Bayesian methods in health technology evaluation: survival extrapolation and structural uncertainty

Chris Jackson
MRC Biostatistics Unit, Cambridge, U.K.

Bayes Pharma, 12 June 2014

Overview

1. Extrapolation of RCT evidence over time (slides 4–23)
   - Estimating expected survival from short-term RCT + long-term population data
   - Importance of modelling different causes of death

   With Tatiana Benaglia (Sao Paulo), Linda Sharples (Leeds)

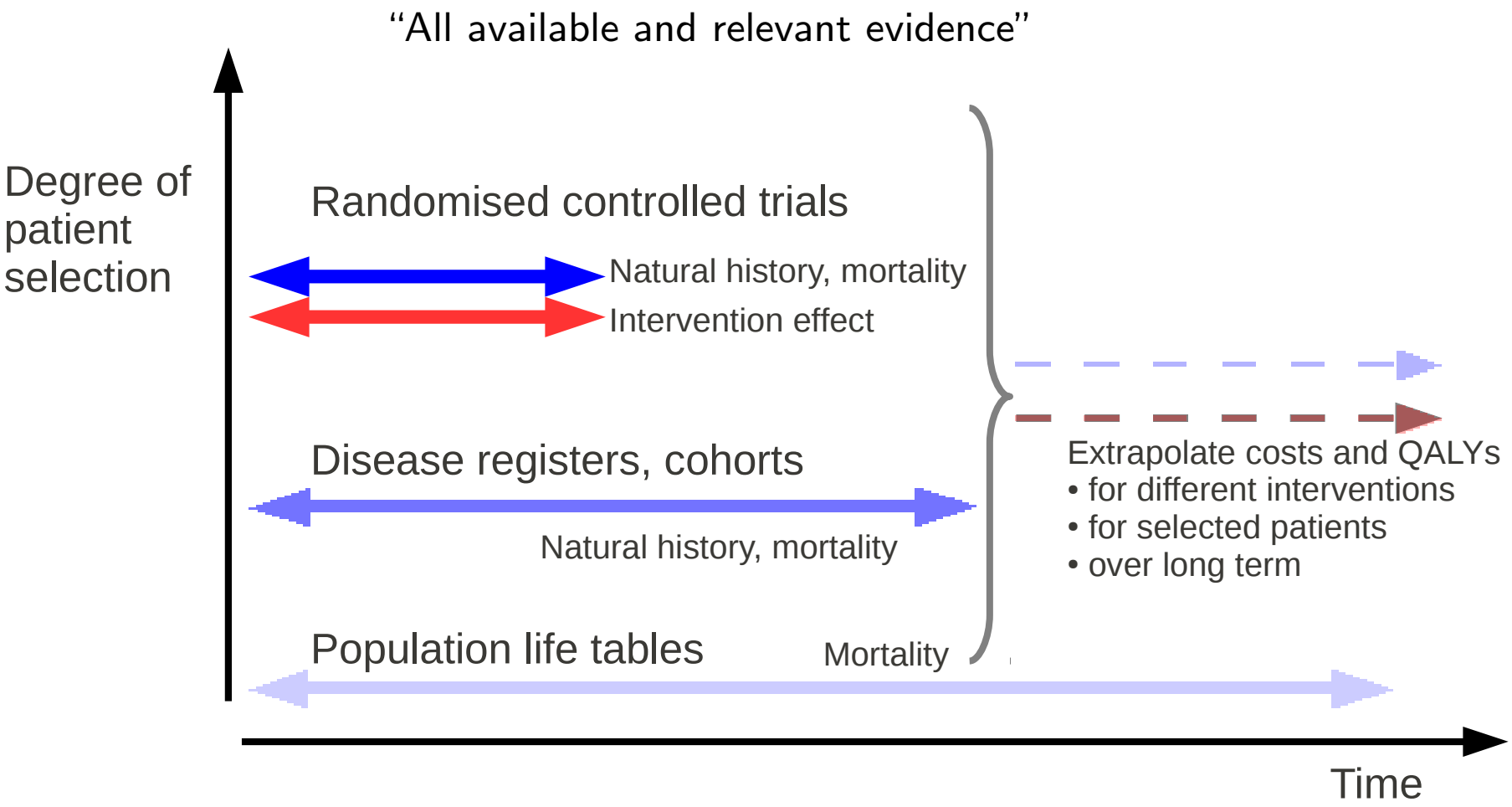
2. Structural uncertainty in decision models (slides 24–33)
   - A brief and broad review of methods

   With Linda Sharples (Leeds), Howard Thom (Bristol), Simon Thompson (Cambridge)

Part I

Extrapolation over time from RCT data for health economic evaluations

Data informing a typical health economic evaluation



- Combine (relevant) short-term + (less-relevant) long-term data.
- Extrapolate over time and to different populations.

Motivating example: ICDs

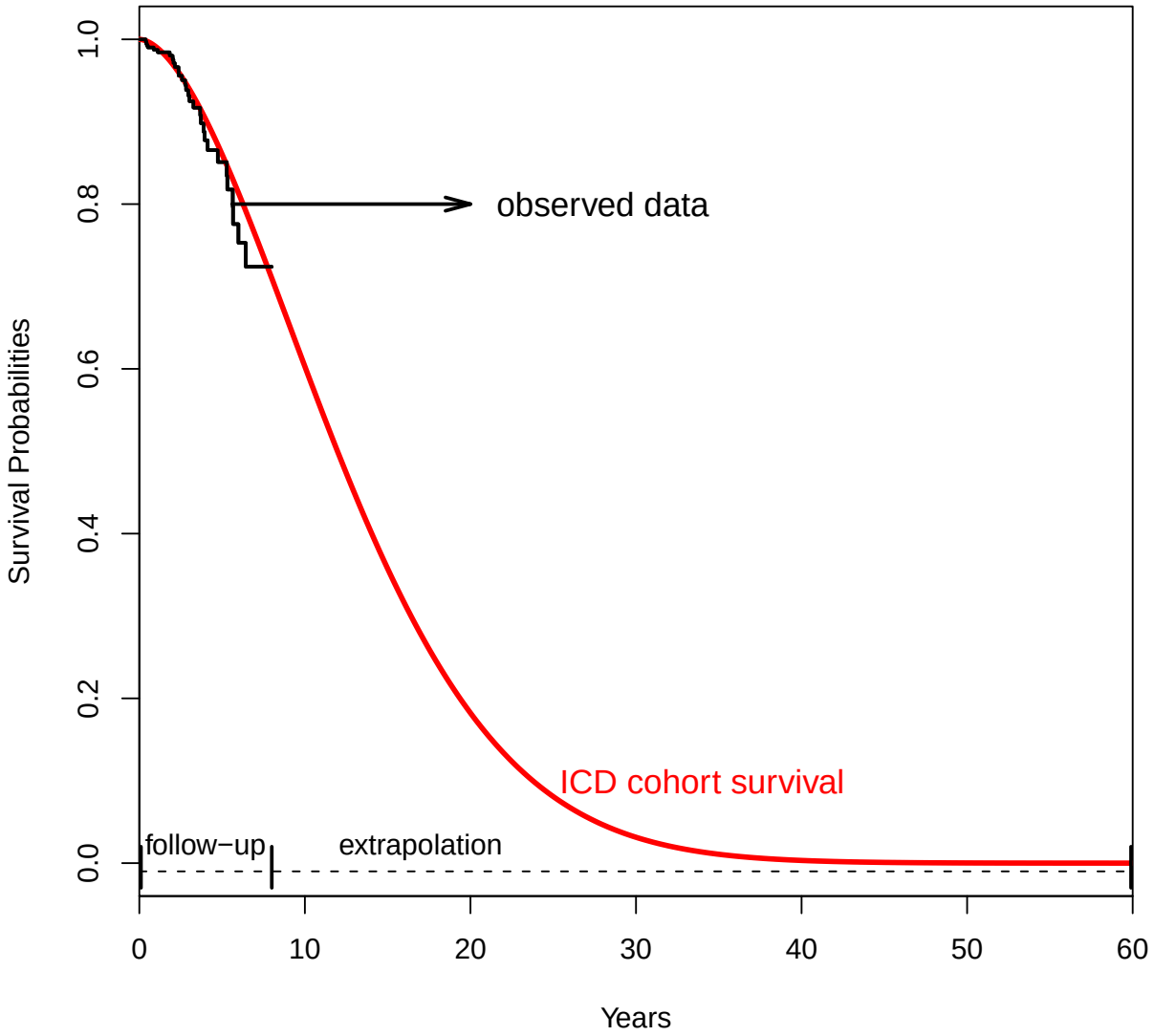
ICD (Implantable Cardioverter Defibrillators) compared to anti-arrhythmic drugs (AAD) for prevention of sudden cardiac death in patients with cardiac arrhythmia.

Data:

- Individual data from cohort of 535 UK cardiac arrhythmia patients implanted with ICDs between 1991 and 2002.
- Meta-analysis of three (non-UK) RCTs (published HRs).
  - Relatively short-term follow-up: approximately 75% of patients followed for less than 5 years, maximum 10 years
- UK population mortality statistics by age, sex, cause of death.

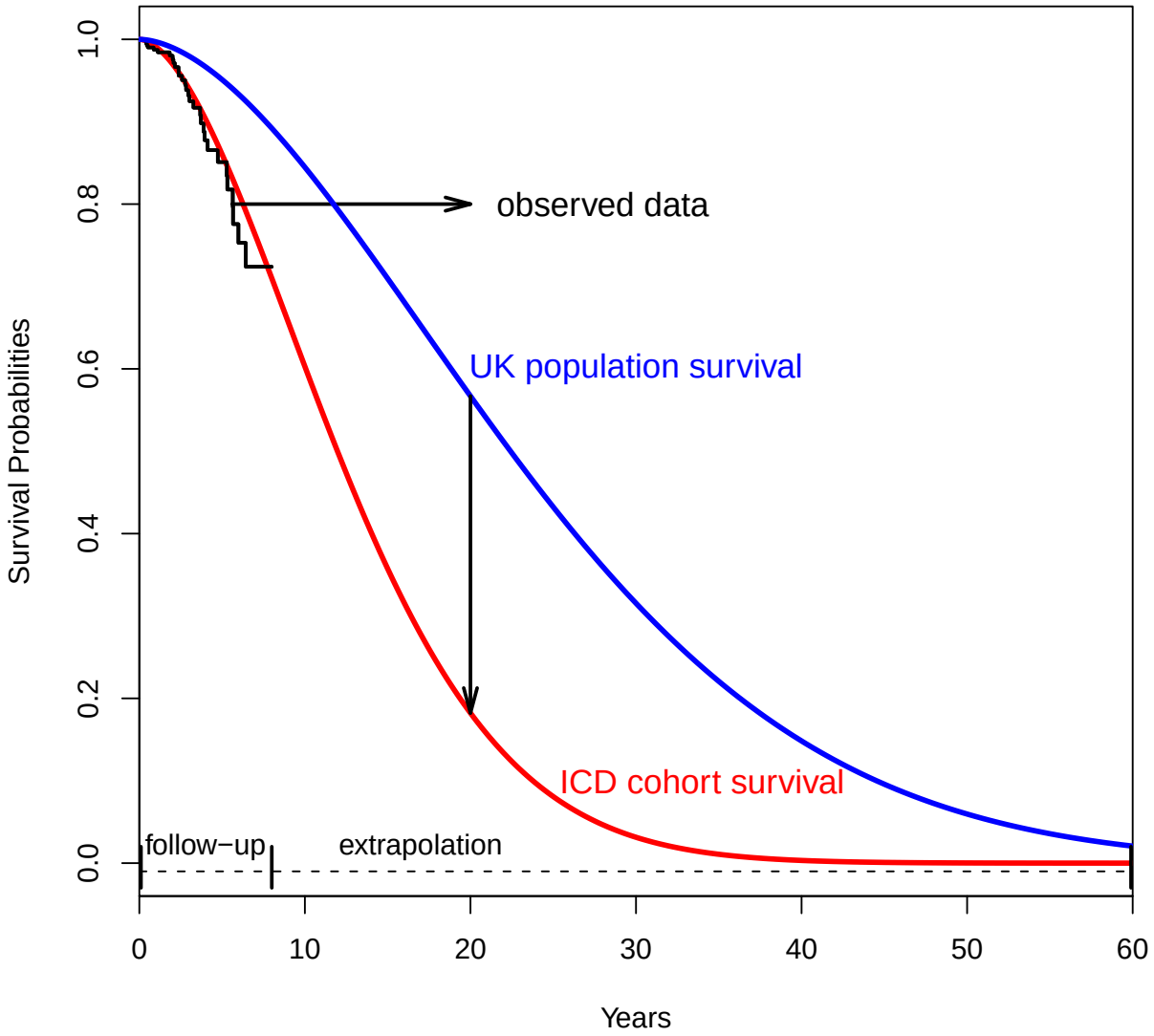
Estimate the survival curve over the lifetime of ICD and AAD patients in UK

Previous Work: central idea (Demiris & Sharples, Stat. Med. 2006)



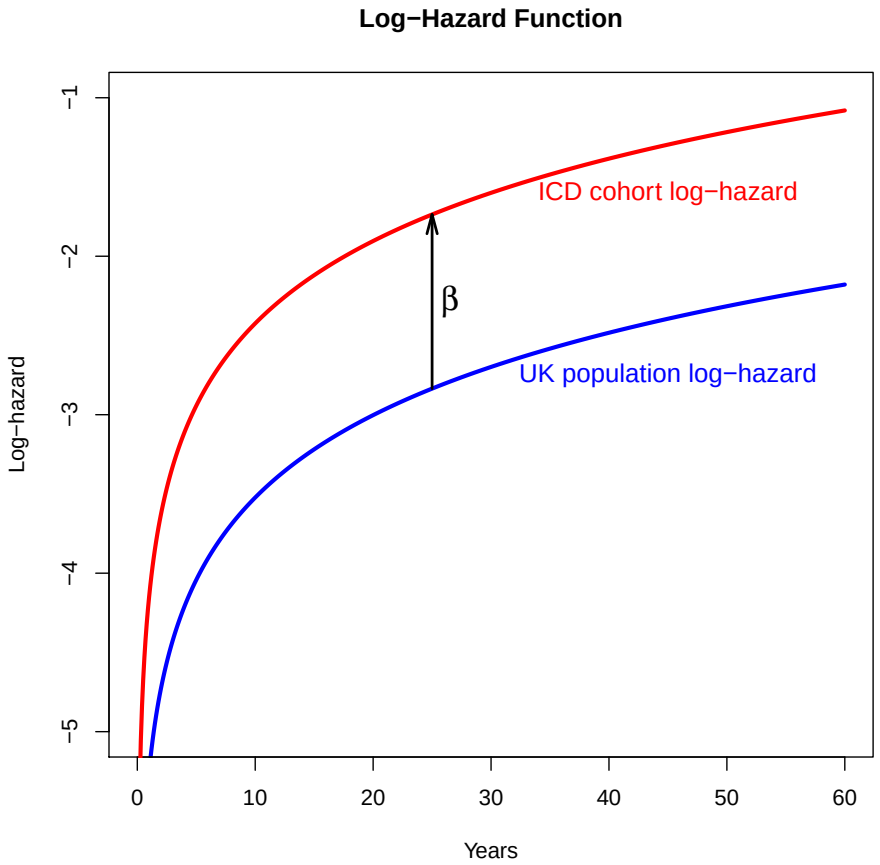
Use UK population data with same age/sex distribution to anchor the ICD population risk

Previous Work: central idea (Demiris & Sharples, Stat. Med. 2006)



Use UK population data with same age/sex distribution to anchor the ICD population risk

Previous Work: key assumption



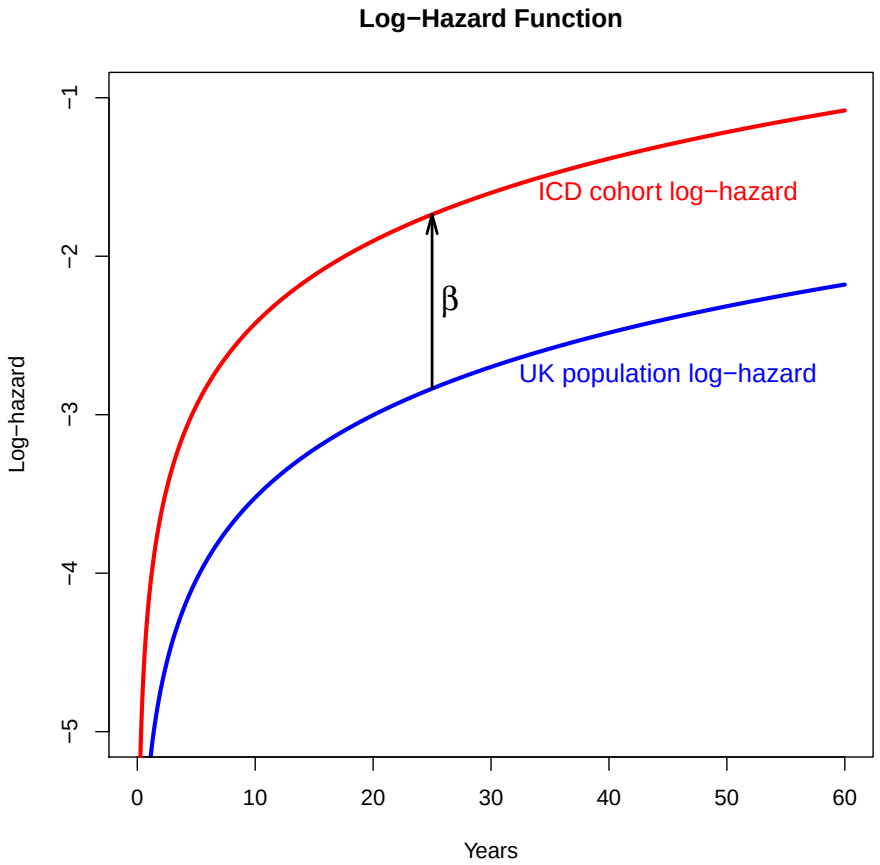
$$h_{ICD}(t) = e^{\beta} h_{UK}(t), \text{ for } t > 0$$

Constant (multiplicative) hazard ratio between ICD and UK population

This seems a strong assumption:

1. ICD patients at greater risk of arrhythmia death
2. If contribution of arrhythmia deaths changes over time, then extrapolating constant HR for all causes of death may be inaccurate

Previous Work: key assumption



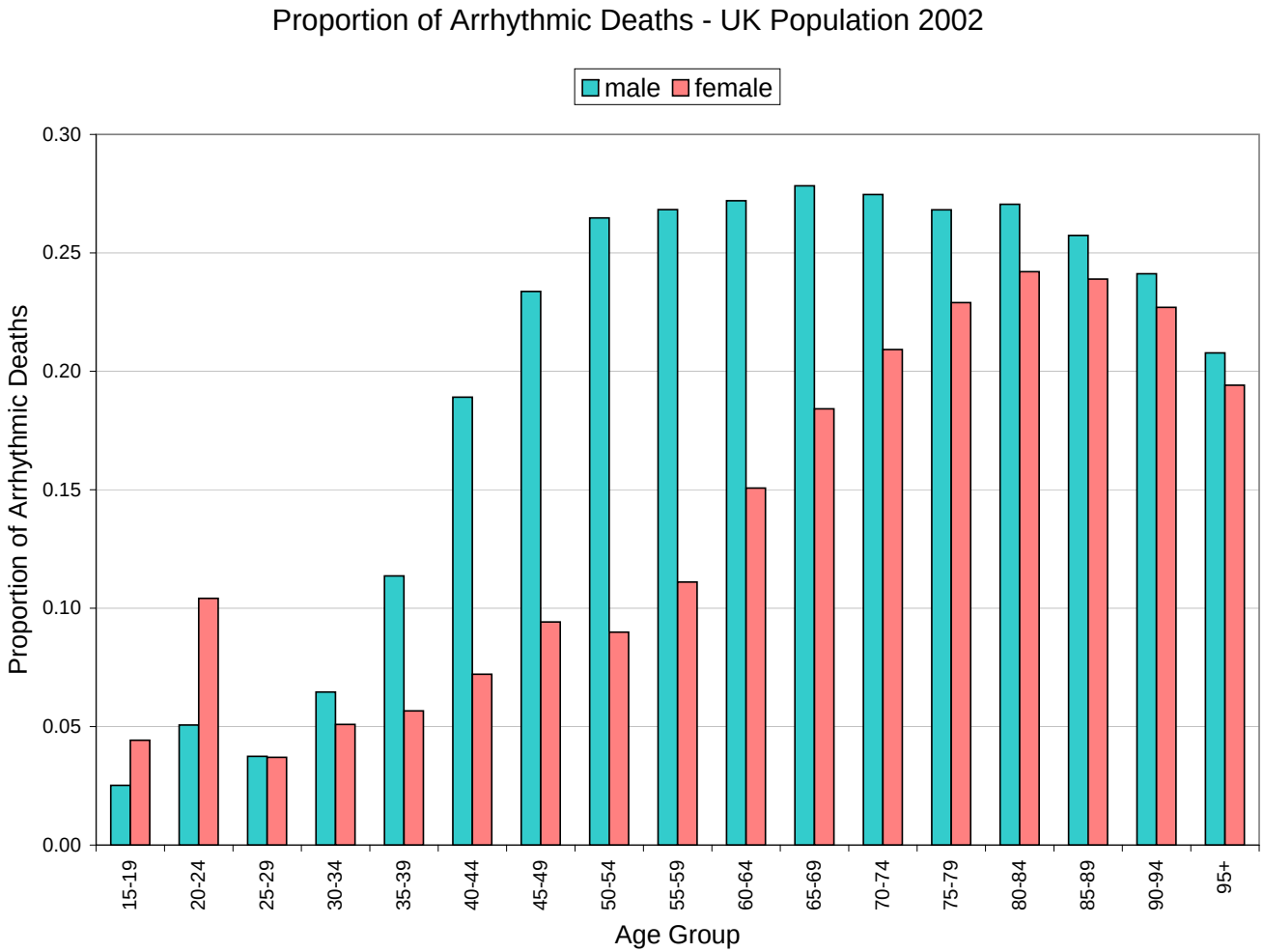
$$h_{ICD}(t) = e^{\beta} h_{UK}(t), \text{ for } t > 0$$

Constant (multiplicative) hazard ratio between ICD and UK population

This seems a strong assumption:

1. ICD patients at greater risk of arrhythmia death
2. If contribution of arrhythmia deaths changes over time, then extrapolating constant HR for all causes of death may be inaccurate

Proportion of UK deaths which are due to arrhythmia



Examining implicit assumptions

Extrapolating e^{β} implicitly assumes arrhythmia hazard is a constant proportion of all-cause hazard.

May be more plausible to extrapolate constant cause-specific hazard ratio.

How much difference is this assumption likely to make?

What parameters affect bias of estimates of mean survival?

1. Simulation study to estimate bias and coverage under different assumptions about model parameters
2. Application to ICD example

Model to extrapolate survival for ICD patients

(not considering AAD control group, RCT data for the moment...)

- General population data: cause of death (k =arrhythmic, non-arrhythmic) known.
  Cause-specific survival is Weibull with hazard:

$$h_{UK}^{(k)}(t) = \alpha_k \lambda_k t^{\alpha_k - 1}$$

- ICD cohort: cause of death unknown
  Overall survival follows a polyhazard model (Louzada-Neto, Biometrics 1999):

$$h_{ICD}(t) = h_{ICD}^{arr}(t) + h_{ICD}^{other}(t)$$

  - t: minimum time to one of 2 possible causes of death
  - Hazard is the sum of 2 cause-specific hazards

Cause-specific proportional hazards assumption

ICD cohort hazard is related to the general population hazard as:

$$\begin{aligned} h_{ICD}(t) &= h^{arr}_{ICD}(t) + h^{other}_{ICD}(t) \\ &= e^{\beta} h^{arr}_{UK}(t) + h^{other}_{UK}(t) \\ &= e^{\beta} \alpha_1 \lambda_1 t^{\alpha_1 - 1} + \alpha_2 \lambda_2 t^{\alpha_2 - 1} \text{ (poly-Weibull)} \end{aligned}$$

Arrhythmia hazard is proportional
Other-cause hazard is identical
} to UK matched population.

- Joint Bayesian model for ICD cohort + UK population data
- Estimate joint posterior of parameters $\alpha_1, \alpha_2, \lambda_1, \lambda_2, \beta$ by MCMC (using WinBUGS).
- WBDev add-on needed to implement the poly-Weibull distribution for the cohort data

Weakly informative prior distributions

Express beliefs on an intuitive scale — exact choice may make a difference for small populations

Weibull rate λ:

- Age around 60 on study entry: patients cannot survive more than 60 additional years. Mean survival $\sim U(0, 60)$.
- $1/\lambda \sim U(0, 100)$, gives a mean $1/\lambda\Gamma(1 + 1/\alpha)$ of < 60, for all plausible α.

Weibull shape α: controls hazard vs. time: $h(t) = \alpha\lambda(\lambda t)^{\alpha-1}$

- Hazard ratio for doubled time t is $2^{\alpha-1}$.
- Prior mean of 1.5 for this, with 95% CI about (0.64, 100)
- implies $\log(\alpha) \sim \mathrm{N}(0.5, \sigma = 0.78)$

Log HR β between ICD patients and general population: 95% CI for HR (1/150,150) $\rightarrow \beta \sim N(0, \sigma = 2.5)$

Weakly informative prior distributions

Express beliefs on an intuitive scale — exact choice may make a difference for small populations

Weibull rate λ:

- Age around 60 on study entry: patients cannot survive more than 60 additional years. Mean survival $\sim U(0, 60)$.
- $1/\lambda \sim U(0, 100)$, gives a mean $1/\lambda\Gamma(1 + 1/\alpha)$ of < 60, for all plausible α.

Weibull shape α: controls hazard vs. time: $h(t) = \alpha\lambda(\lambda t)^{\alpha-1}$

- Hazard ratio for doubled time t is $2^{\alpha-1}$.
- Prior mean of 1.5 for this, with 95% CI about (0.64, 100)
- implies $\log(\alpha) \sim \mathrm{N}(0.5, \sigma = 0.78)$

Log HR β between ICD patients and general population: 95% CI for HR (1/150,150) $\rightarrow \beta \sim N(0, \sigma = 2.5)$

Weakly informative prior distributions

Express beliefs on an intuitive scale — exact choice may make a difference for small populations

Weibull rate λ:

- Age around 60 on study entry: patients cannot survive more than 60 additional years. Mean survival $\sim U(0, 60)$.
- $1/\lambda \sim U(0, 100)$, gives a mean $1/\lambda\Gamma(1 + 1/\alpha)$ of < 60, for all plausible α.

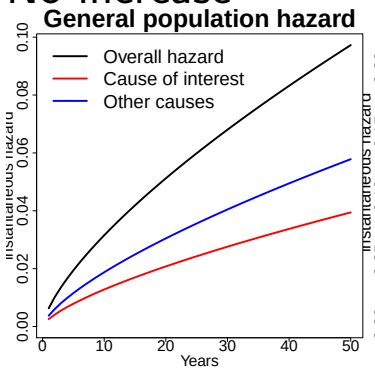
Weibull shape α: controls hazard vs. time: $h(t) = \alpha\lambda(\lambda t)^{\alpha-1}$

- Hazard ratio for doubled time t is $2^{\alpha-1}$.
- Prior mean of 1.5 for this, with 95% CI about (0.64, 100)
- implies $\log(\alpha) \sim \mathrm{N}(0.5, \sigma = 0.78)$

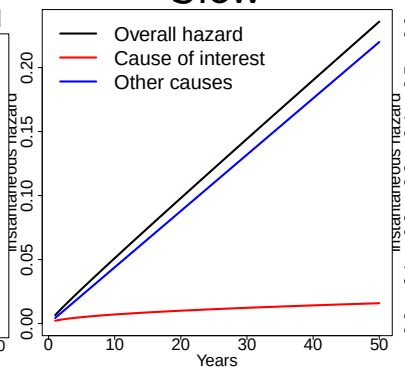
Log HR β between ICD patients and general population: 95% CI for HR (1/150,150) $\rightarrow \beta \sim N(0, \sigma = 2.5)$

Simulation study — generate data of same design

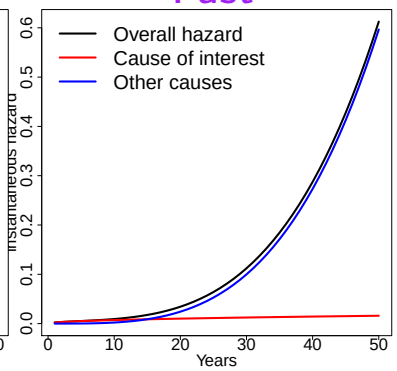
No increase



Slow



Fast



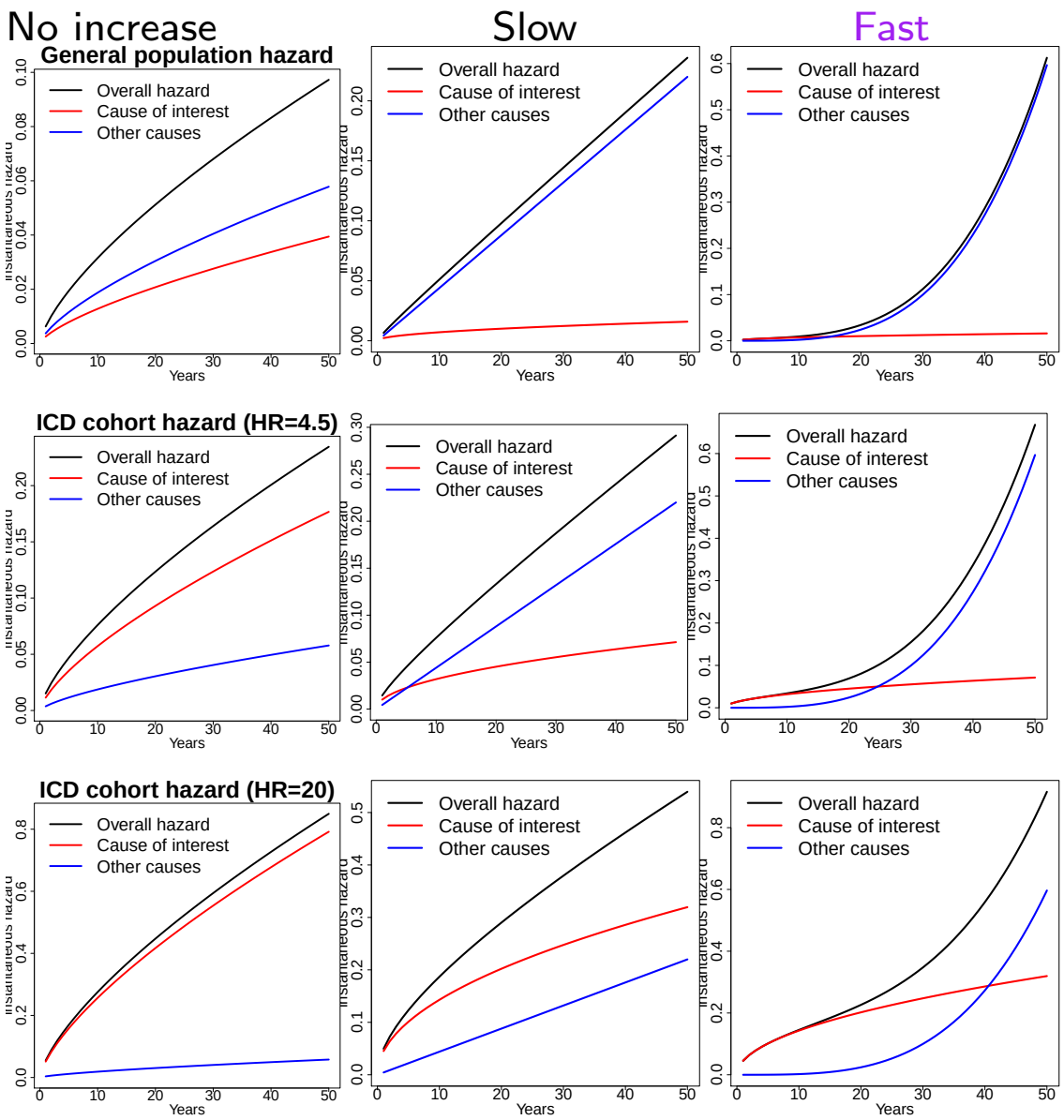
Parameters varied:

- Increase in other-cause hazard relative to cause-specific (other causes may dominate as people age)
  - Hazard ratio between the ICD cohort and population

Alternative models fitted

- Correct (poly-Weibull) model
- Weibull model which ignores cause of death

Simulation study — generate data of same design



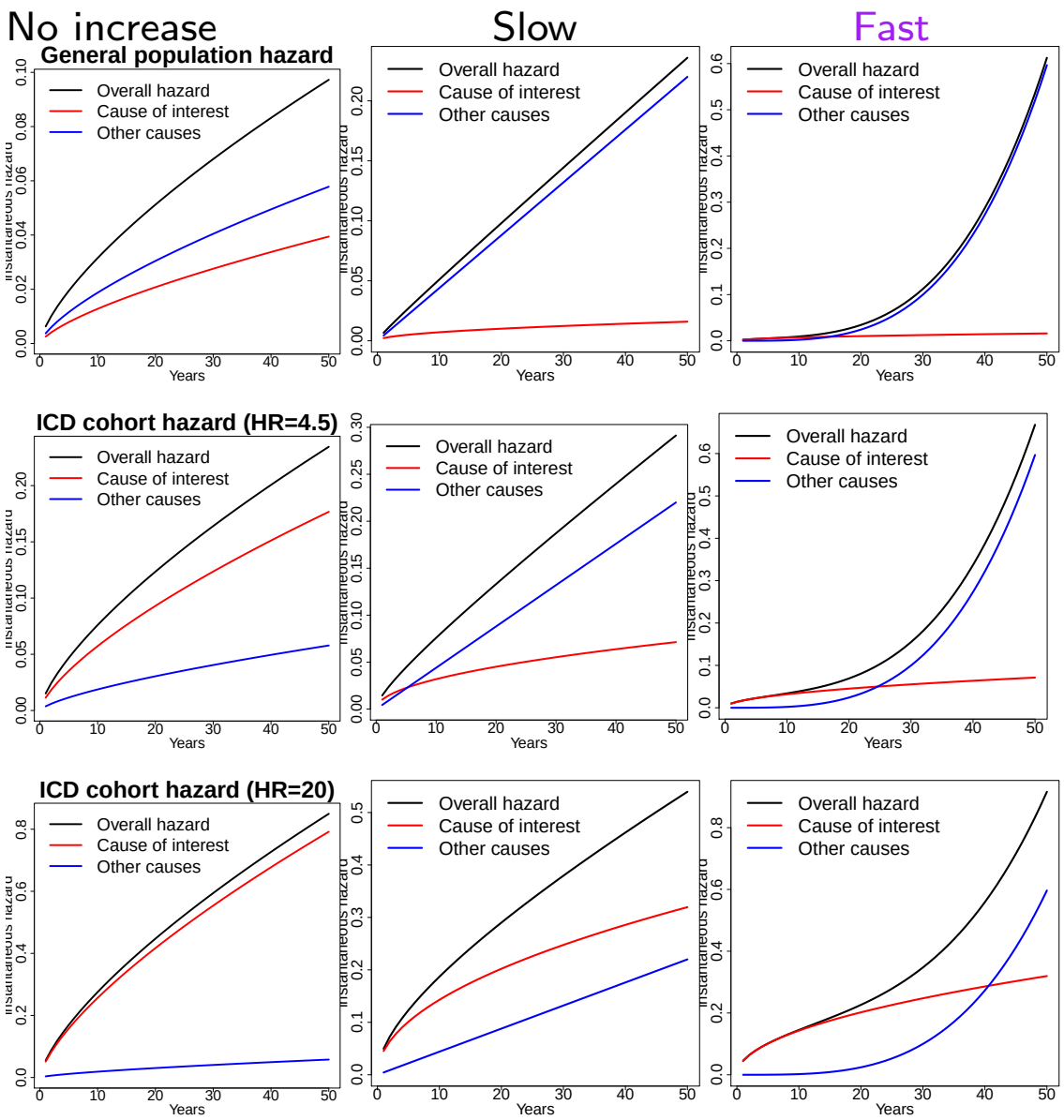
Parameters varied:

- Increase in other-cause hazard relative to cause-specific (other causes may dominate as people age)
- Hazard ratio between the ICD cohort and population

Alternative models fitted

- Correct (poly-Weibull) model
- Weibull model which ignores cause of death

Simulation study — generate data of same design



Parameters varied:

- Increase in other-cause hazard relative to cause-specific (other causes may dominate as people age)
- Hazard ratio between the ICD cohort and population

Alternative models fitted

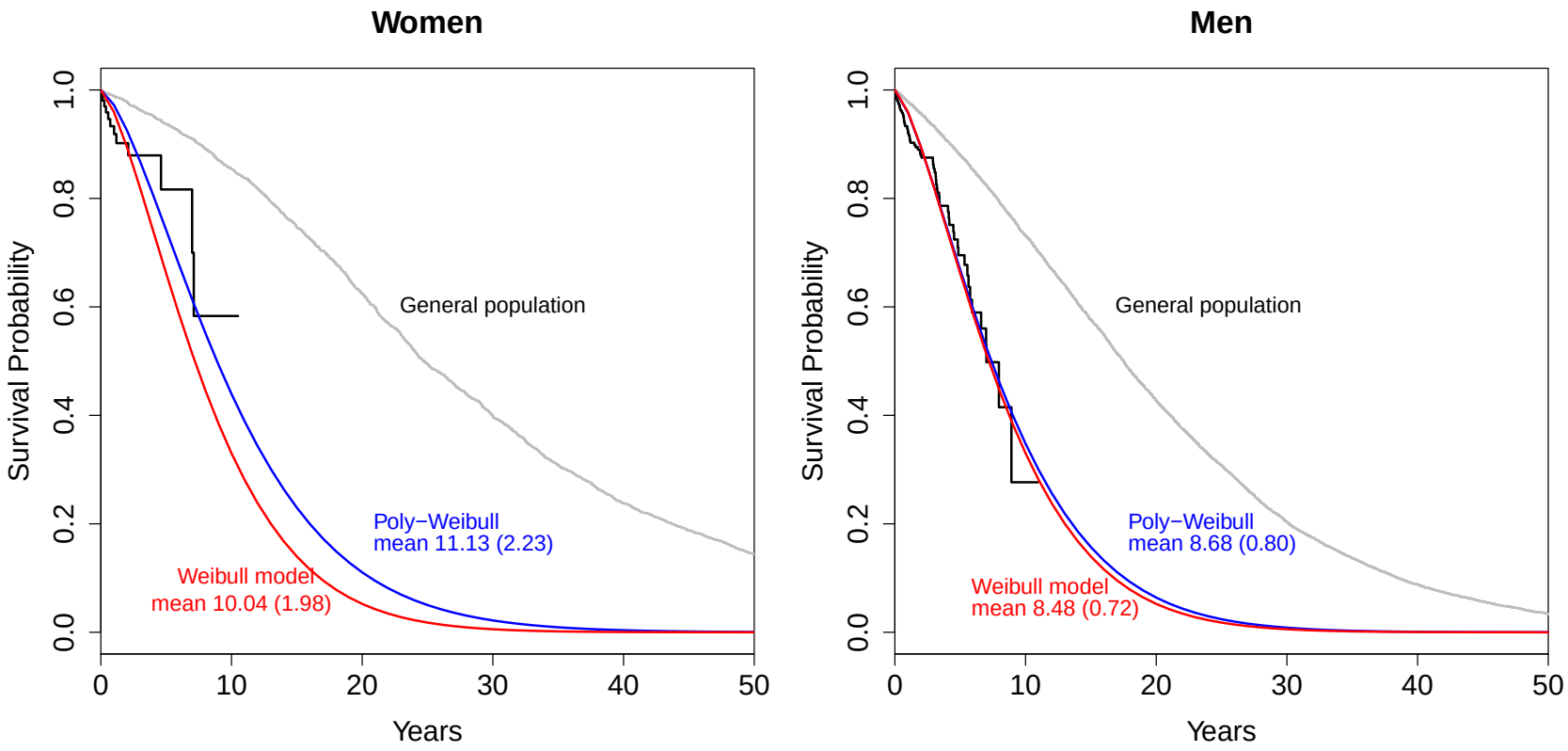
- Correct (poly-Weibull) model
- Weibull model which ignores cause of death

Simulated data and results: Weibull, HR = 4.5 / 20

	No increase	Slow	Fast
ICD:UK arr HR=4.5			
Bias	-0.1%	-5.8%	-28.4%
Coverage	94%	78%	0%
ICD:UK arr HR=20			
Bias	0.6%	-12.5%	-26.4%
Coverage	92%	52%	0%

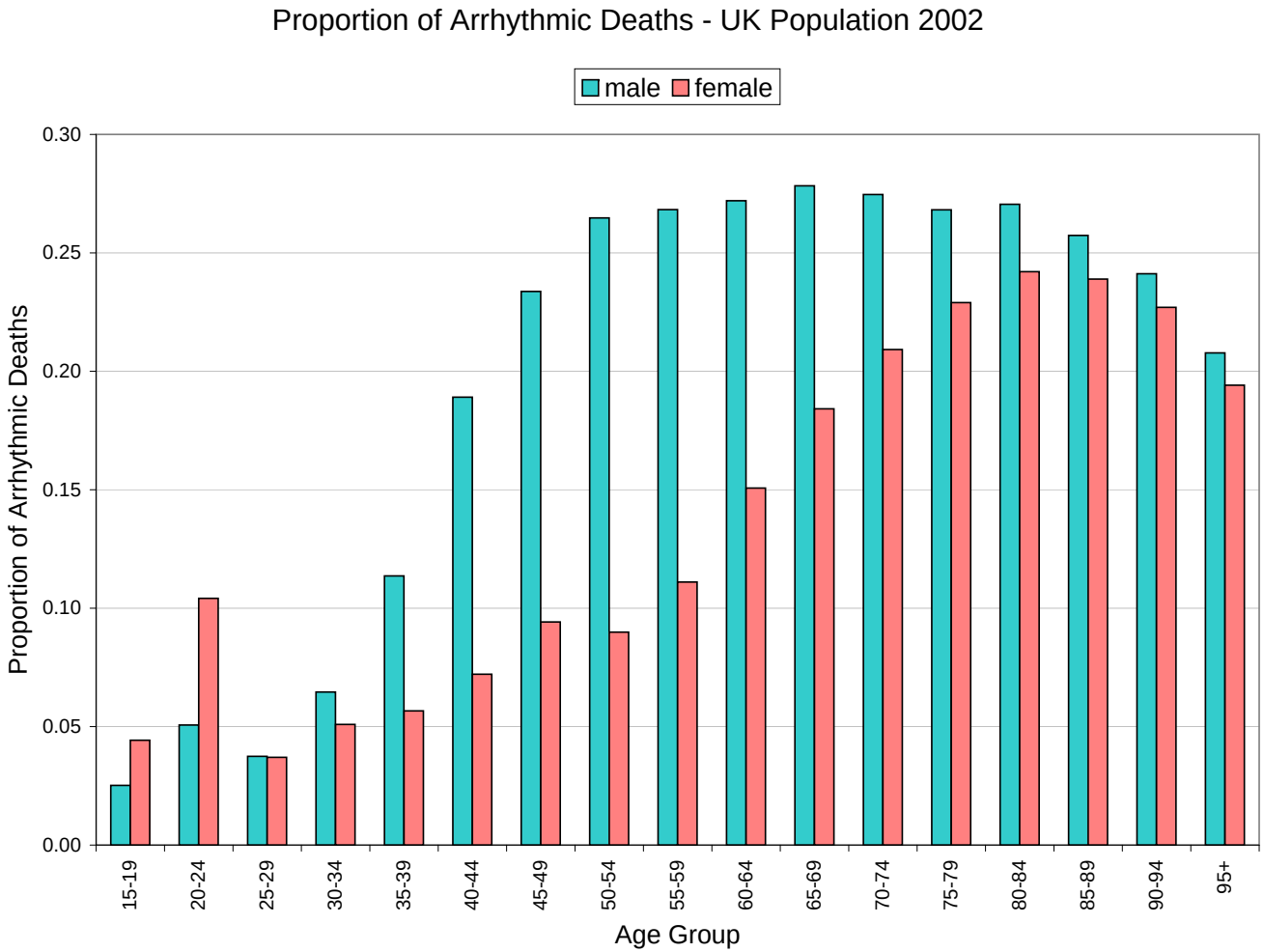
- Ignoring cause-specific hazard by fitting plain Weibull model gives bias in mean survival
  - particularly when hazard increases much quicker for other cause
- When HR between disease / general population bigger (HR 20 vs 4.5): similar bias from ignoring cause
  - Any changes through time are more marked ($\rightarrow$ more bias)
  - Overall HR better estimate of cause-specific HR ($\rightarrow$ less bias)

Extrapolating real ICD cohort data



- More bias for women when using Weibull instead of Poly-Weibull.
  - due to time-varying proportion of deaths due to arrhythmia.

Proportion of UK deaths which are due to arrhythmia



Including an intervention effect from literature

Hazards for three groups under Poly-Weibull model:

$$\begin{aligned} h_{\mathrm{UK}}(t) &= h^{arr}_{\mathrm{UK}}(t) + h^{other}_{\mathrm{UK}}(t) \\ h_{\mathrm{ICD}}(t) &= e^{\beta} h^{arr}_{\mathrm{UK}}(t) + h^{other}_{\mathrm{UK}}(t) \\ h_{\mathrm{AAD}}(t) &= e^{\gamma_a + \beta} h^{arr}_{\mathrm{UK}}(t) + h^{other}_{\mathrm{UK}}(t), \end{aligned}$$

Meta-analysis of ICD vs AAD trials, published HR for arrhythmia mortality, gives a prior for γ_a.

For the (probably biased) Weibull model we have:

$$\begin{aligned} h_{\mathrm{UK}}(t) &= \mu_1 \alpha t^{\alpha-1} = e^{\beta_0} \alpha t^{\alpha-1} \\ h_{\mathrm{ICD}}(t) &= e^{\beta_1} h_{\mathrm{UK}}(t) = e^{\beta_0 + \beta_1} \alpha t^{\alpha-1} \\ h_{\mathrm{AAD}}(t) &= e^{\gamma} h_{\mathrm{ICD}}(t) = e^{\beta_0 + \beta_1 + \gamma} \alpha t^{\alpha-1}, \end{aligned}$$

Prior for γ from published meta-analysis HR for all-cause mortality.

Outcome of interest $\rightarrow$ life years gained (LYG) by ICDs vs AADs.

Including an intervention effect from literature

Hazards for three groups under Poly-Weibull model:

$$\begin{aligned} h_{\text{UK}}(t) &= h^{arr}_{\text{UK}}(t) + h^{other}_{\text{UK}}(t) \\ h_{\text{ICD}}(t) &= e^{\beta} h^{arr}_{\text{UK}}(t) + h^{other}_{\text{UK}}(t) \\ h_{\text{AAD}}(t) &= e^{\gamma_a + \beta} h^{arr}_{\text{UK}}(t) + h^{other}_{\text{UK}}(t), \end{aligned}$$

Meta-analysis of ICD vs AAD trials, published HR for arrhythmia mortality, gives a prior for γ_a.

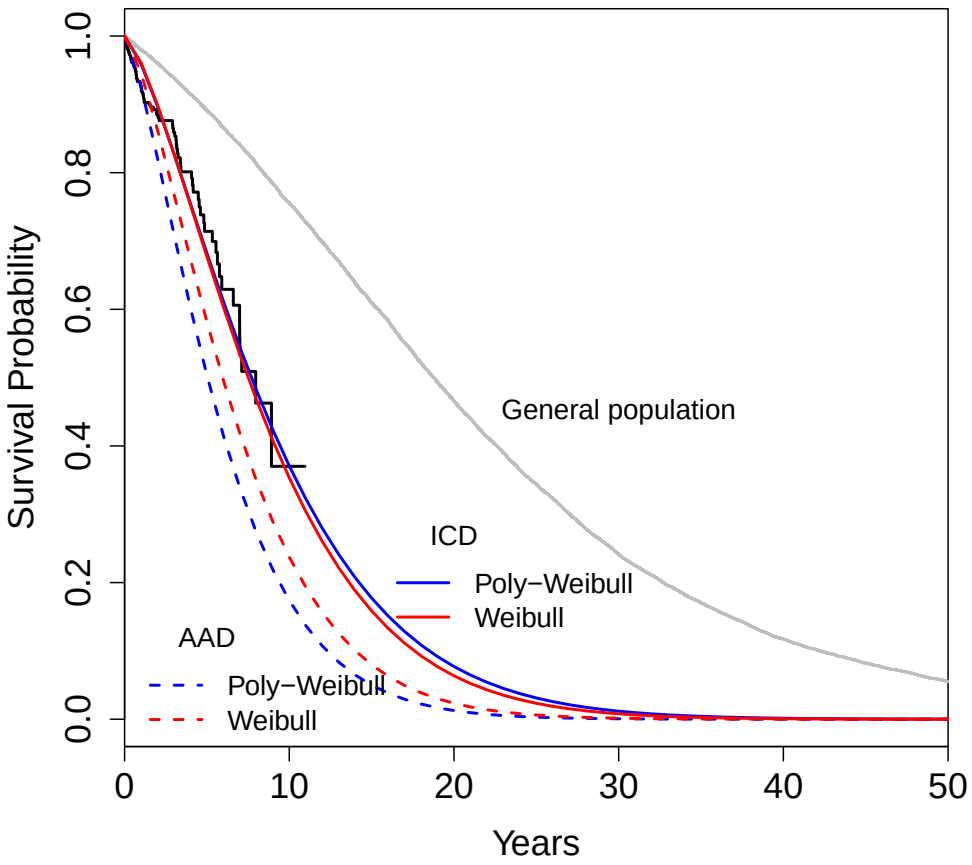
For the (probably biased) Weibull model we have:

$$\begin{aligned} h_{\text{UK}}(t) &= \mu_1 \alpha t^{\alpha-1} = e^{\beta_0} \alpha t^{\alpha-1} \\ h_{\text{ICD}}(t) &= e^{\beta_1} h_{\text{UK}}(t) = e^{\beta_0 + \beta_1} \alpha t^{\alpha-1} \\ h_{\text{AAD}}(t) &= e^{\gamma} h_{\text{ICD}}(t) = e^{\beta_0 + \beta_1 + \gamma} \alpha t^{\alpha-1}, \end{aligned}$$

Prior for γ from published meta-analysis HR for all-cause mortality.

Outcome of interest $\rightarrow$ life years gained (LYG) by ICDs vs AADs.

Extrapolating incremental survival between interventions

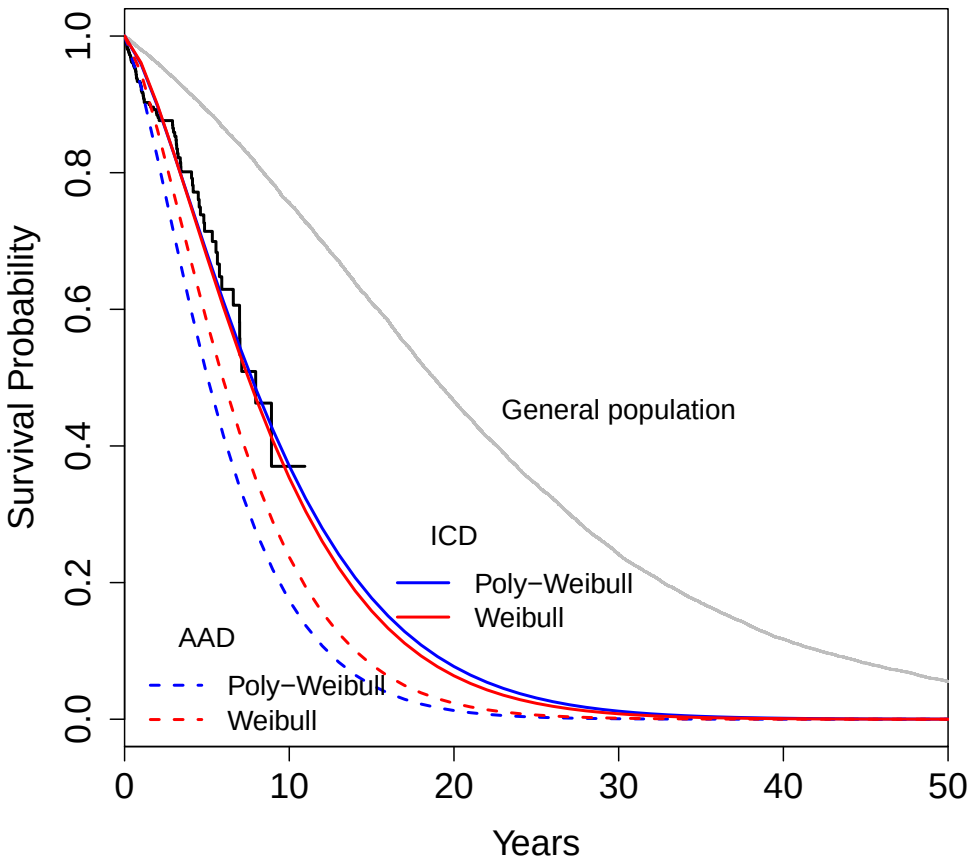


- ICD cohort extrapolated using population data
- AAD survival generated with aid of meta-analysis.
- Life-years gained from ICD appears biased if use Weibull
- Slightly more apparent bias for women

Life-years gained from ICD	Weibull	Poly-Weibull
Overall	1.82 (0.49)	3.12 (0.61)
Women	1.89 (0.62)	3.11 (0.76)
Men	1.73 (0.47)	2.91 (0.58)

. . . Still bias for men

Extrapolating incremental survival between interventions

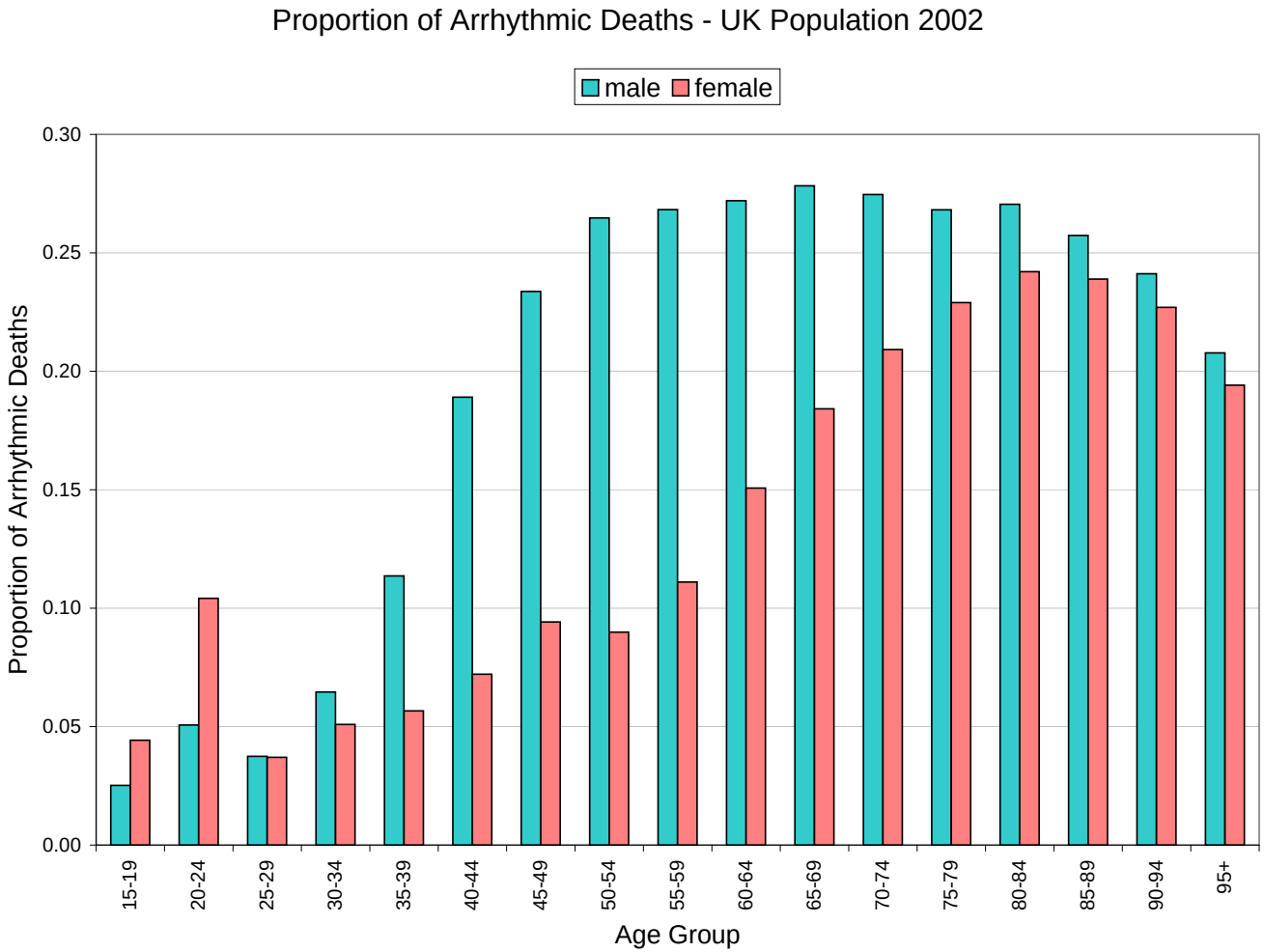


- ICD cohort extrapolated using population data
- AAD survival generated with aid of meta-analysis.
- Life-years gained from ICD appears biased if use Weibull
- Slightly more apparent bias for women

Life-years gained from ICD	Weibull	Poly-Weibull
Overall	1.82 (0.49)	3.12 (0.61)
Women	1.89 (0.62)	3.11 (0.76)
Men	1.73 (0.47)	2.91 (0.58)

. . . Still bias for men

Proportion of UK deaths which are due to arrhythmia



Risks of combining data (1) – data inconsistency

Causes of death may be recorded inconsistently between

- meta-analysis of ICD vs drug trials "HR for arrhythmia deaths"
- population mortality data

Sensitivity analysis — assume 10%-20% "arrhythmia" / "non-arrhythmia" deaths are misclassified.

- e.g. if fewer deaths actually affected by treatment, expected survival gains from treatment lower
- Still doesn't remove the bias for men.

Risks of combining data (2) — assumptions about hazards

We assumed ICD patients had

arrhythmia hazard proportional(=greater)
other-cause hazard identical
} to general population.

- What if ICD patients at greater risk from some other causes (other heart disease), as well as arrhythmia?
- May have led to biases in survival
  (underestimation of AAD-specific survival in poly-Weibull model. . .reasoning for this in paper)

Other issues: goodness of fit, prior sensitivity

- Fit of Weibull distribution — OK for our data — better than alternative Gompertz.
- Model baseline flexibly using semi-parametric model?
  - piecewise-constant hazard, Bayesian Cox-like model
  - doesn't alleviate bias from ignoring cause of death
- Tried "flat" prior for Weibull rate $\log(\lambda) \sim N(0, 1000)$
  - Expected survival 10% higher for women, compared to weakly informative prior.
  - Small sample: only 12 out of 104 died.
  - Flat prior includes unnaturally high survival times — better to include real information about human survival.

Other issues: goodness of fit, prior sensitivity

- Fit of Weibull distribution — OK for our data — better than alternative Gompertz.
- Model baseline flexibly using semi-parametric model?
  - piecewise-constant hazard, Bayesian Cox-like model
  - doesn't alleviate bias from ignoring cause of death
- Tried "flat" prior for Weibull rate $\log(\lambda) \sim N(0, 1000)$
  - Expected survival 10% higher for women, compared to weakly informative prior.
  - Small sample: only 12 out of 104 died.
  - Flat prior includes unnaturally high survival times — better to include real information about human survival.

Conclusions: survival extrapolation example

- Bayesian models useful for combining short-term RCT / cohort and longer-term survival data.
- Ignoring cause-specific hazard, thus misspecifying the underlying model, introduces bias in survival estimates.
  - may underestimate or overestimate overall survival.
- Bias can be alleviated by modelling cause-specific hazards
  - but requires cause-specific survival data / treatment effects
  - and information about which causes will be affected by disease status and / or treatment
- Bias for treatment comparisons may be less if bias acts in the same way in all treatment groups.
- Sensitivity analysis to model / data assumptions important

T. Benaglia, C. Jackson, L. Sharples *"Survival Extrapolation in the Presence of Cause Specific Hazards"*: revision under review for Stat. Med.

Conclusions: survival extrapolation example

- Bayesian models useful for combining short-term RCT / cohort and longer-term survival data.
- Ignoring cause-specific hazard, thus misspecifying the underlying model, introduces bias in survival estimates.
  - may underestimate or overestimate overall survival.
- Bias can be alleviated by modelling cause-specific hazards
  - but requires cause-specific survival data / treatment effects
  - and information about which causes will be affected by disease status and / or treatment
- Bias for treatment comparisons may be less if bias acts in the same way in all treatment groups.
- Sensitivity analysis to model / data assumptions important

T. Benaglia, C. Jackson, L. Sharples *"Survival Extrapolation in the Presence of Cause Specific Hazards"*: revision under review for Stat. Med.

Conclusions: survival extrapolation example

- Bayesian models useful for combining short-term RCT / cohort and longer-term survival data.
- Ignoring cause-specific hazard, thus misspecifying the underlying model, introduces bias in survival estimates.
  - may underestimate or overestimate overall survival.
- Bias can be alleviated by modelling cause-specific hazards
  - but requires cause-specific survival data / treatment effects
  - and information about which causes will be affected by disease status and / or treatment
- Bias for treatment comparisons may be less if bias acts in the same way in all treatment groups.
- Sensitivity analysis to model / data assumptions important

T. Benaglia, C. Jackson, L. Sharples *"Survival Extrapolation in the Presence of Cause Specific Hazards"*: revision under review for Stat. Med.

Conclusions: survival extrapolation example

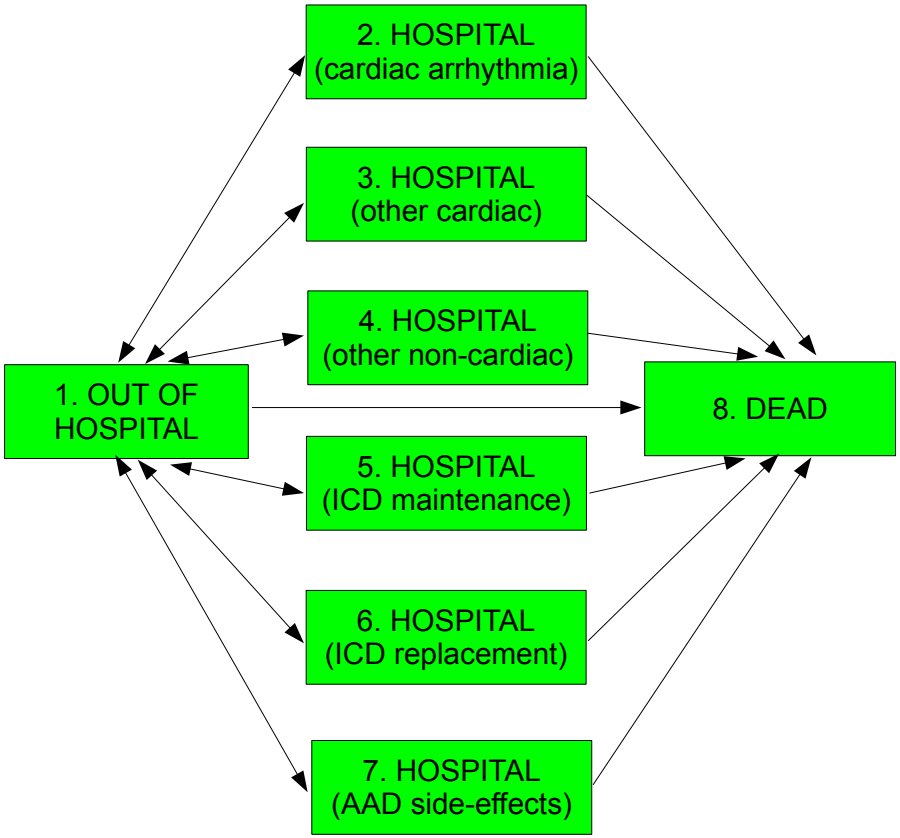
- Bayesian models useful for combining short-term RCT / cohort and longer-term survival data.
- Ignoring cause-specific hazard, thus misspecifying the underlying model, introduces bias in survival estimates.
  - may underestimate or overestimate overall survival.
- Bias can be alleviated by modelling cause-specific hazards
  - but requires cause-specific survival data / treatment effects
  - and information about which causes will be affected by disease status and / or treatment
- Bias for treatment comparisons may be less if bias acts in the same way in all treatment groups.
- Sensitivity analysis to model / data assumptions important

T. Benaglia, C. Jackson, L. Sharples *"Survival Extrapolation in the Presence of Cause Specific Hazards"*: revision under review for Stat. Med.

Part II

Structural uncertainty in health economic models

Health economic models to estimate costs and effects

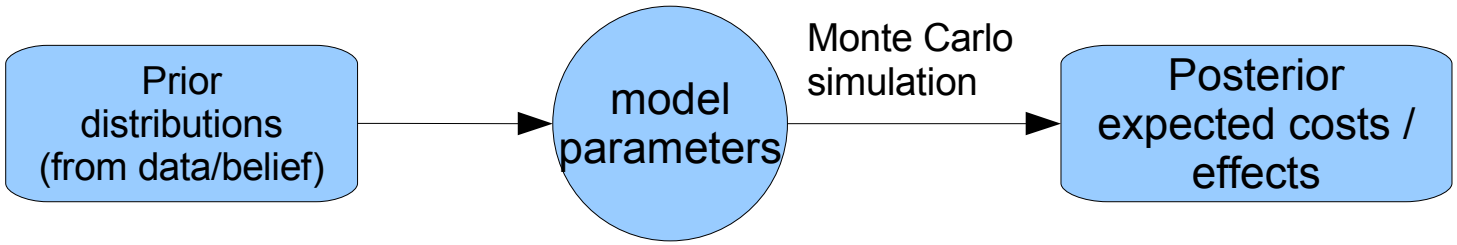


(Example: cost-effectiveness of implantable defibrillators for cardiac arrhythmia)

- Commonly Markov models for clinical history
- Each state / event associated with a cost or detriment to quality of life
- Combine all relevant evidence on disease and treatment – randomised trials → meta-analyses, observational data, national registries...

Standard procedure for economic modelling

- Choose states to represent important events
  - Which are relevant to the decision?
- Identify the parameters of the model
  - transition rates between states
  - cost and quality of life for each state
- Identify how these parameters vary
  - between patients and through time
  - What covariates? What time-dependence?
- Estimate parameters from data or expert belief
- Account for ensuing parameter uncertainty probabilistically:



Standard procedure for economic modelling

- Choose states to represent important events
  – Which are relevant to the decision?
- Identify the parameters of the model
  - transition rates between states
  - cost and quality of life for each state
- Identify how these parameters vary
  - between patients and through time
  – What covariates? What time-dependence?
- Estimate parameters from data or expert belief
  Which data are relevant? What if no data?
- Account for ensuing parameter uncertainty probabilistically

But what should be modelled?

Structural uncertainty / model uncertainty

Accounting for structural uncertainty informally

Cost-effectiveness often presented for "best case" assumptions ...

- understates uncertainty – may be biased

often alongside alternative scenarios

- with little indication of plausibility of each one

If possible, should express structural uncertainty in a formal probabilistic way

Accounting for structural uncertainty informally

Cost-effectiveness often presented for "best case" assumptions ...

- understates uncertainty – may be biased

often alongside alternative scenarios

- with little indication of plausibility of each one

If possible, should express structural uncertainty in a formal probabilistic way

Setting up the model

- Set up and parameterise model
- Expand model to encompass structural uncertainties
  - Extra parameters whose values represent structural choices
  - Include as many states / events as might affect the decision
  - Allow parameters to vary with as many covariates as might be relevant

Problem...

no data / not enough data to form confident (posterior) distributions on some parameters

Setting up the model

- Set up and parameterise model
- Expand model to encompass structural uncertainties
  - Extra parameters whose values represent structural choices
  - Include as many states / events as might affect the decision
  - Allow parameters to vary with as many covariates as might be relevant

Problem...

no data / not enough data to form confident (posterior) distributions on some parameters

Examples of structural uncertainty (1): Variable selection

Is there any treatment effect on a particular health event?

Fix treatment effect at $\beta = 0$?
Or use a weakly-informed posterior $\beta \sim f()$?

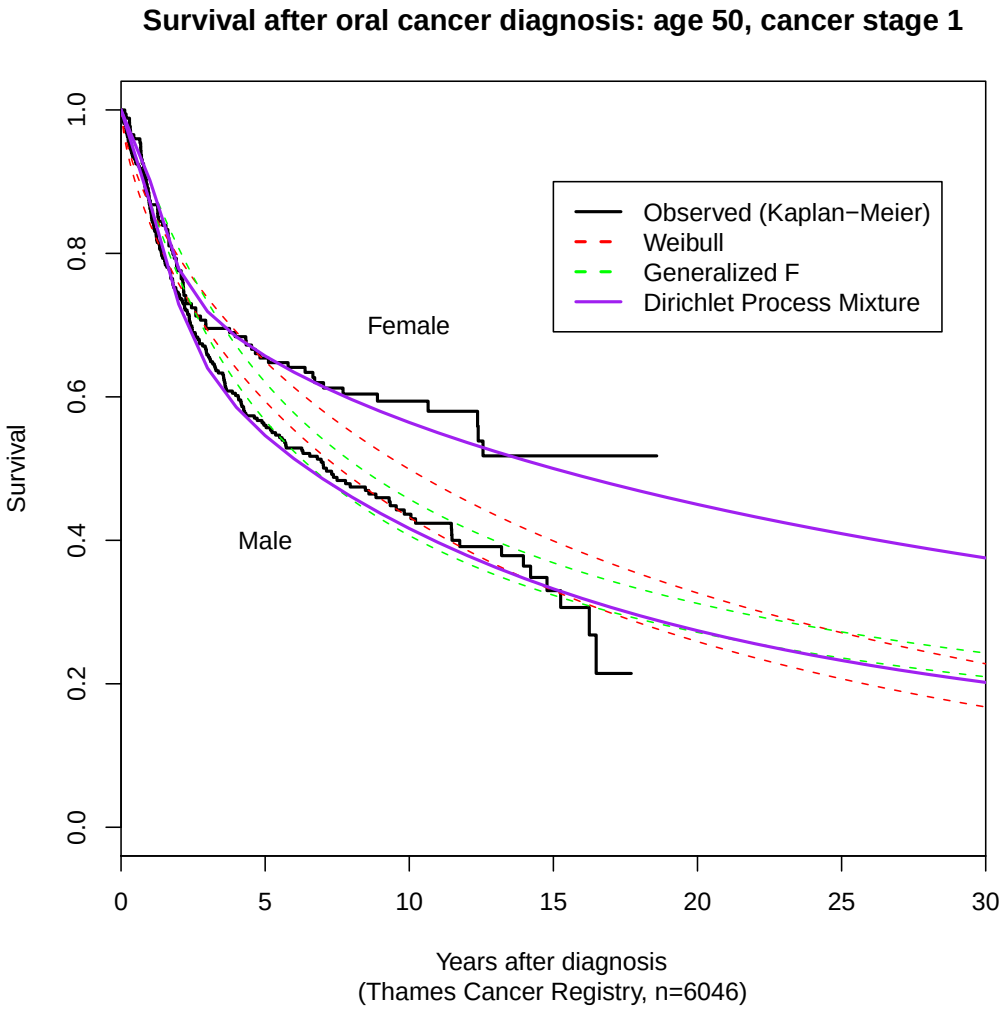
1. Model averaging:
   - compute model choice criteria (e.g. DIC, AIC, Bayes factors)
   - express as weights w_1, w_2 to form mixture posterior

$$\beta \sim w_1 I_{\beta=0} + w_2 f()$$

(Jackson et al, (2010) J.R. Stat. Soc. C 59(2); (2009) J. R. Stat. Soc A 172(2))

2. Shrinkage:
   - use $f()$ obtained from a prior that "smoothly" shrinks covariate effect(s) β towards zero. (lasso, elastic net etc.)

Examples (2): Dependence of parameters on time

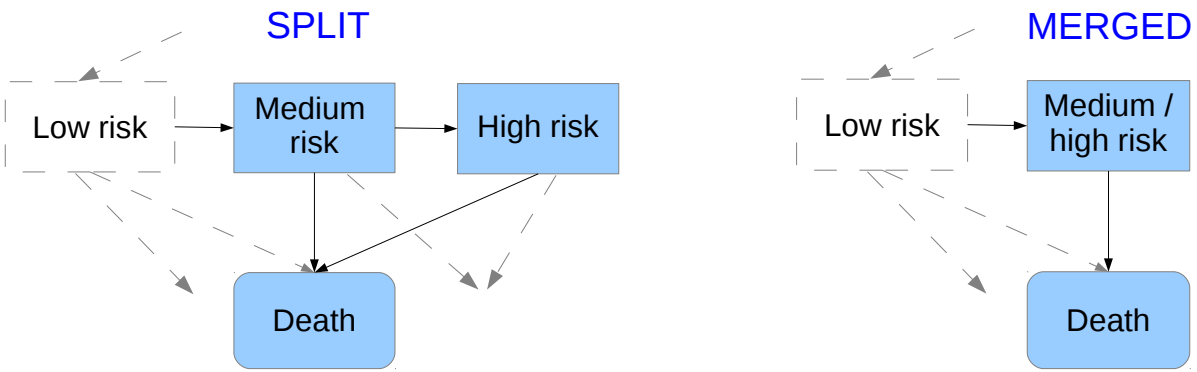


(data from Jackson et al, (2010) Int. J. Biostat. 6(1))

What parametric survival model?

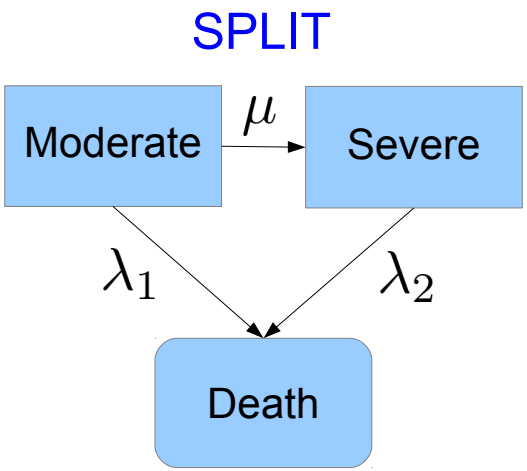
- Flexible distributions (3-4 parameters) have e.g. Weibull as special cases
- Spline-based models (Royston & Parmar, Stat. Med. 2002)
- Bayesian non-parametrics: Dirichlet process (e.g. De Iorio et al., Biometrics 2009)

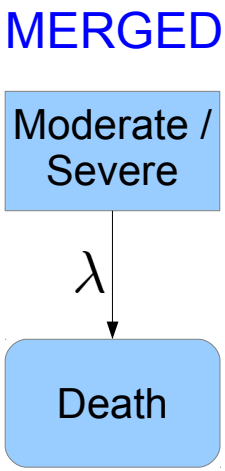
Examples (3): uncertainty about state structure



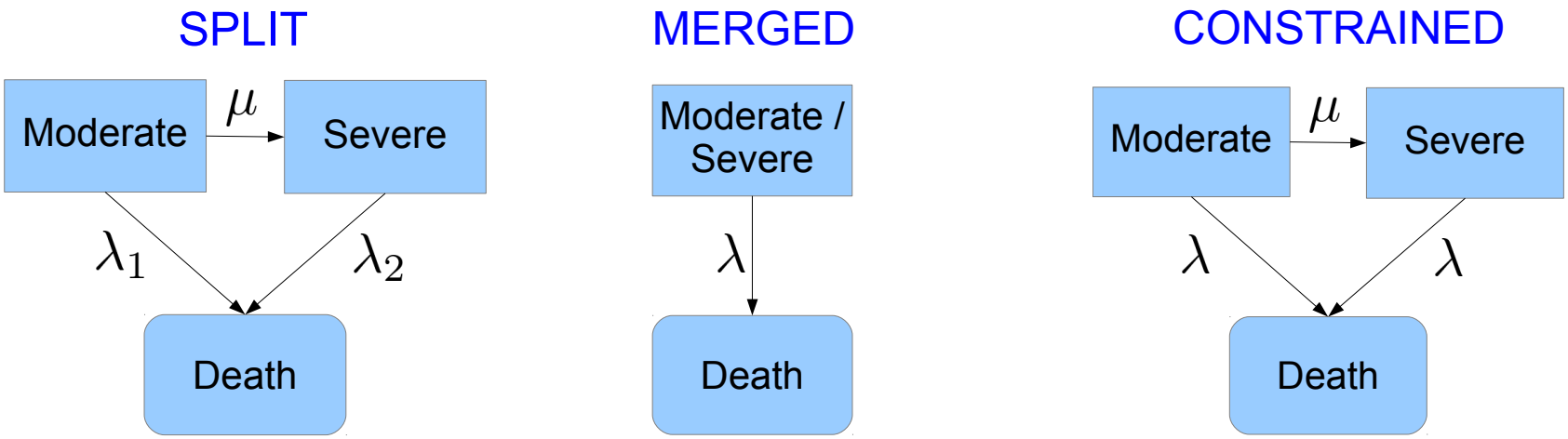
- Split or merge medium / high risk coronary disease?
- Which model gives better estimates of expected quality-adjusted survival, lifetime cost?
- Are costs / quality of life / death rates in medium / high risk sufficiently different?
  $\leftrightarrow$ Enough data to distinguish them?

Choice of states expressed as parameter uncertainty





Choice of states expressed as parameter uncertainty



CONSTRAINED: risk of death λ whether in medium or high

- Same survival distribution as MERGED
- $\rightarrow$ practically equivalent – gives same answer of interest
- even though theoretically different
  - different sample space / likelihood

Can compare SPLIT vs CONSTRAINED by standard methods (DIC, AIC, Bayes factors, shrinkage...)

Conclusions: structural uncertainty

- If different plausible structures give different answers (expected cost / survival) then there's structural uncertainty.
  - conversely: if they give same answers $\rightarrow$ no uncertainty!
- Structural uncertainty can be accounted for by parameterising the uncertainty in a bigger model
  - e.g. include all potential covariates, or use very flexible distributions
  - statistical model uncertainty methods lead to posterior for "structural" parameter(s).
  - posterior expresses strength of belief / evidence for each choice.
  - Jackson et al. (2011) *A framework for addressing structural uncertainty in decision models*, Med. Decis. Making 31(4).
- Choice of what evidence should be included? No good evidence on some parameter?
  - "Softer" methods important here (sensitivity analysis, elicitation).

Conclusions: structural uncertainty

- If different plausible structures give different answers (expected cost / survival) then there's structural uncertainty.
  - conversely: if they give same answers $\rightarrow$ no uncertainty!
- Structural uncertainty can be accounted for by parameterising the uncertainty in a bigger model
  - e.g. include all potential covariates, or use very flexible distributions
  - statistical model uncertainty methods lead to posterior for "structural" parameter(s).
  - posterior expresses strength of belief / evidence for each choice.
  - Jackson et al. (2011) *A framework for addressing structural uncertainty in decision models*, Med. Decis. Making 31(4).
- Choice of what evidence should be included? No good evidence on some parameter?
  - "Softer" methods important here (sensitivity analysis, elicitation).