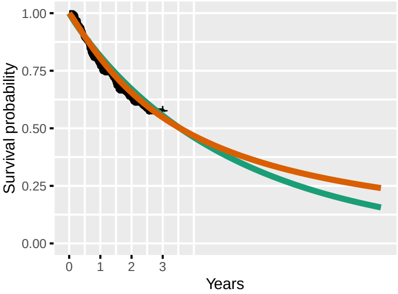
$h(t|\lambda, \alpha) = \lambda \alpha t^{\alpha-1}$



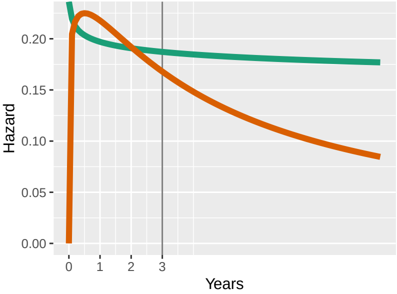
Parametric survival models: examples

Log-logistic distribution

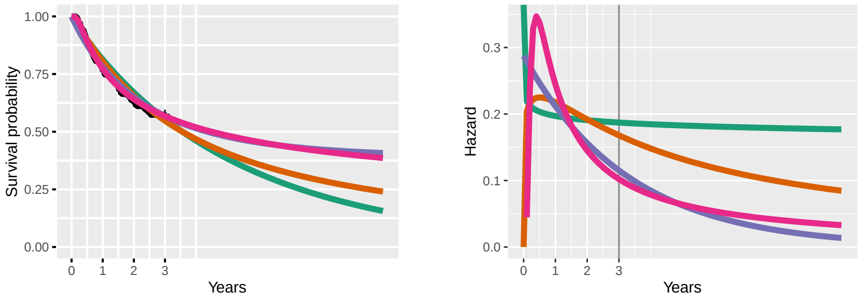
Survivor function
$S(t|\lambda,\alpha) = 1/(1+(t/b)^a)$



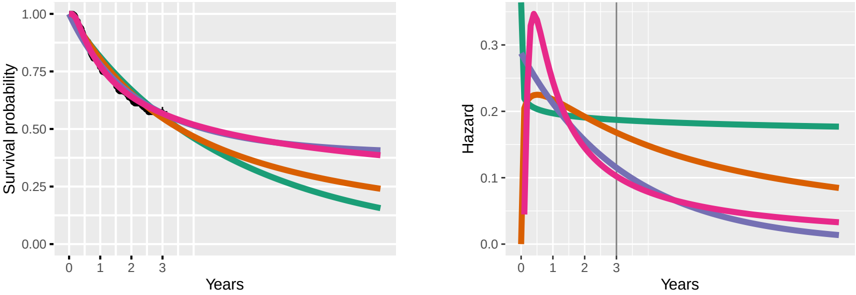
Hazard function $h(t|\lambda,\alpha) =$



Parametric survival models: examples



Parametric survival models: examples



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?

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?

Clinical judgements about the mechanism

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

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

Reviews for methods for extrapolation with external data

Bullement et al. (Medical Decision Making 2023)

A Systematic Review of Methods to Incorporate External Evidence into Trial-Based Survival Extrapolations for Health Technology Assessment

Medical Decision Making
1–11

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0272989X231168618
journals.sagepub.com/home/mdm
SAGE

Ash Bullement, Matthew D. Stevenson, Gianluca Baio, Gemma E. Shields, and Nicholas R. Latimer

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

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

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

In particular, Bayesian evidence synthesis methods

- are comprehensive, flexible, principled
- but have needed specialised programming (BUGS / JAGS), advanced statistical expertise

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!

In particular, Bayesian evidence synthesis methods

- are comprehensive, flexible, principled
- but have needed specialised programming (BUGS / JAGS), advanced statistical expertise

survextrap R package

https://chjackson.github.io/survextrap

Paper just out in BMC Med Res Meth https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-023-02094-1

Package for Bayesian survival modelling

- Multiple sources of external data to aid extrapolation
- Flexible parametric spline model for the hazard
- Multiparameter evidence synthesis, MCMC estimation (Stan)
- Principle: data and influential assumptions made as transparent as possible.
  - "you say what you know, then the computer does the hard work of converting that to answers"!

Tour of its features. . .

survextrap R package

https://chjackson.github.io/survextrap

Paper just out in BMC Med Res Meth https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-023-02094-1

Package for Bayesian survival modelling

- Multiple sources of external data to aid extrapolation
- Flexible parametric spline model for the hazard
- Multiparameter evidence synthesis, MCMC estimation (Stan)
- Principle: data and influential assumptions made as transparent as possible.
  - "you say what you know, then the computer does the hard work of converting that to answers"!

Tour of its features. . .

survextrap R package

https://chjackson.github.io/survextrap

Paper just out in BMC Med Res Meth https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-023-02094-1

Package for Bayesian survival modelling

- Multiple sources of external data to aid extrapolation
- Flexible parametric spline model for the hazard
- Multiparameter evidence synthesis, MCMC estimation (Stan)
- Principle: data and influential assumptions made as transparent as possible.
  - "you say what you know, then the computer does the hard work of converting that to answers"!

Tour of its features...

survextrap R package

https://chjackson.github.io/survextrap

Paper just out in BMC Med Res Meth https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-023-02094-1

Package for Bayesian survival modelling

- Multiple sources of external data to aid extrapolation
- Flexible parametric spline model for the hazard
- Multiparameter evidence synthesis, MCMC estimation (Stan)
- Principle: data and influential assumptions made as transparent as possible.
  - "you say what you know, then the computer does the hard work of converting that to answers"!

Tour of its features...

survextrap R package

https://chjackson.github.io/survextrap

Paper just out in BMC Med Res Meth https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-023-02094-1

Package for Bayesian survival modelling

- Multiple sources of external data to aid extrapolation
- Flexible parametric spline model for the hazard
- Multiparameter evidence synthesis, MCMC estimation (Stan)
- Principle: data and influential assumptions made as transparent as possible.
  - "you say what you know, then the computer does the hard work of converting that to answers"!

Tour of its features...

survextrap R package

https://chjackson.github.io/survextrap

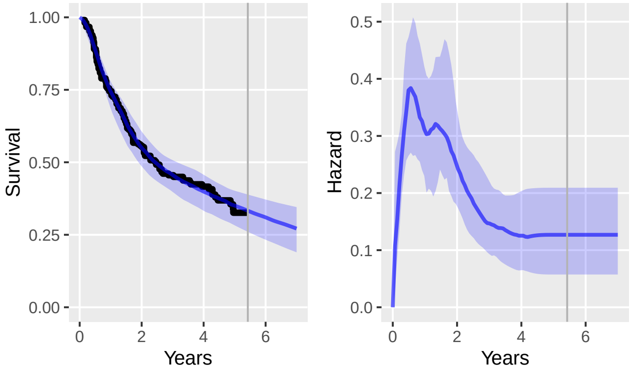
Paper just out in BMC Med Res Meth https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-023-02094-1

Package for Bayesian survival modelling

- Multiple sources of external data to aid extrapolation
- Flexible parametric spline model for the hazard
- Multiparameter evidence synthesis, MCMC estimation (Stan)
- Principle: data and influential assumptions made as transparent as possible.
  - "you say what you know, then the computer does the hard work of converting that to answers"!

Tour of its features...

Basic survextrap model for one short-term dataset



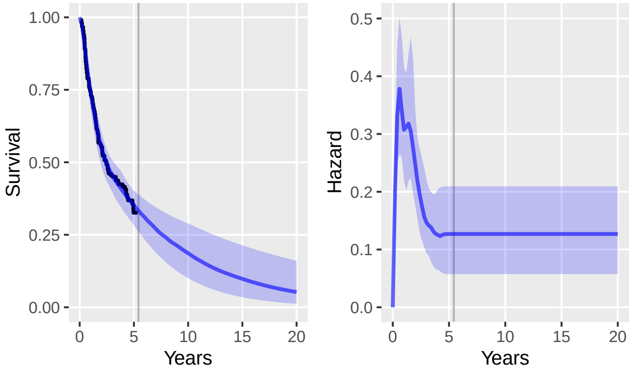
Running example: head and neck cancer trial data from Guyot et al. 2017

Penalized M-spline model

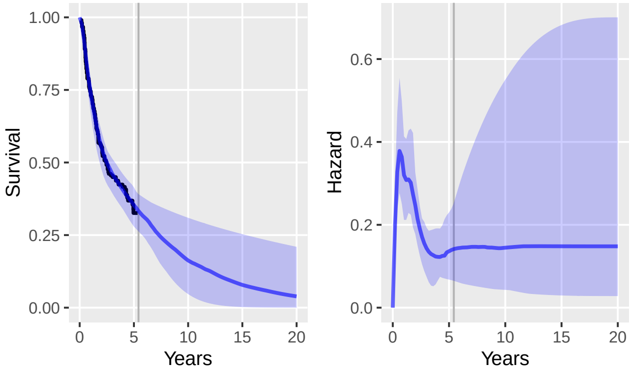
Flexible hazard within data, constant beyond data by default

```
mod1 <- survextrap(Surv(years, d) ~ 1, data=dat)
```

Basic `survextrap` model for one short-term dataset



Basic survextrap model for one short-term dataset



Allow hazard to vary up to 20 years, by adding a spline "knot".
Assumes only that the hazard is smooth

- short-term extrapolation influenced weakly by latest data.

Posterior distribution represents uncertainty due to lack of data

```
mod2 <- survextrap(Surv(years, d) ~ 1, add_knots=20)
```

M-spline survival model: how it works

Hazard $h(t)$ is weighted average of basis functions, local to time periods defined by "knots"

$$h(t) = \eta \sum_{i=1}^{n} p_i b_i(t) \qquad \eta > 0, \ \sum p_i = 1$$

Example: $n = 5$ basis functions $b_i(t)$ shown as grey lines

Parameters: weights p_i determining the shape. η is *scale*

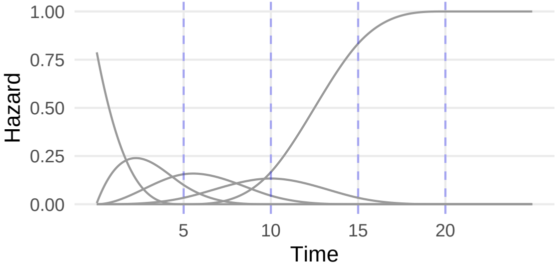
Hazard constant after the last knot

M-spline survival model: how it works

Hazard $h(t)$ is weighted average of basis functions, local to time periods defined by "knots"

$$h(t) = \eta \sum_{i=1}^{n} p_i b_i(t) \qquad \eta{>}0, \ \sum p_i{=}1$$

Example: $n = 5$ basis functions $b_i(t)$ shown as grey lines



Parameters: weights p_i determining the shape. η is *scale*

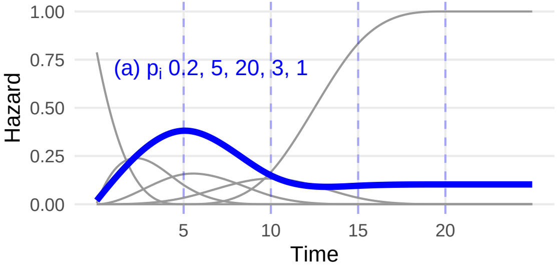
Hazard constant after the last knot

M-spline survival model: how it works

Hazard $h(t)$ is weighted average of basis functions, local to time periods defined by "knots"

$$h(t) = \eta \sum_{i=1}^{n} p_i b_i(t) \qquad \eta > 0, \; \sum p_i = 1$$

Example: $n = 5$ basis functions $b_i(t)$ shown as grey lines



Parameters: weights p_i determining the shape. η is *scale*

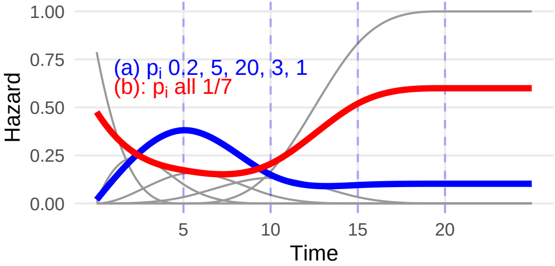
Hazard constant after the last knot

M-spline survival model: how it works

Hazard $h(t)$ is weighted average of basis functions, local to time periods defined by "knots"

$$h(t) = \eta \sum_{i=1}^{n} p_i b_i(t) \qquad \eta > 0, \ \sum p_i = 1$$

Example: $n = 5$ basis functions $b_i(t)$ shown as grey lines



Parameters: weights p_i determining the shape. η is *scale*

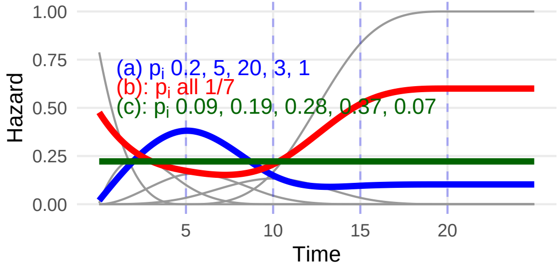
Hazard constant after the last knot

M-spline survival model: how it works

Hazard $h(t)$ is weighted average of basis functions, local to time periods defined by "knots"

$$h(t) = \eta \sum_{i=1}^{n} p_i b_i(t) \qquad \eta>0, \ \sum p_i=1$$

Example: $n = 5$ basis functions $b_i(t)$ shown as grey lines



Parameters: weights p_i determining the shape. η is *scale*

Hazard constant after the last knot

Modelling data with a penalised M-spline

- By default, knots chosen as quantiles of observed survival times, so similar amounts of information in each time interval.
- Recommendation: use lots of knots / basis coefficients
- Over-fitting controlled by a hierarchical model on basis coefficients p_i (similar to "penalised likelihood").

$$p_i \sim MNlogit\left[\text{mean: constant hazard; variance: } \sigma\right]$$

σ defines the smoothness of the hazard function
$\sigma = 0$: constant hazard. σ higher $\rightarrow$ more wiggly hazard
σ given a prior, updated to posterior given data

Principle

- learn how much smoothness is needed to represent the data
- "shrink" towards a constant hazard if data are weak

Modelling data with a penalised M-spline

- By default, knots chosen as quantiles of observed survival times, so similar amounts of information in each time interval.
- Recommendation: use lots of knots / basis coefficients
- Over-fitting controlled by a hierarchical model on basis coefficients p_i (similar to "penalised likelihood").

$$p_i \sim MNlogit\left[\text{mean: constant hazard; variance: } \sigma\right]$$

σ defines the smoothness of the hazard function
$\sigma = 0$: constant hazard. σ higher $\rightarrow$ more wiggly hazard
σ given a prior, updated to posterior given data

Principle

- learn how much smoothness is needed to represent the data
- "shrink" towards a constant hazard if data are weak

Modelling data with a penalised M-spline

- By default, knots chosen as quantiles of observed survival times, so similar amounts of information in each time interval.
- Recommendation: use lots of knots / basis coefficients
- Over-fitting controlled by a hierarchical model on basis coefficients p_i (similar to "penalised likelihood").

$$p_i \sim MNlogit\left[\text{mean: constant hazard; variance: } \sigma\right]$$

σ defines the smoothness of the hazard function
$\sigma = 0$: constant hazard. σ higher $\rightarrow$ more wiggly hazard
σ given a prior, updated to posterior given data

Principle

- learn how much smoothness is needed to represent the data
- "shrink" towards a constant hazard if data are weak

Modelling data with a penalised M-spline

- By default, knots chosen as quantiles of observed survival times, so similar amounts of information in each time interval.
- Recommendation: use lots of knots / basis coefficients
- Over-fitting controlled by a hierarchical model on basis coefficients p_i (similar to "penalised likelihood").

$$p_i \sim MNlogit\left[\text{mean: constant hazard; variance: } \sigma\right]$$

σ defines the smoothness of the hazard function
$\sigma = 0$: constant hazard. σ higher $\rightarrow$ more wiggly hazard
σ given a prior, updated to posterior given data

Principle

- learn how much smoothness is needed to represent the data
- "shrink" towards a constant hazard if data are weak

Use of external data (needed for the long term)

Supplied as data frame of aggregate counts of survivors over a series of intervals, with covariates. Example:

Follow-up period		Number		Covariates	
From t	To u	Alive at t	Still alive at u	Treatment	...
5	6	358	325	Control	
6	7	308	285	Control	
8	9	221	198	Control	
etc.					

Use of external data (needed for the long term)

Supplied as data frame of aggregate counts of survivors over a series of intervals, with covariates. Example:

Follow-up period		Number		Covariates	
				Treatment	...
From t	To u	Alive at t	Still alive at u		
5	6	358	325	Control	
6	7	308	285	Control	
8	9	221	198	Control	
etc.					

- Registry data, population data, and elicited judgements can often be expressed in this form
- Each count of survivors is a binomal outcome, with probability defined by the spline model and covariates
- Bayesian evidence synthesis: posterior for parameters determined given individual and external data together

Use of external data (needed for the long term)

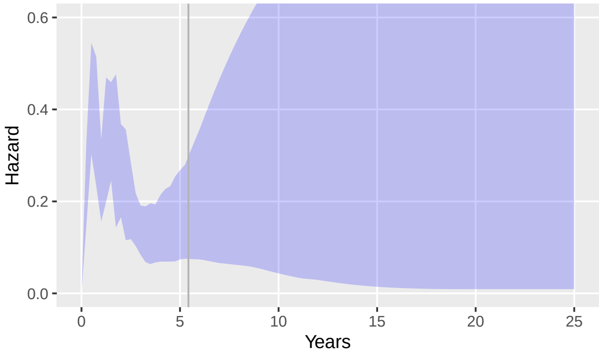
Supplied as data frame of aggregate counts of survivors over a series of intervals, with covariates. Example:

Follow-up period		Number		Covariates	
From t	To u	Alive at t	Still alive at u	Treatment	...
5	6	358	325	Control	
6	7	308	285	Control	
8	9	221	198	Control	
etc.					

Biased external data? — population with different risks from trial?

- Can model via a "dataset" covariate with proportional hazards (and/or other covariates) — assuming hazard shape common
  - can estimate hazard ratio if datasets observed at same times
  - but take care with extrapolating outside those times

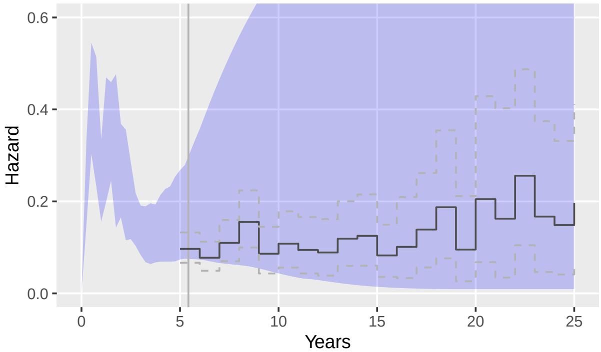
Survival modelling with external data: example



With no external data, and just a smoothness assumption, estimated long-term hazard is extremely uncertain

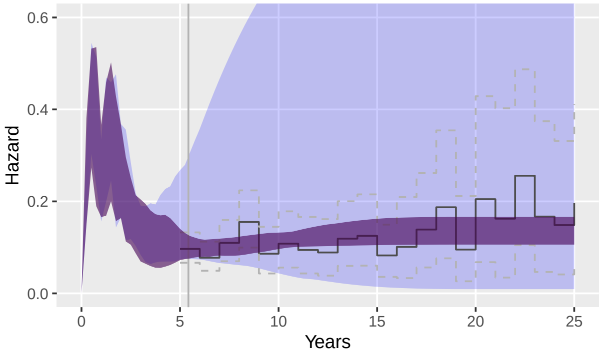
```
mod_seer <- survextrap(Surv(years, d) ~ 1, data=dat,
                       add_knots=20)
```

Survival modelling with external data: example



Registry data give annual survival rates from 5 to 25 years.

Survival modelling with external data: example

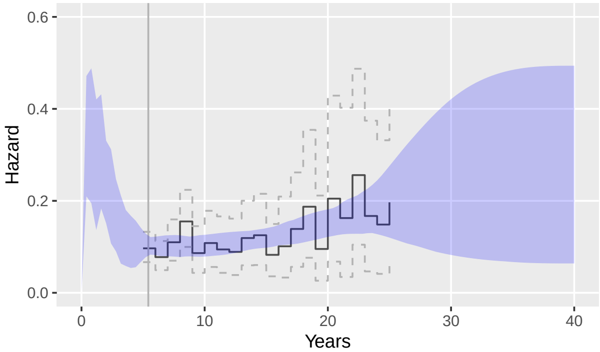


Registry data give annual survival rates from 5 to 25 years.

- Enables a confident "extrapolation" of the hazard for control group survival

```
mod_seer <- survextrap(Surv(years, d) ~ 1, data=dat,
                       external=ext_dat, add_knots=20)
```

Survival modelling with external data: example



Extrapolating up to 40 years, need longer-term data...

```
mod_seer <- survextrap(Surv(years, d) ~ 1, data=dat,
                       external=ext_dat,
                       add_knots=c(20,30,40))
```

Additive hazards model

Can supply a known background hazard $h_b(t)$

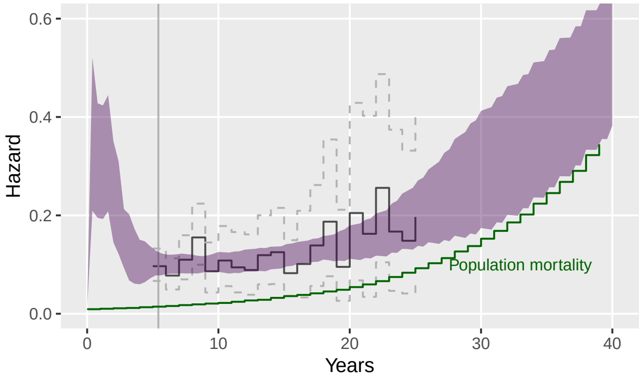
- e.g. national mortality statistics for people of comparable age/sex

Hazard for trial patients modelled as additive:

$$h(t) = h_b(t) + h_e(t)$$

i.e. no lower than population hazard
$h_b()$ assumed known, M-spline model placed on excess hazard $h_e()$

Additive hazards model




```
mod_seer <- survextrap(Surv(years, d) ~ 1, data=dat,
                       external=ext_dat, bh=back_dat,
                       add_knots=c(20,30,40))
```

Including expert judgements on long-term survival

Suppose we judge by 40 years, survival will be similar to general population (though no evidence of this here!)

- i.e. (conditional annual) survival probability 0.72

We might elicit a distribution (Beta, say) with

- mean 0.72
- 95% credible interval of 0.69–0.75 (say)

Equivalent information to having observed 724 survivors out of 1000 people

- posterior we would get with these data and an uniform prior. Bigger denominator → narrower credible interval

Expressing elicited info as a count allows inclusion in model as "external" data

Including expert judgements on long-term survival

Suppose we judge by 40 years, survival will be similar to general population (though no evidence of this here!)

- i.e. (conditional annual) survival probability 0.72

We might elicit a distribution (Beta, say) with

- mean 0.72
- 95% credible interval of 0.69–0.75 (say)

Equivalent information to having observed 724 survivors out of 1000 people

- posterior we would get with these data and an uniform prior. Bigger denominator → narrower credible interval

Expressing elicited info as a count allows inclusion in model as "external" data

Including expert judgements on long-term survival

Suppose we judge by 40 years, survival will be similar to general population (though no evidence of this here!)

- i.e. (conditional annual) survival probability 0.72

We might elicit a distribution (Beta, say) with

- mean 0.72
- 95% credible interval of 0.69–0.75 (say)

Equivalent information to having observed 724 survivors out of 1000 people

- posterior we would get with these data and an uniform prior.
  Bigger denominator $\rightarrow$ narrower credible interval

Expressing elicited info as a count allows inclusion in model as "external" data

Including expert judgements on long-term survival

Suppose we judge by 40 years, survival will be similar to general population (though no evidence of this here!)

- i.e. (conditional annual) survival probability 0.72

We might elicit a distribution (Beta, say) with

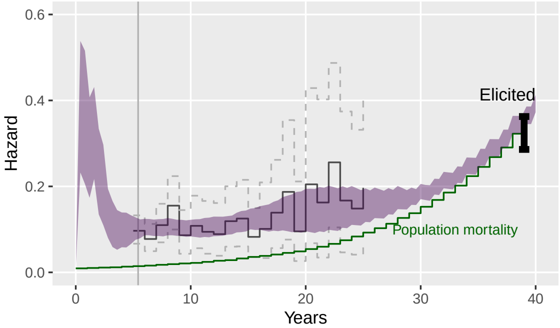
- mean 0.72
- 95% credible interval of 0.69–0.75 (say)

Equivalent information to having observed 724 survivors out of 1000 people

- posterior we would get with these data and an uniform prior. Bigger denominator $\rightarrow$ narrower credible interval

Expressing elicited info as a count allows inclusion in model as "external" data

Model including long term expert judgement



Full evidence synthesis: trial, registry, population and elicited data all combined

```
mod_seer <- survextrap(Surv(years, d) ~ 1, data=dat,
                       external = ext_elic_dat,
                       add_knots=c(20,30,40))
```

Mixture cure models

Common to extrapolate survival by assuming disease-specific hazard diminishes to zero.

Mixture cure model: $$S(t) = p + (1-p)S_0(t)$$

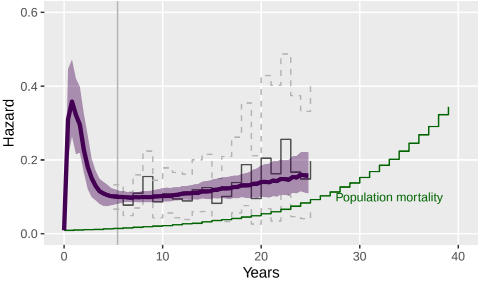
Long-term survival converges to p (probability of "cure"), excess hazard converges to 0, S_0 is survival for "uncured" person.

Mixture cure models

Common to extrapolate survival by assuming disease-specific hazard diminishes to zero.

$$\text{Mixture cure model:} \qquad S(t) = p + (1-p)S_0(t)$$

Long-term survival converges to p (probability of "cure"), excess hazard converges to 0, S_0 is survival for "uncured" person.



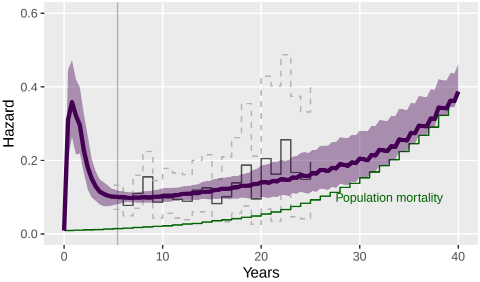
Little evidence of "cure" in data.

Mixture cure models

Common to extrapolate survival by assuming disease-specific hazard diminishes to zero.

$$\text{Mixture cure model:} \qquad S(t) = p + (1-p)S_0(t)$$

Long-term survival converges to p (probability of "cure"), excess hazard converges to 0, S_0 is survival for "uncured" person.



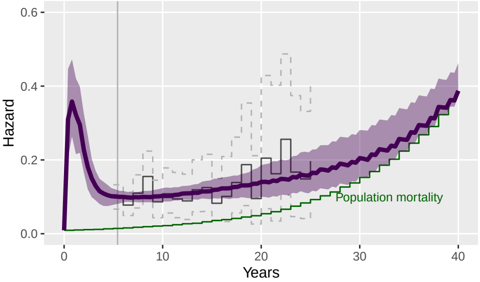
Little evidence of "cure" in data.
Extrapolation relies on parametric shape of cure model.

Mixture cure models

Common to extrapolate survival by assuming disease-specific hazard diminishes to zero.

Mixture cure model: $$S(t) = p + (1-p)S_0(t)$$

Long-term survival converges to p (probability of "cure"), excess hazard converges to 0, S_0 is survival for "uncured" person.



Little evidence of "cure" in data.
Extrapolation relies on parametric shape of cure model.

More transparent to express "cure" assumption using elicited data

Treatment or covariate effects

Proportional hazards model

```
mod_seer <- survextrap(Surv(years, d) ~ treat + age,
                       data=dat, external=ext_dat)
```

Flexible non-proportional hazards model (hazard shape parameters p_i depend on covariates)

```
mod_seer <- survextrap(Surv(years, d) ~ treat + age,
                       data=dat, external=ext_dat,
                       nonprop=TRUE)
```

Additionally can impose a waning effect when predicting

- log hazard ratio for a treatment decreases from modelled value to zero over some time period
- used for sensitivity analysis to what happens after a trial

Treatment or covariate effects

Proportional hazards model

```
mod_seer <- survextrap(Surv(years, d) ~ treat + age,
                       data=dat, external=ext_dat)
```

Flexible non-proportional hazards model (hazard shape parameters p_i depend on covariates)

```
mod_seer <- survextrap(Surv(years, d) ~ treat + age,
                       data=dat, external=ext_dat,
                       nonprop=TRUE)
```

Additionally can impose a waning effect when predicting

- log hazard ratio for a treatment decreases from modelled value to zero over some time period
- used for sensitivity analysis to what happens after a trial

Treatment or covariate effects

Proportional hazards model

```
mod_seer <- survextrap(Surv(years, d) ~ treat + age,
                       data=dat, external=ext_dat)
```

Flexible non-proportional hazards model (hazard shape parameters p_i depend on covariates)

```
mod_seer <- survextrap(Surv(years, d) ~ treat + age,
                       data=dat, external=ext_dat,
                       nonprop=TRUE)
```

Additionally can impose a waning effect when predicting

- log hazard ratio for a treatment decreases from modelled value to zero over some time period
- used for sensitivity analysis to what happens after a trial

Prior distributions: general strategy

Use simulation to determine consequences of a prior on an interpretable scale

Example: smoothness of hazard shape driven by a parameter σ without a direct interpretation

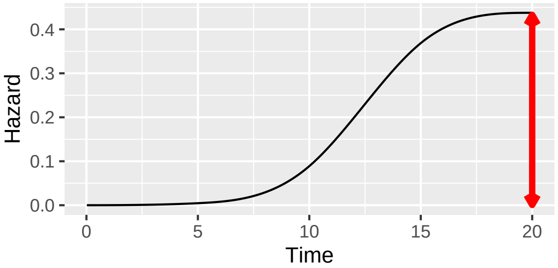
- Simulate hazard curves from prior predictive distribution
- Find e.g. 10% and 90% quantile of $h(t)$ over time t
- Leads to a distribution for the ratio between high and low values
- Find prior that gives belief such as "highest hazards (over time) unlikely (e.g. 5% prob) to be $10\times$ the lowest"

Prior distributions: general strategy

Use simulation to determine consequences of a prior on an interpretable scale

Example: smoothness of hazard shape driven by a parameter σ without a direct interpretation

- Simulate hazard curves from prior predictive distribution
- Find e.g. 10% and 90% quantile of $h(t)$ over time t
- Leads to a distribution for the ratio between high and low values
- Find prior that gives belief such as "highest hazards (over time) unlikely (e.g. 5% prob) to be 10 × the lowest"

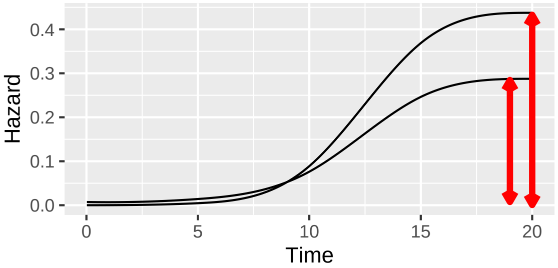


Prior distributions: general strategy

Use simulation to determine consequences of a prior on an interpretable scale

Example: smoothness of hazard shape driven by a parameter σ without a direct interpretation

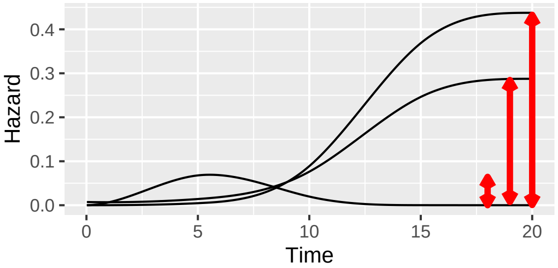
- Simulate hazard curves from prior predictive distribution
- Find e.g. 10% and 90% quantile of $h(t)$ over time t
- Leads to a distribution for the ratio between high and low values
- Find prior that gives belief such as "highest hazards (over time) unlikely (e.g. 5% prob) to be 10 $\times$ the lowest"



Use simulation to determine consequences of a prior on an interpretable scale

Example: smoothness of hazard shape driven by a parameter σ without a direct interpretation

- Simulate hazard curves from prior predictive distribution
- Find e.g. 10% and 90% quantile of $h(t)$ over time t
- Leads to a distribution for the ratio between high and low values
- Find prior that gives belief such as "highest hazards (over time) unlikely (e.g. 5% prob) to be 10 $\times$ the lowest"



Final discussion

Further info in the paper, full doc and worked examples on
https://chjackson.github.io/survextrap

Many details I haven't mentioned today, e.g:

- "Post-estimation" outputs e.g. restricted mean survival
- Model comparison by efficient cross-validation

Software status: "beta": not expected to change drastically before "full release" (CRAN), but more testing planned

Ongoing further work

- Simulation studies to assess if defaults are sensible
- Listen to users. Usable? Understandable? What is most challenging?

Final discussion

Further info in the paper, full doc and worked examples on
`https://chjackson.github.io/survextrap`

Many details I haven't mentioned today, e.g:

- "Post-estimation" outputs e.g. restricted mean survival
- Model comparison by efficient cross-validation

Software status: "beta": not expected to change drastically before "full release" (CRAN), but more testing planned

Ongoing further work

- Simulation studies to assess if defaults are sensible
- Listen to users. Usable? Understandable? What is most challenging?

Final discussion

Further info in the paper, full doc and worked examples on
https://chjackson.github.io/survextrap

Many details I haven't mentioned today, e.g:

- "Post-estimation" outputs e.g. restricted mean survival
- Model comparison by efficient cross-validation

Software status: "beta": not expected to change drastically before "full release" (CRAN), but more testing planned

Ongoing further work

- Simulation studies to assess if defaults are sensible
- Listen to users. Usable? Understandable? What is most challenging?

MRC Biostatistics Unit

UNIVERSITY OF CAMBRIDGE

Thank you