W11: Science, art or arbitrariness? Evaluating the risk of treatment effect waning for novel oncology therapies

Raquel Aguiar-Ibáñez MSc
Merck Canada Inc., Kirkland, QC, Canada
Dawn Lee MMath, MSc
BresMed Health Solutions, Sheffield, UK
Gianluca Baio PhD
University College London, London, UK
Nicholas Latimer PhD
University of Sheffield, Sheffield, UK

Conflicts of interest and disclosures

No external funding was received for the development of this workshop

Raquel Aguiar-Ibáñez	Employee of Merck Canada Inc., a subsidiary of Merck & Co., Inc., Kenilworth, NJ, USA.
Dawn Lee	No conflicts of interest relevant to the content of this workshop. Employed by BresMed Health Solutions, UK.
Gianluca Baio	No conflicts of interest relevant to this presentation.
Nicholas Latimer	No conflicts of interest relevant to the content of this workshop.

Agenda

Raquel Aguiar-Ibáñez	Treatment effect waning: the why and the what in the HTA-related cost-effectiveness modelling of immuno-oncology therapies
Dawn Lee	Incorporating uncertainty around long-term treatment effects into economic analyses
Gianluca Baio	Who wants to be a Bayesian (and why you should)
Nicholas Latimer	A decision-making perspective

Treatment effect waning
The why and the what

Raquel Aguiar-Ibáñez, Senior Principal Scientist – Oncology

Center for Observation and Real World Evidence (CORE),
Economic and Data Sciences (EDS), Merck Canada Inc., Kirkland, QC, Canada

Disclaimer

- I am an employee of Merck Canada Inc., a subsidiary of Merck & Co., Inc., Kenilworth, NJ, USA.
- The views and opinions presented here are my own and do not necessarily reflect those of Merck Canada Inc., Kirkland, QC, Canada.

What is treatment effect waning?

The potential for a therapy that is initially effective to become less effective over time when compared to an alternative

Why is treatment effect waning an issue in CEMs for immuno-oncology (IO) therapies?

- Classical chemotherapies are problematic due to resistance
- IOs have high therapeutic value, but some potential for resistance
- Limited trial follow up with unknown long-term effect (yet…)
- Need to extrapolate to a lifetime (in many HTAs)

What are HTA agencies advising/requesting regarding treatment effect waning?

- Describe and justify assumptions
- Present alternative scenarios on benefit duration
- Present alternative time horizons

Long-term treatment effect scenarios across HTAs

NICE

- Same as treatment phase
- Diminishes in the long term
- Nil
- Threshold analysis for duration of treatment effect*

*Under consultation

CADTH

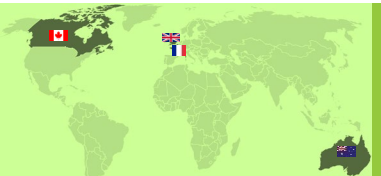
- Continues while data available
- Declines over time until:
  - Stabilizing
  - No further effective
- Treatment effect while on treatment

HAS

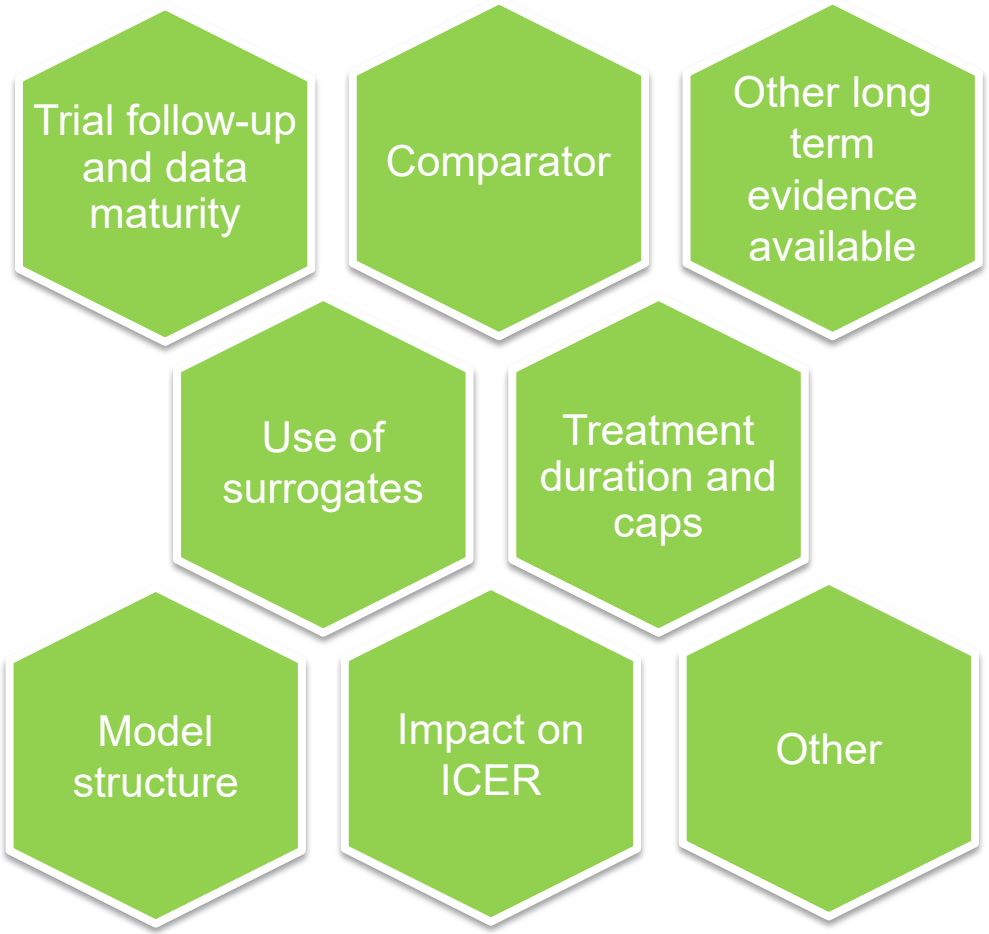
- Nil after observation period
- Nil after treatment period
- Diminishes over time

PBS

- HR for intervention and comparator to converge at plausible point
- Assess plausibility and link to time horizon
- Force convergence from median follow up to a clinically plausible time
- Consider clinical decision regarding treatment cessation or continuation



What factors may affect considerations of treatment effect waning?



Main take-away messages

Agreement across main HTA agencies on assessing treatment effect waning

Limited guidance on when (and how?) to implement it

Realistic scenario analyses are needed on alternative treatment effect durations and time horizons

Our people.
Your success.

BresMed™

Incorporating uncertainty around long-term treatment effects into economic analyses

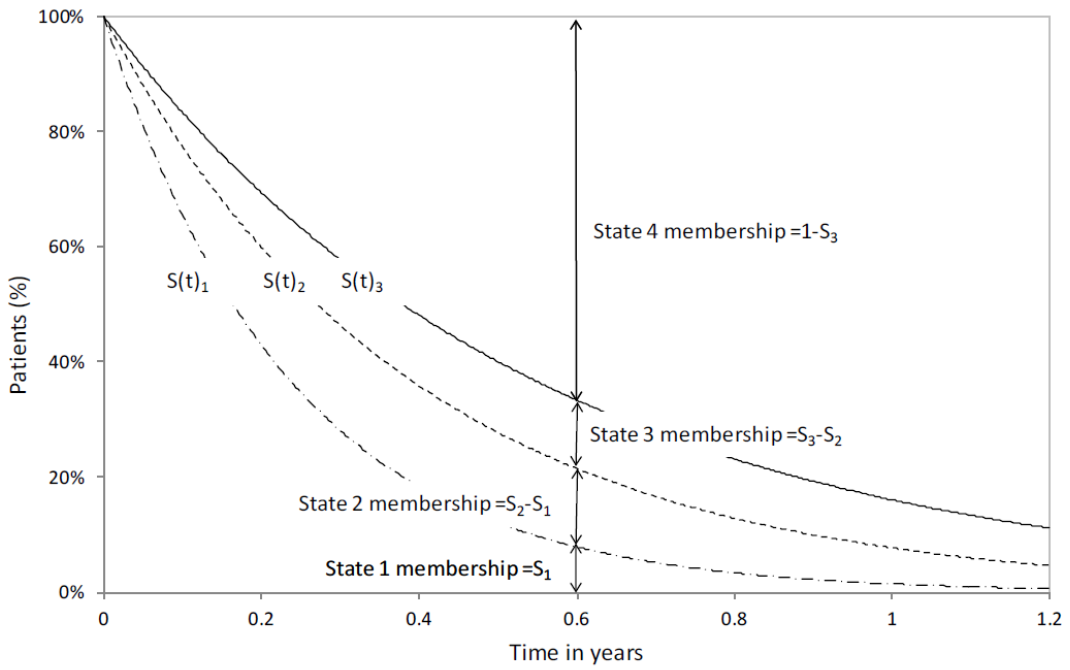
Dawn Lee
BresMed

What drives long term effectiveness?

> *"Careful thought should be given to the biological and clinical justification to any statistical approach selected"*
>
> NICE TSD21

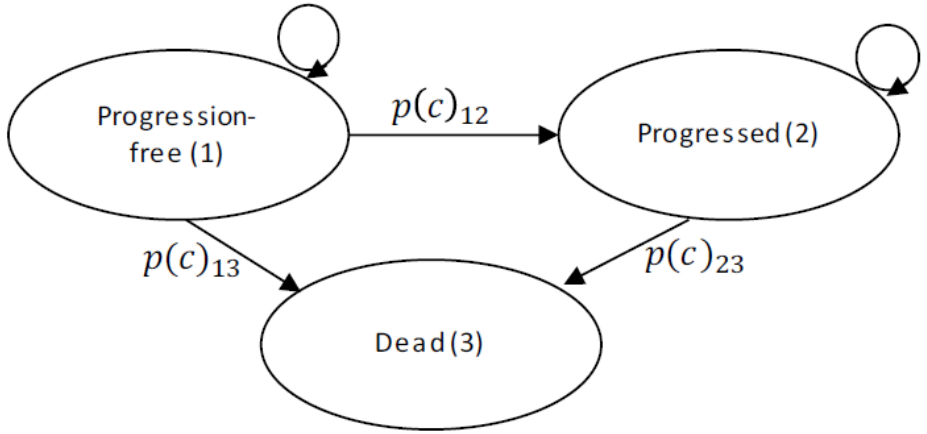
Application of treatment-waning differs by model structure

Partitioned Survival Analysis



Transitions are independent – potential for curves crossing

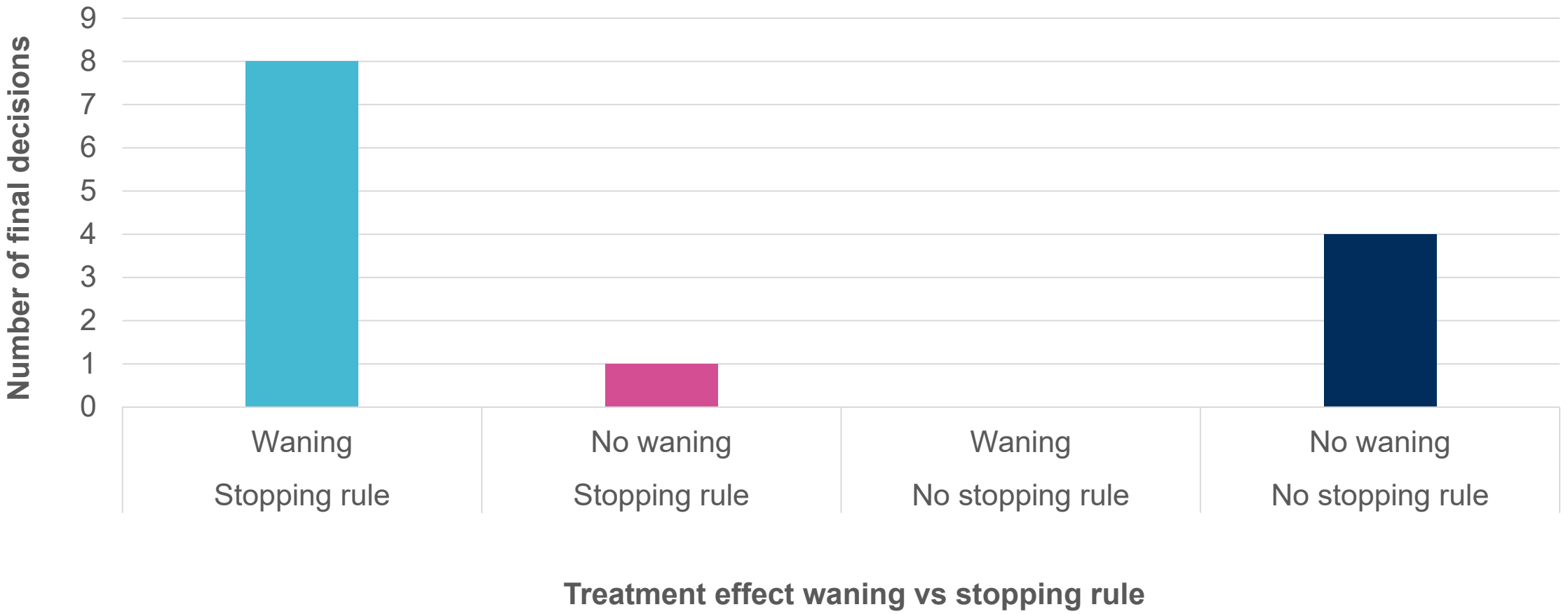
State Transition Model



Dependency incorporated

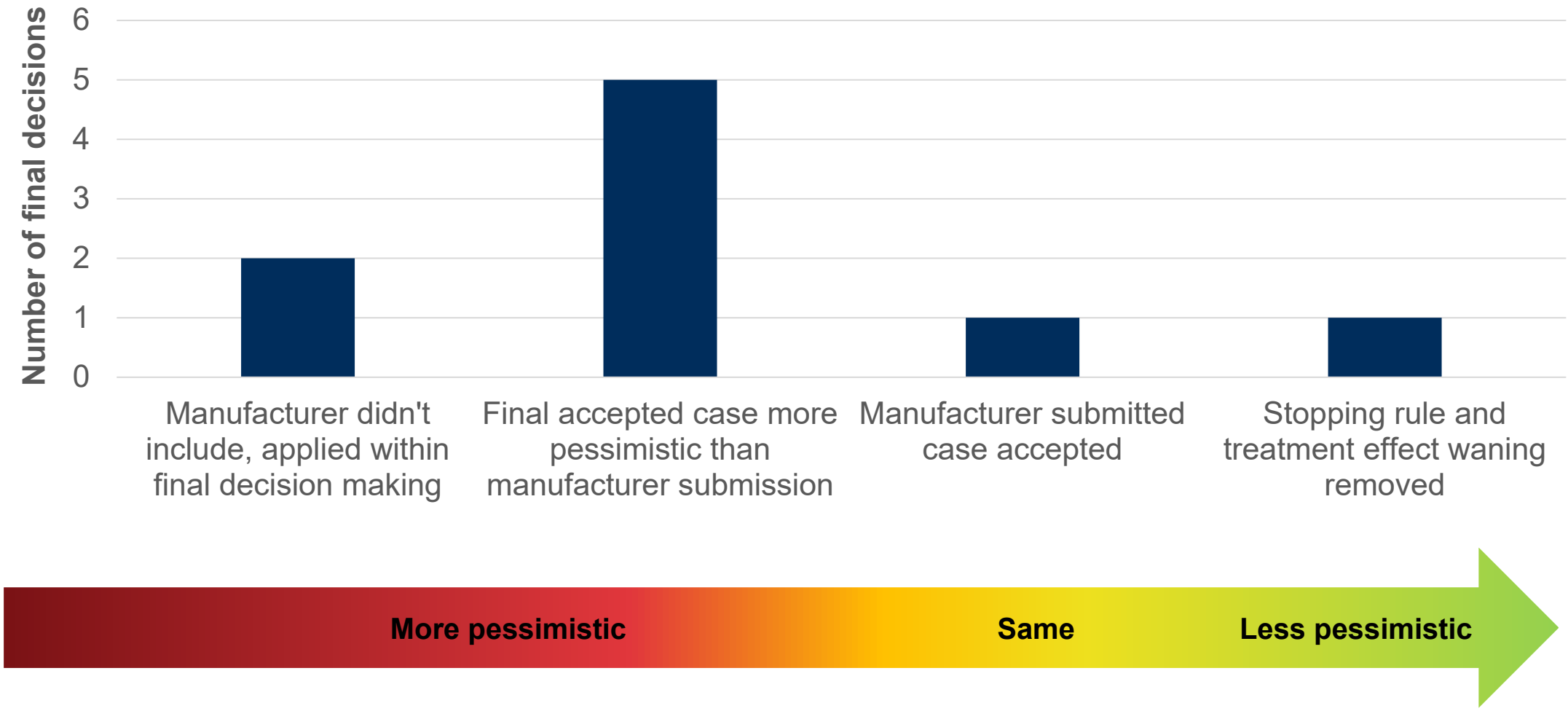
↘ Carefully consider: (1) which transitions any waning effect should be applied to; (2) the potential for curves crossing in a PartSA analysis

Treatment effect waning is commonly assumed in IO-models submitted to NICE where a stopping rule applies



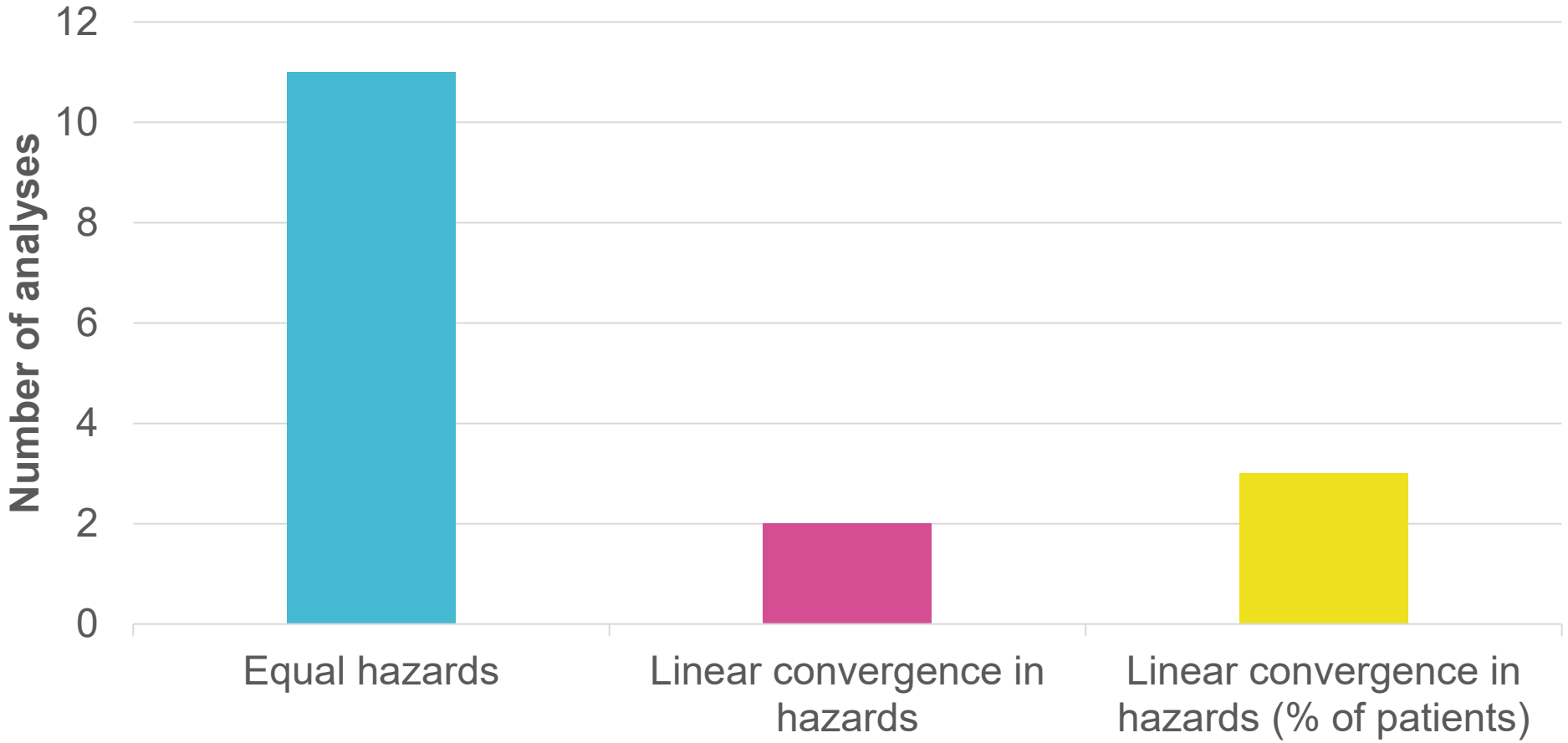
Review of 13 NICE immuno-oncology Technology Appraisals completed and not superceded between 1 Jan 2019 and 23 Feb 2021

However, NICE are typically more pessimistic than manufacturers



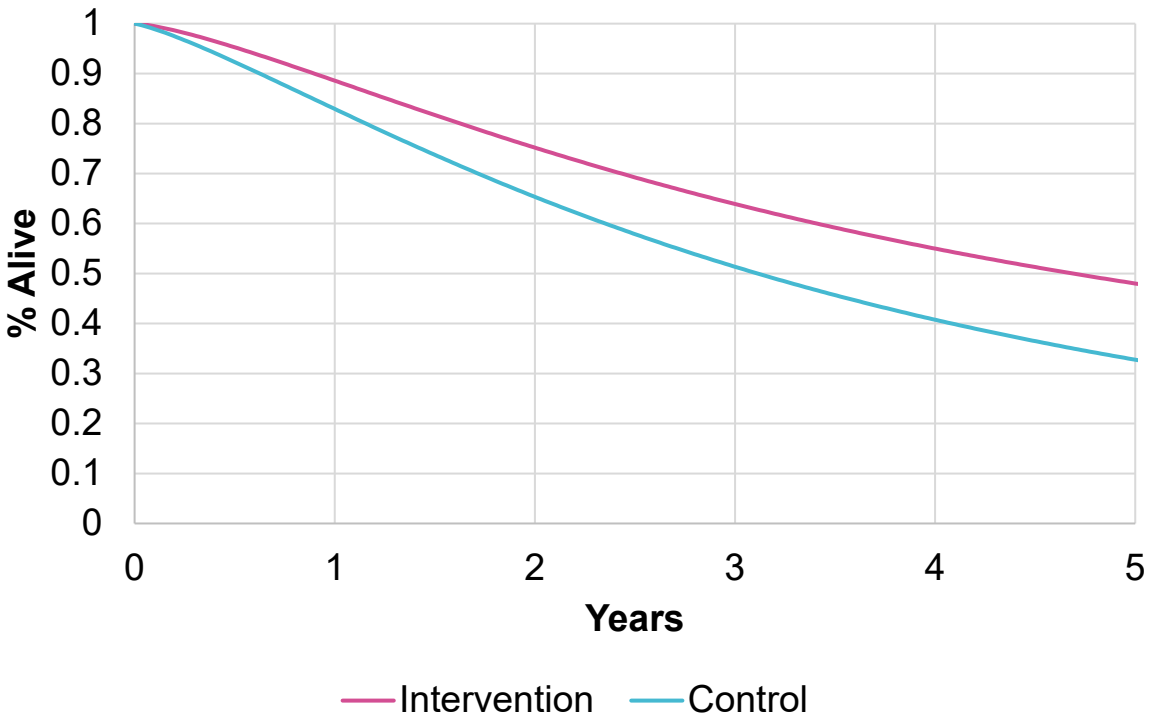
Review of 13 NICE immuno-oncology Technology Appraisals completed and not superceded between 1 Jan 2019 and 23 Feb 2021

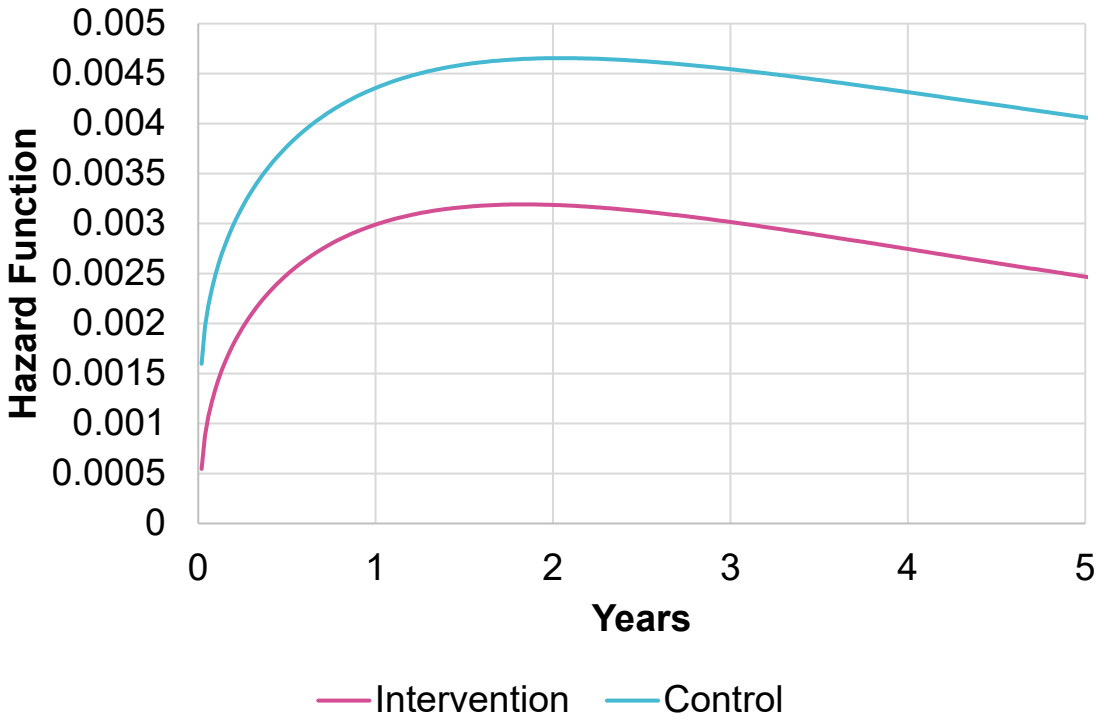
Typically, modelers assume equal hazards post-treatment effect (i.e. HR=1)



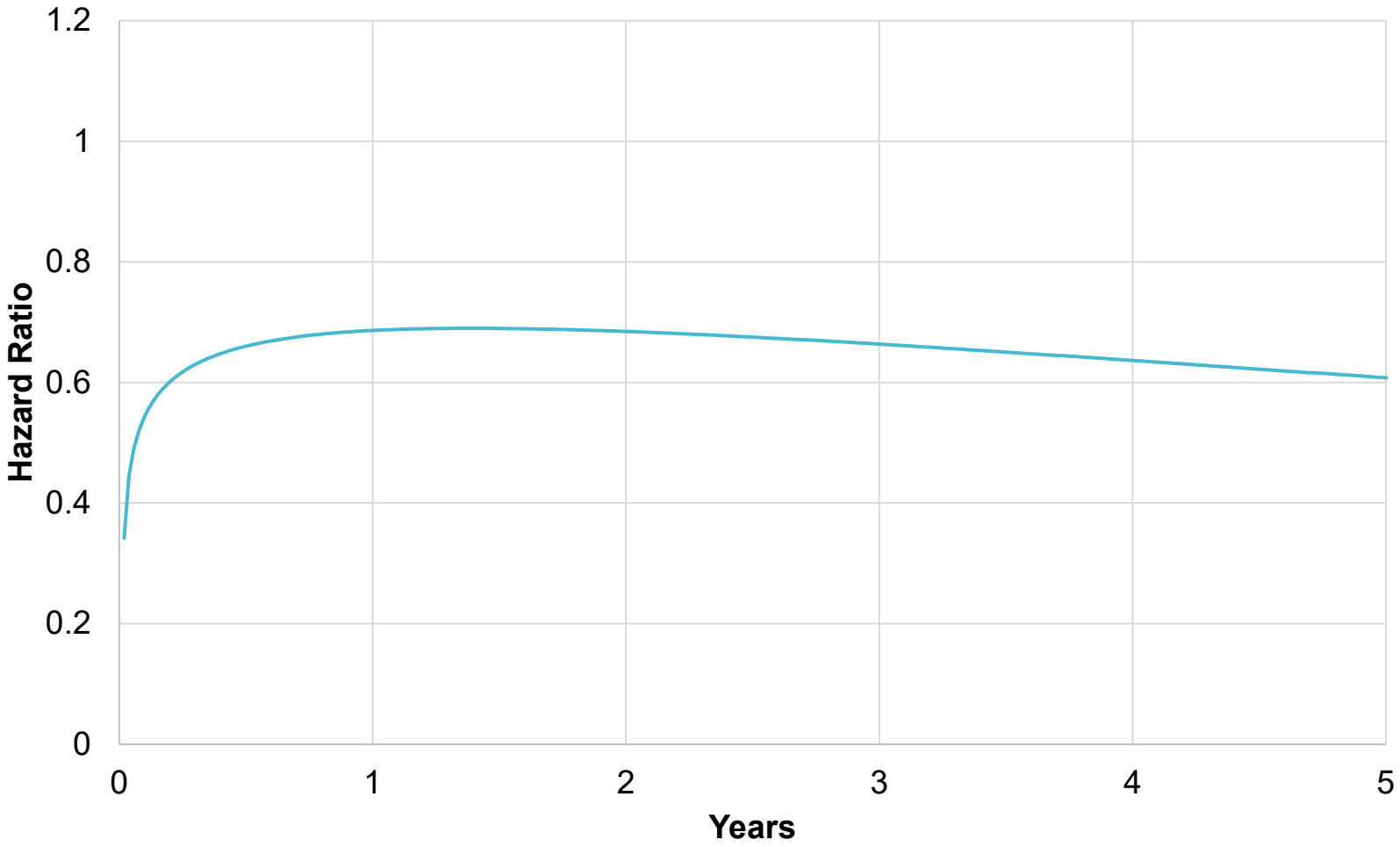
Review of 13 NICE immuno-oncology Technology Appraisals between 1 Jan 2019 and 23 Feb 2021

A typical example: immature trial data





A typical example: immature trial data



What hazard function do we expect for IOs?

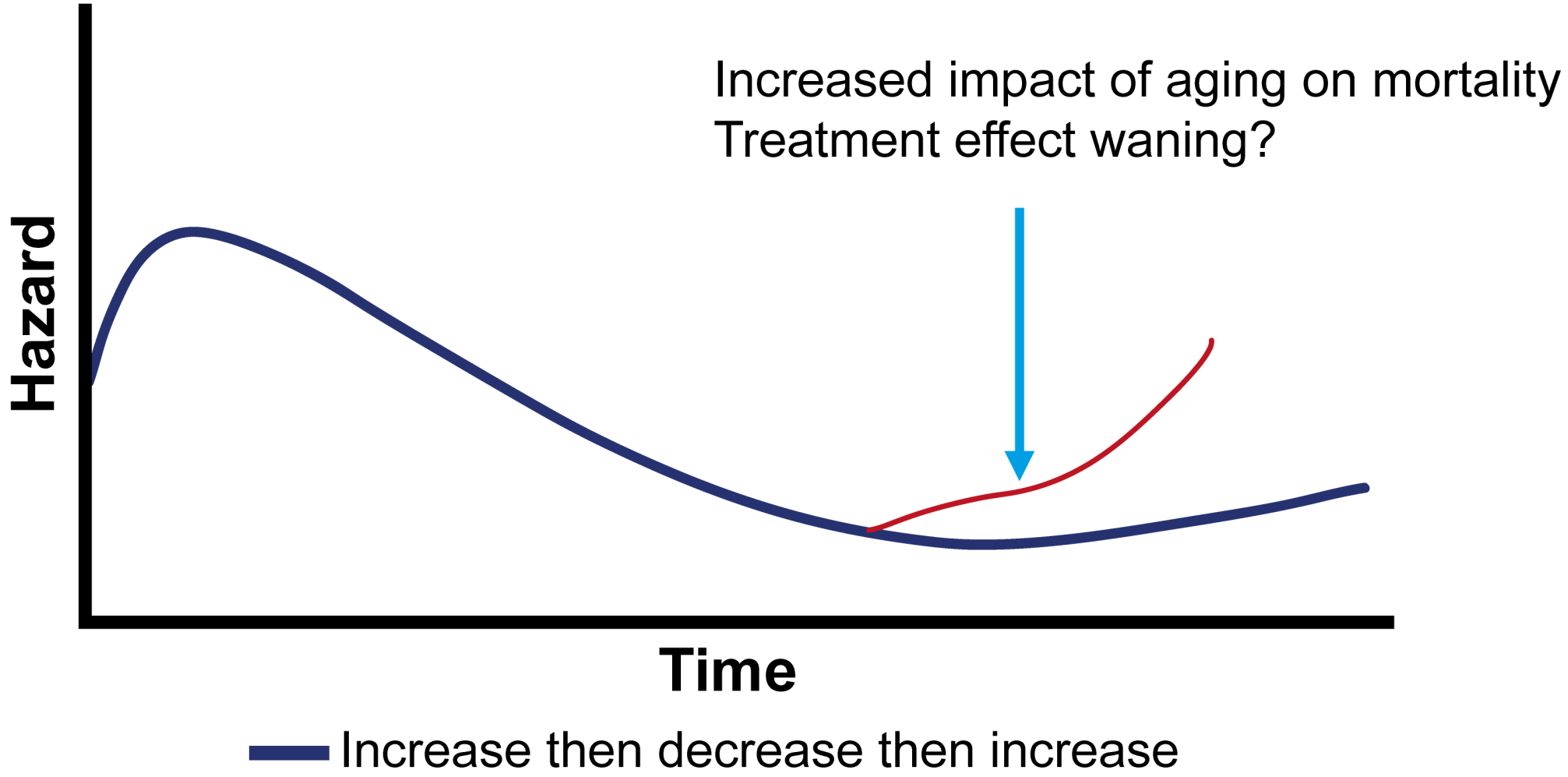
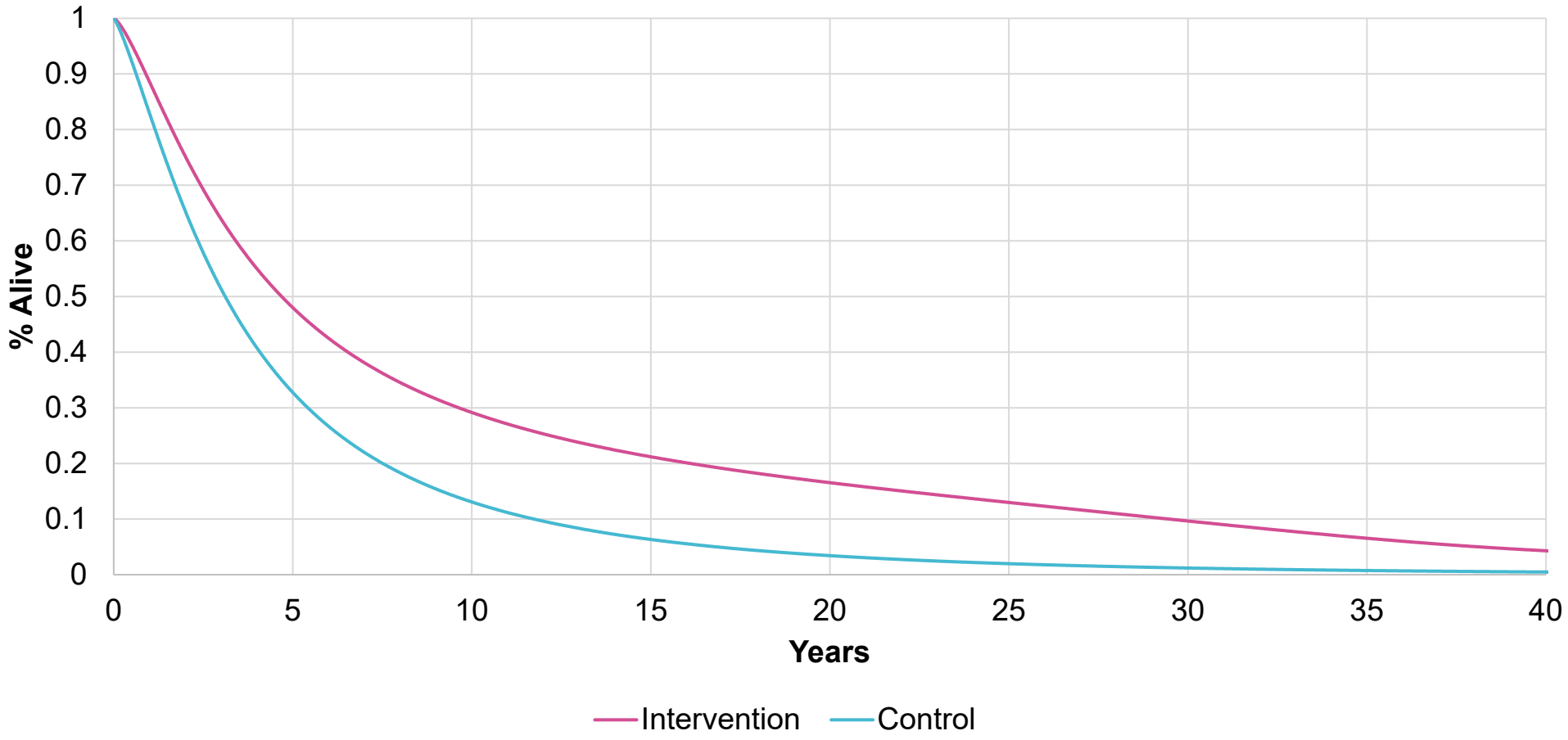


Diagram amended from TSD21

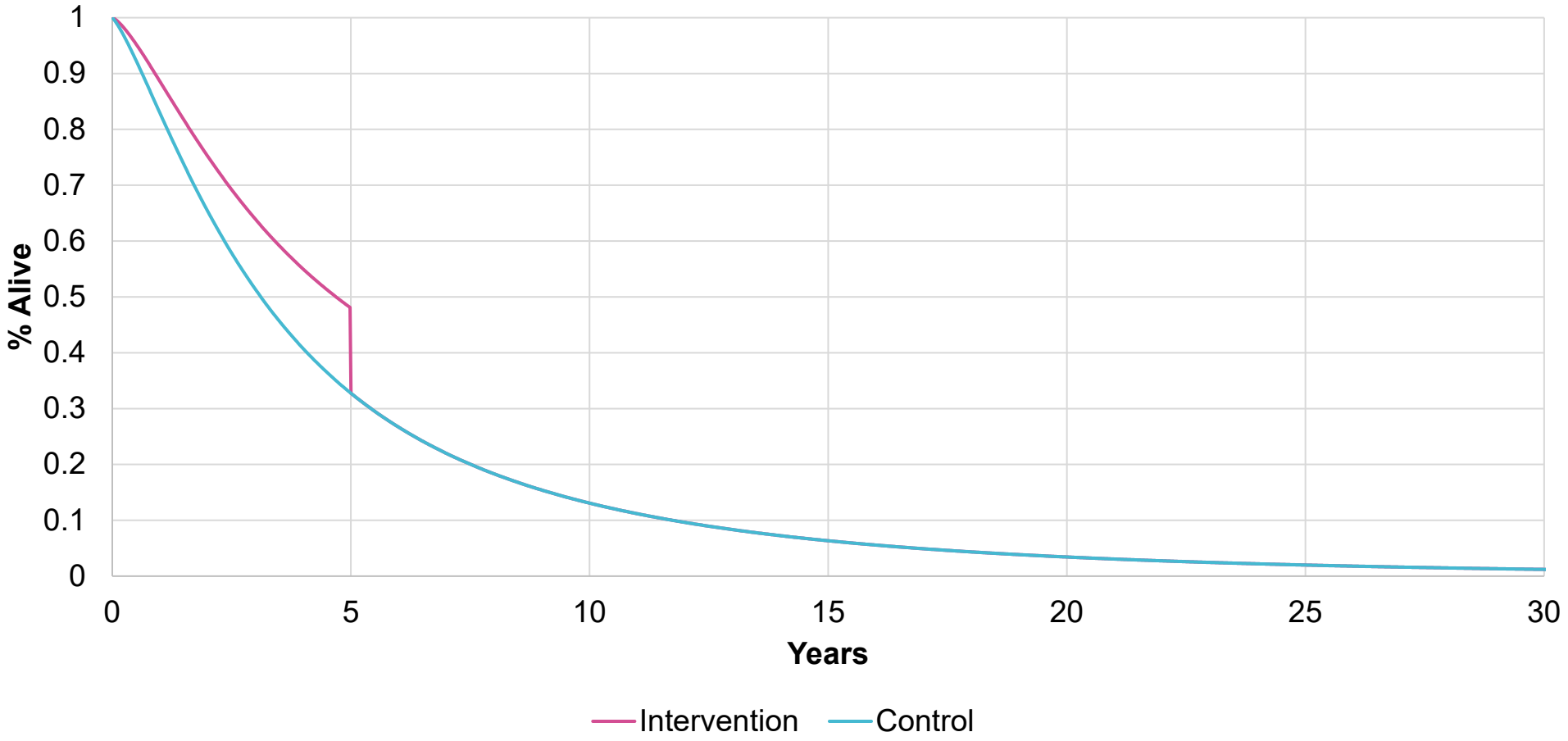
If I just continue my survivor functions (accounting only for age)….



ΔLY = 5.1

Large benefit in life years gained here

Most of the benefit comes from the extrapolated period…

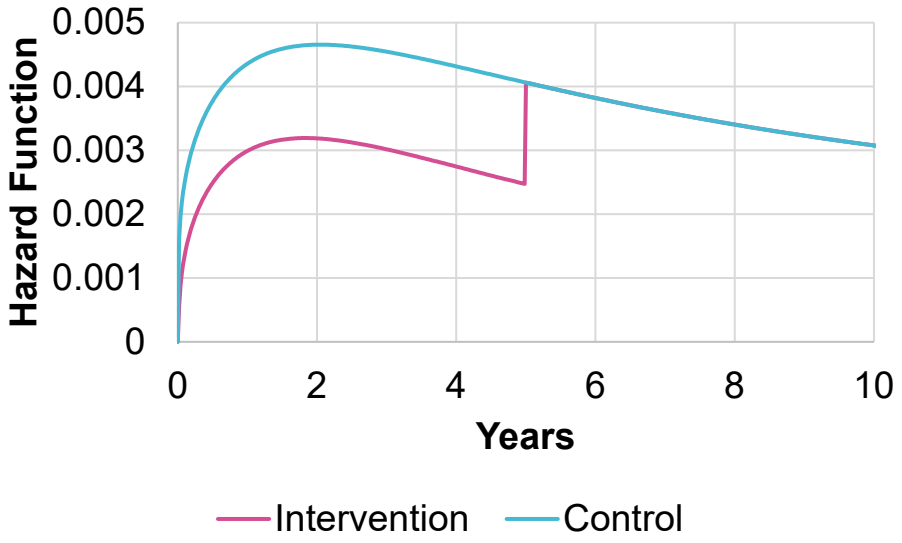


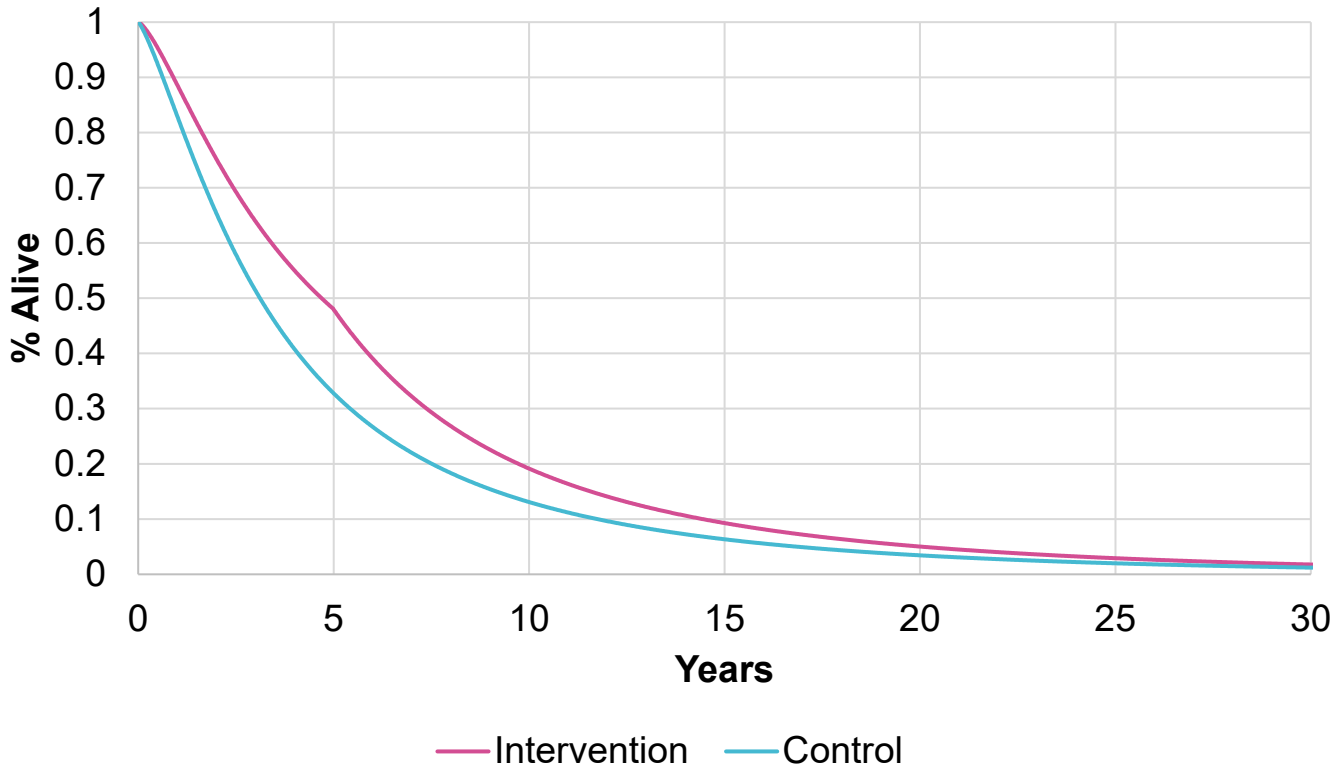
$\Delta LY = 0.5$

Not realistic but informative

Setting hazards equal after 5 years

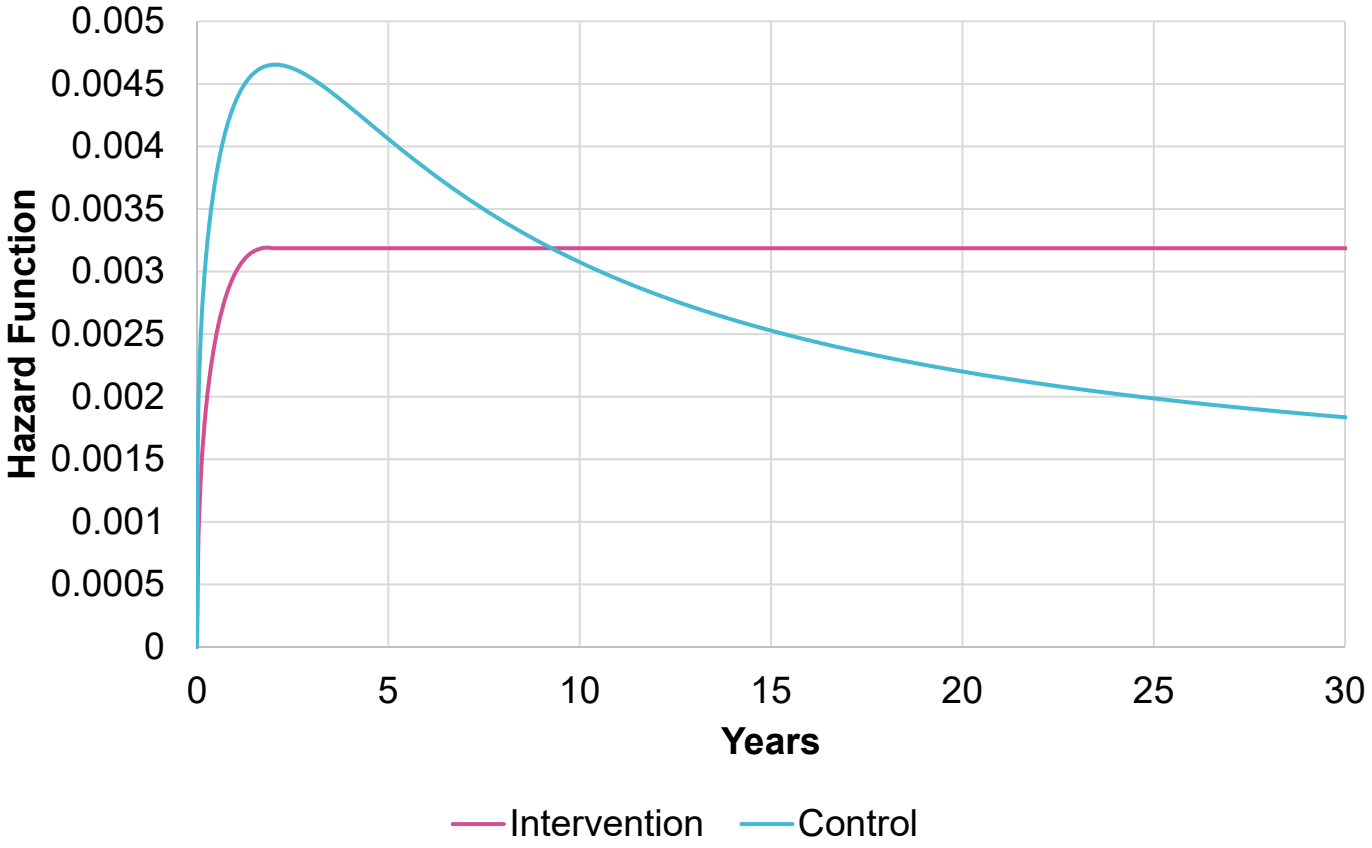
ΔLY = 1.5





Again not very realistic: most common method used in NICE submissions currently

What if comparator & intervention hazard functions cross?



"Setting the hazard ratio to one at the chosen cut-off can cause counter-intuitive results, if the per-period hazard in the comparator arm is below that of the hazard in the intervention arm"

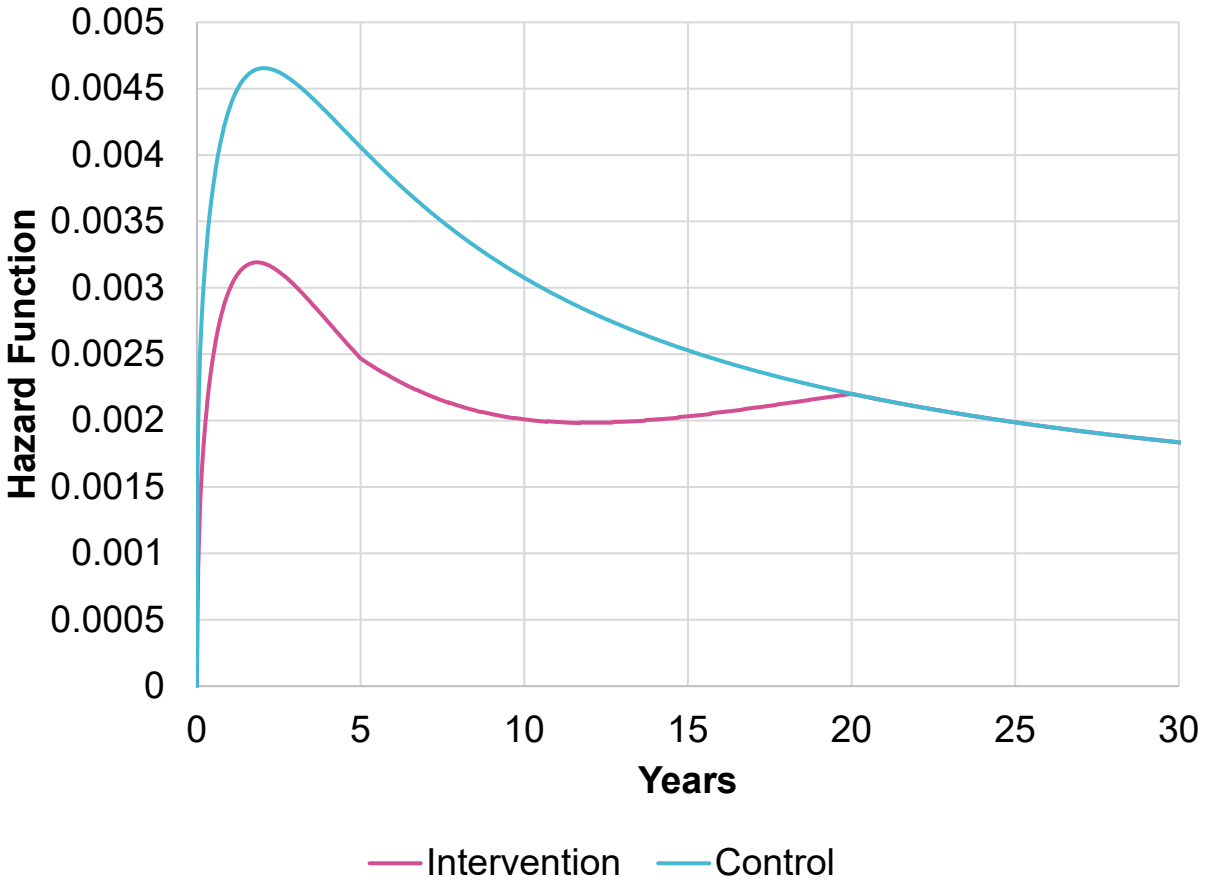
ERG report TA578

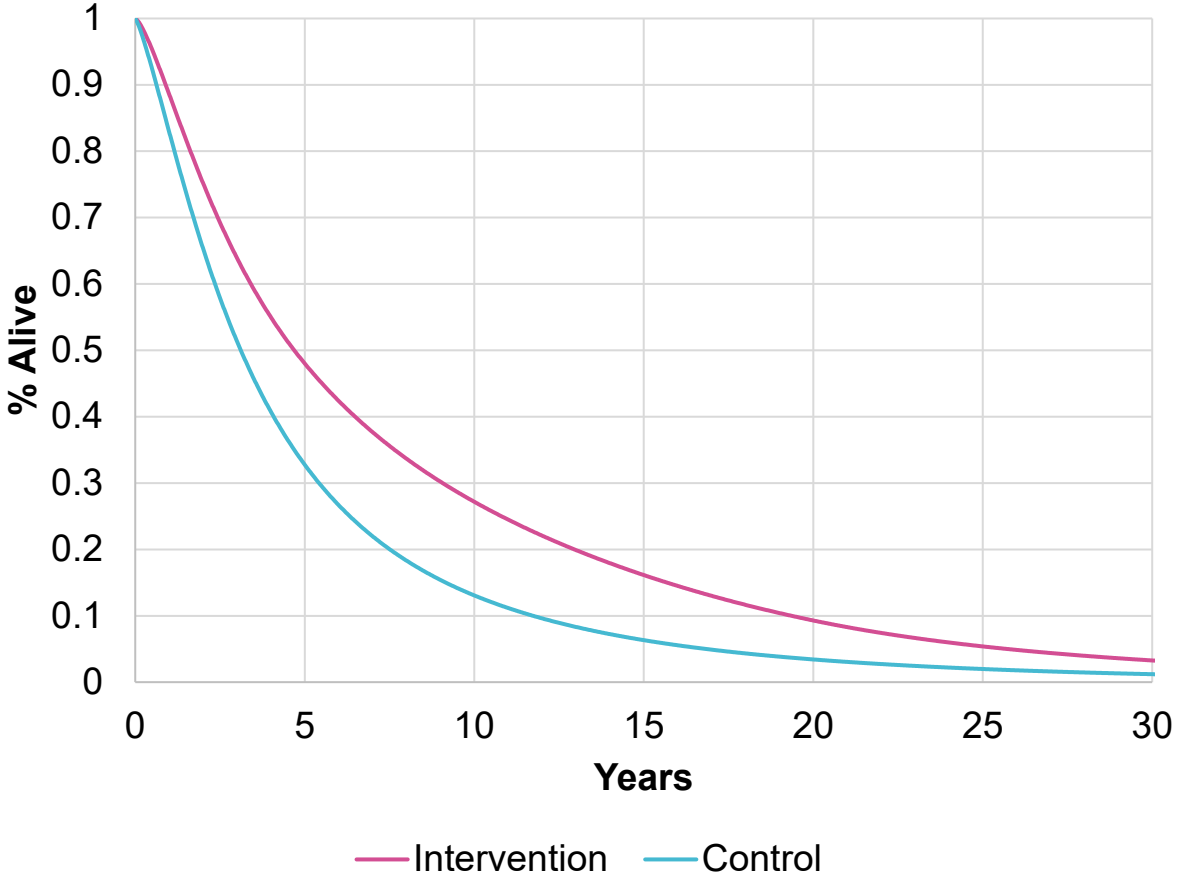
Hazards set equal at	ΔLY
5 years	1.3
10 years	1.5
15 years	1.4
Lifetime	1.1

Results can often be counterintuitive

Gradual waning: linear interpolation

ΔLY = 2.8

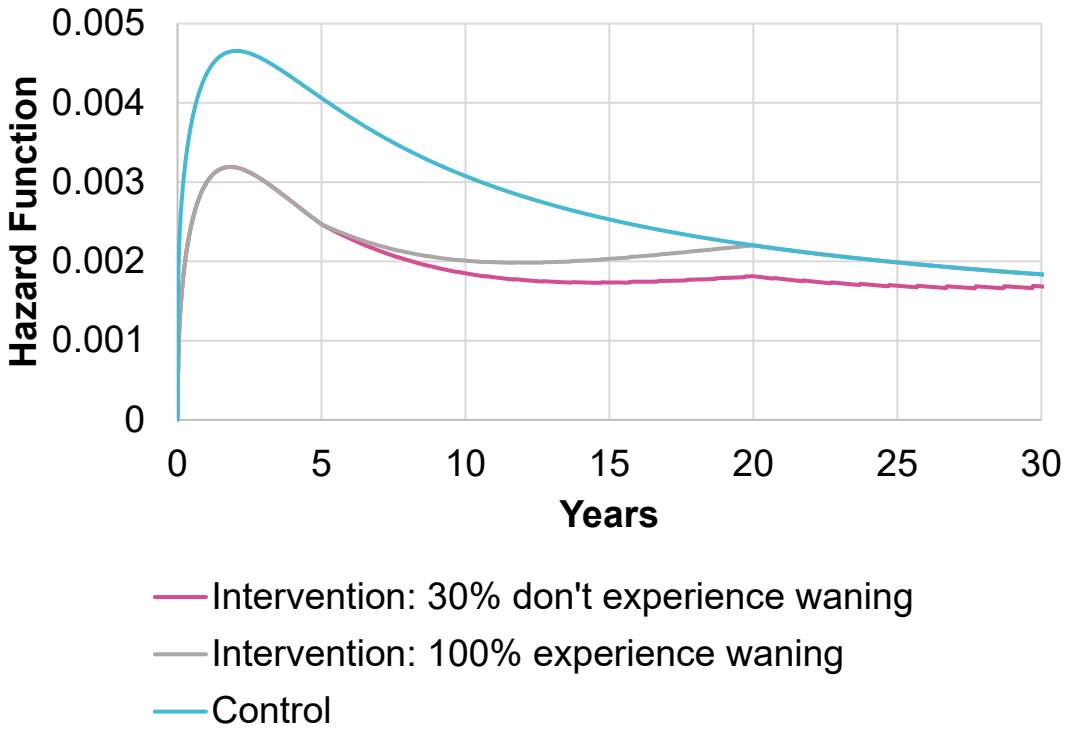




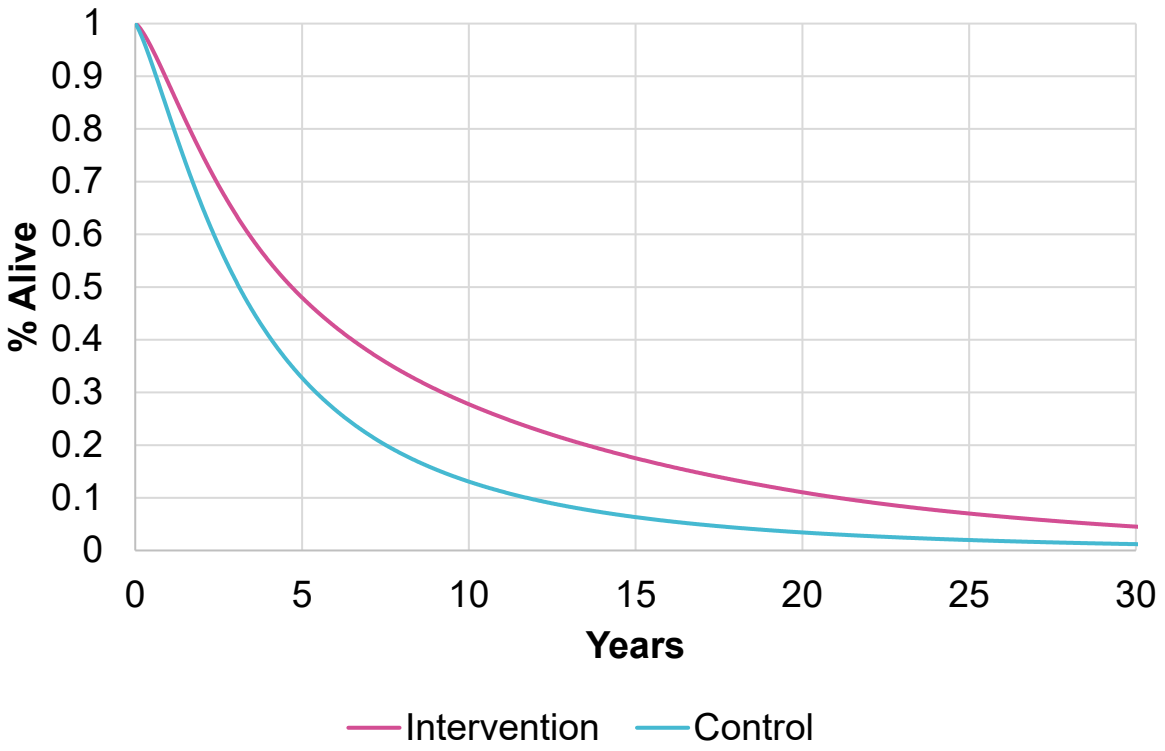
A bit more realistic but requires a good idea of times for start and end of decline in effect: still pretty arbitrary

What if some patients are expected to maintain their treatment effect?

For example complete responders



ΔLY = 3.3



More complex

Some alternatives: all of which rely on a good model for comparator survival

Link treatment effect on OS to treatment duration

- More complex structure needed
- Competing risks
- What happens when patients stop treatment?
  - Similar to comparator arm? From which timepoint (beginning of that curve or now)?
  - Worse (disease now resistant)?

Direct consideration of likely future relative treatment effects

- Expert elicitation
- Using external or earlier phase datasets

Implicitly capture through selection of more realistic hazard functions

Directly model impact of lines of treatment / progressions / response

Should we be incorporating the uncertainty probabilistically?

- Enables us to better quantify the impact of uncertainty on the ICER
- Correlations need to be accounted for
- Model averaging methods could be used + informed priors
- Combined with value of information techniques could lead to increased understanding of the value & usefulness of managed access

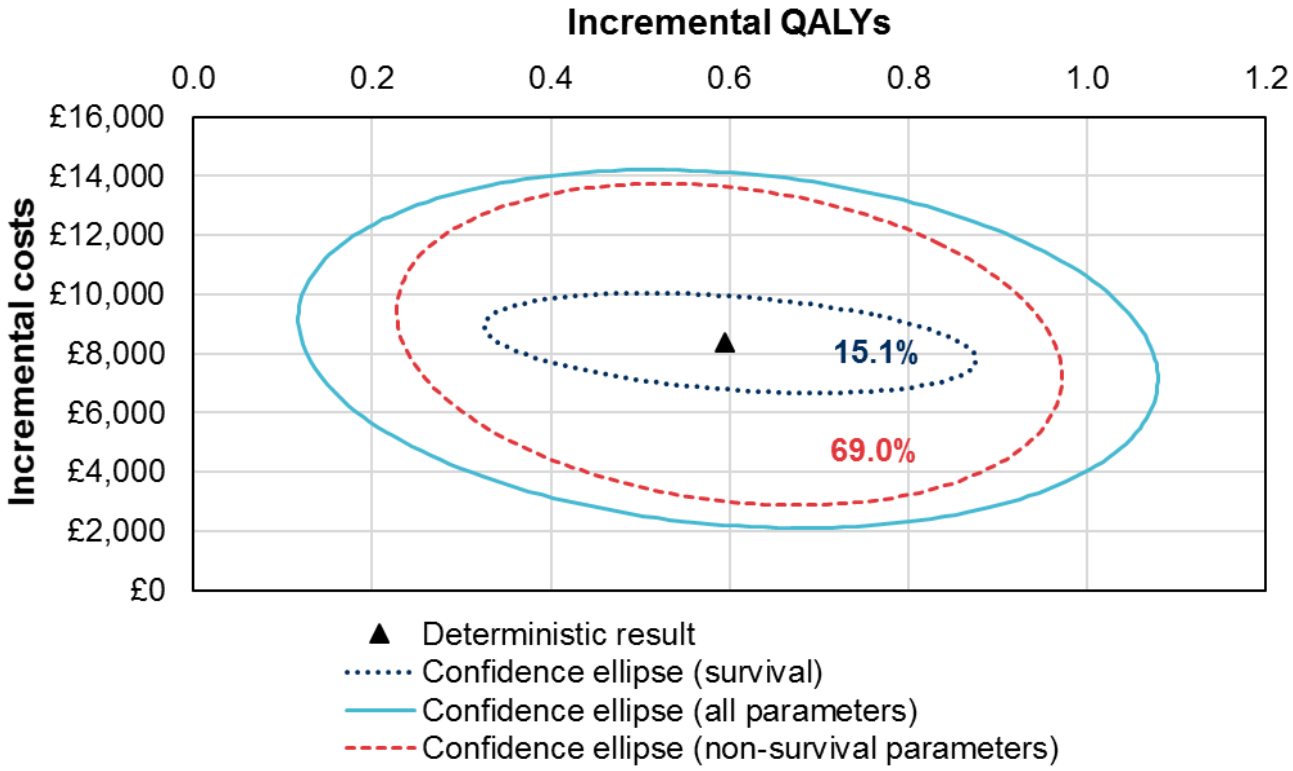


Figure from PRM119, An exploration of techniques for addressing uncertainty in survival estimates used within partitioned-survival models, ISPOR Europe 2016

Who wants to be a Bayesian? (and why you should...)

Gianluca Baio

Department of Statistical Science | University College London

g.baio@ucl.ac.uk
http://www.statistica.it/gianluca/
https://egon.stats.ucl.ac.uk/research/statistics-health-economics/
https://github.com/giabaio
https://github.com/StatisticsHealthEconomics
@gianlubaio

Virtual ISPOR US Conference, The Internet

20 May 2021

This presentation is available at www.statistica.it/gianluca/slides/ispor-2021

What are we talking about?...

... Well, there are **many** problems!

Data

1. We may (or may not!) access **individual level data** for "our" trial, but not for the competitors'
2. The trial data have a very limited follow up, which implies large amount of censoring
   - This is often OK(-ish!) for "medical stats" analysis. But **HORRIBLE** for economic evaluation! $\Rightarrow$ Extrapolation
3. Often the data are manipulated by the stats team within the sponsor and the economic modellers only get summaries/estimates
   - It is **ALWAYS** good to leave things to statisticians. But the modellers can (should?!) be statisticians too, so they could handle the data!...

Models

1. Which model is the "best fit" – how to judge that?
2. Is modelling even enough? (How to make the most of "external data")
3. Should you be Bayesians about this?
   - (Spoiler alert: the answer is *always* Yes!...)

Combining composite data sources

Increasingly popular

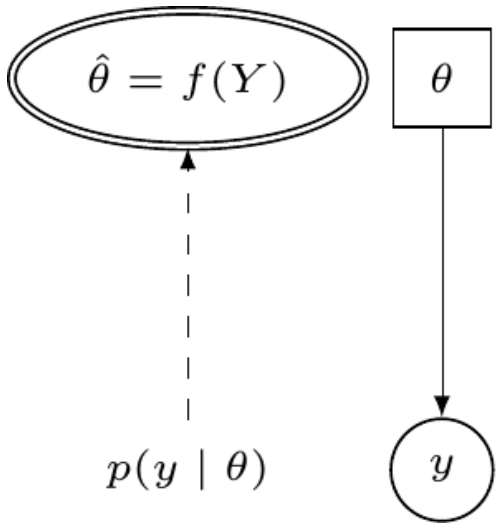
- 13 Technology Assessments (TAs) in immuno-oncology in the period 2019-2021
- 7 formally included external data, of various form
- Sources used to support treatment effect waning (or lack of it) included:
  - Other non-pivotal clinical trials and published sources with specific % of patients alive at a time point
  - Flatiron (or other registries such as SEER)
  - Clinical expert opinion ("soft" vs "hard" data... $\Rightarrow$ more on this later)
    - On % of patients surviving at a specific time t
    - On clinical implausibility of hazards crossing and becoming higher for intervention vs comparator

Challenges

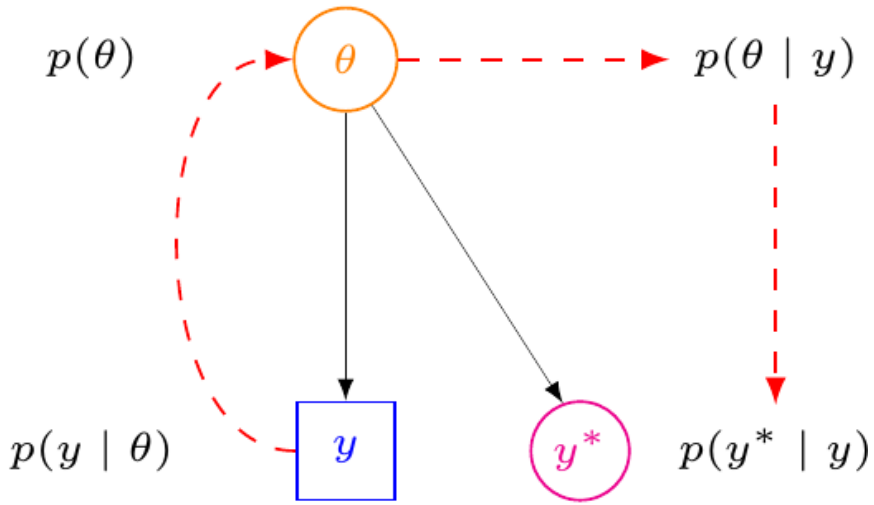
- Heterogeneity/representativeness
  - "Exchangeability"
- Afterthought vs plan ahead...
- KOL/Expert opinion/soft evidence: elicitation, formal modelling?...

To be or not to be (a Bayesian)?...

Frequentist ("standard")



Bayesian

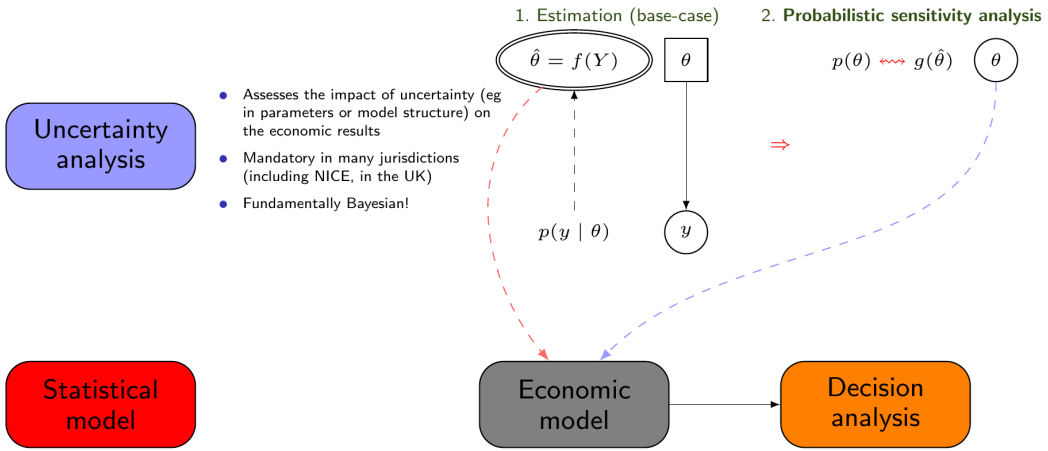


- A Bayesian only speaks one language: probability distributions to describe
  – Sampling variability (relevant for observ***ed*** data)
  – Epistemic uncertainty (relevant for ***un***observ***able*** parameters + yet ***un***observ***ed*** future data)
- Contextual (="prior") information to be formally included in the construction of the model
  – Almost irrelevant when evidence is "definitive" (large and consistent data)
  – Crucial when data are sparse! (... But this isn't preposterous, is it?...)

To be or not to be (a Bayesian)?...

In HTA

Frequentist ("standard")



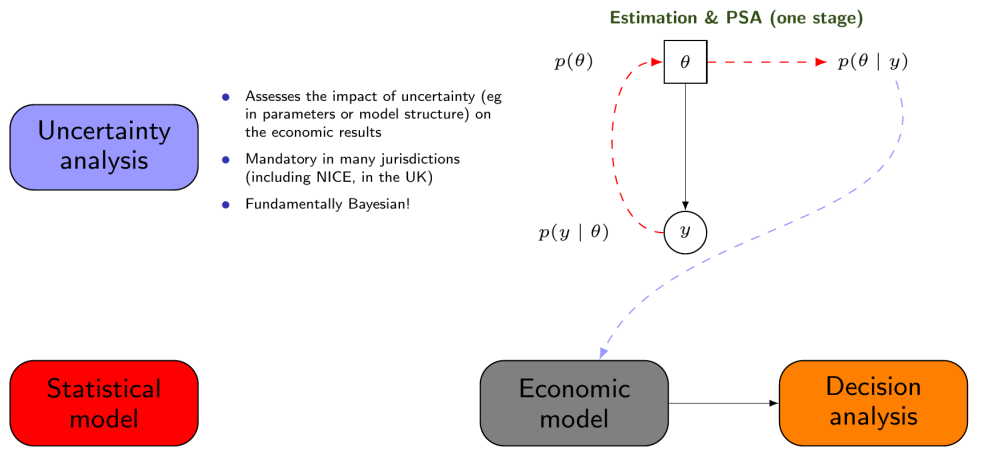
- Assesses the impact of uncertainty (eg in parameters or model structure) on the economic results
- Mandatory in many jurisdictions (including NICE, in the UK)
- Fundamentally Bayesian!

- Estimates relevant **population** parameters $\boldsymbol{\theta}$
- Varies with the type of available data (& statistical approach!)

- Combines the parameters to obtain a population average measure for costs and clinical benefits
- Varies with the type of available data & statistical model used

- Summarises the economic model by computing suitable measures of "cost-effectiveness"
- Dictates the best course of actions, given current evidence
- Standardised process

Bayesian



- Assesses the impact of uncertainty (eg in parameters or model structure) on the economic results
- Mandatory in many jurisdictions (including NICE, in the UK)
- Fundamentally Bayesian!

- Estimates relevant **population** parameters $\boldsymbol{\theta}$
- Varies with the type of available data (& statistical approach!)

- Combines the parameters to obtain a population average measure for costs and clinical benefits
- Varies with the type of available data & statistical model used

- Summarises the economic model by computing suitable measures of "cost-effectiveness"
- Dictates the best course of actions, given current evidence
- Standardised process

Survival analysis in HTA

General structure

$$t \sim f(\mu(\boldsymbol{x}), \alpha(\boldsymbol{x})), \qquad t \geq 0$$

- $\boldsymbol{x} =$ vector of covariates (potentially influencing survival)
- $\mu(\boldsymbol{x}) =$ **location** parameter
  - Scale or mean – usually main objective of the (biostats!) analysis
  - Typically depends on the covariates $\boldsymbol{x}$
- $\alpha(\boldsymbol{x}) =$ **ancillary** parameters
  - Shape, variances, etc
  - May depend on $\boldsymbol{x}$, but often assume they don't (see NICE TSD 14)
- **NB**: $S(t)$ and $h(t)$ are functions of $\mu(\boldsymbol{x}), \alpha(\boldsymbol{x})$
- Typically use generalised linear model

$$g(\mu_i) = \beta_0 + \sum_{j=1}^{J} \beta_j x_{ij} [+ \ldots]$$

  - since $t > 0$, usually, $g(\cdot) = \log$
- In a Bayesian setting, complete by putting suitable priors on $\boldsymbol{\beta}$ and α

Bayesian Survival analysis in HTA

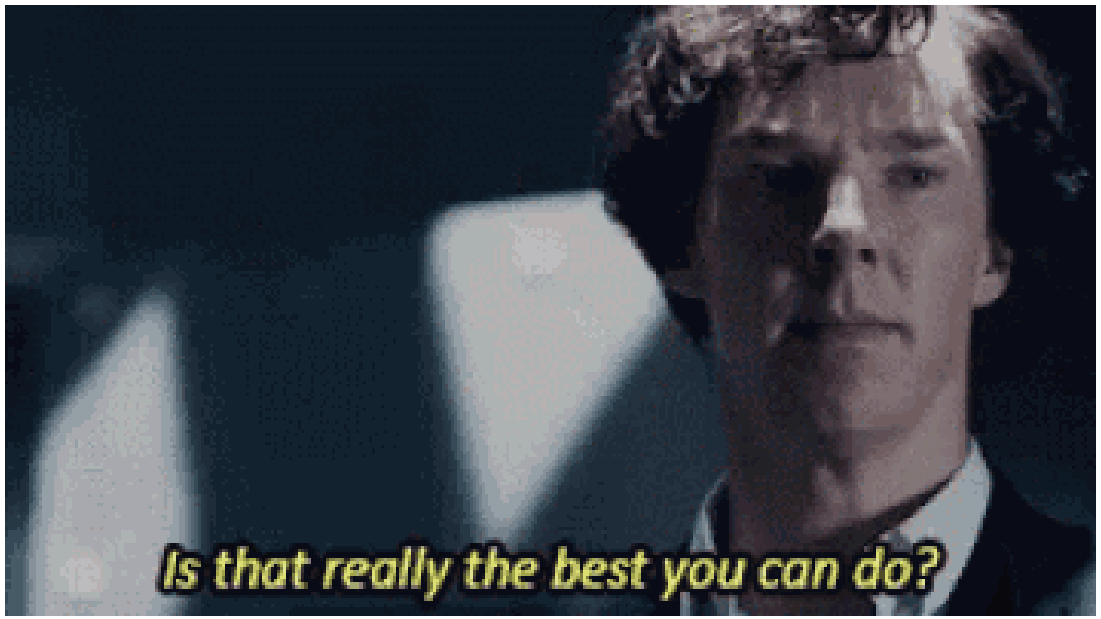
Data model	Location parameter	Ancillary parameter
$t_i \sim$ Exponential(μ_i)	Rate: $\mu_i = \exp\left(\beta_0 + \sum_{j=1}^{J} \beta_j x_{ij}\right)$	–
$t_i \sim$ Weibull(μ_i, α)	Scale: $\mu_i = \exp\left(\beta_0 + \sum_{j=1}^{J} \beta_j x_{ij}\right)$	Shape: $\alpha \sim$ Gamma$(0.1, 0.1)$
$t_i \sim$ logNormal(μ_i, α)	log-mean: $\mu_i = \beta_0 + \sum_{j=1}^{J} \beta_j x_{ij}$	log-sd: $\alpha \sim$ Uniform$(0, 5)$
$t_i \sim$ logLogistic(μ_i, α)	Rate: $\mu_i = \exp\left(\beta_0 + \sum_{j=1}^{J} \beta_j x_{ij}\right)$	Shape: $\alpha \sim$ Gamma$(0.1, 0.1)$
$t_i \sim$ Gamma(μ_i, α)	Rate: $\mu_i = \exp\left(\beta_0 + \sum_{j=1}^{J} \beta_j x_{ij}\right)$	Shape: $\alpha \sim$ Gamma$(0.1, 0.1)$
$t_i \sim$ Gompertz(μ_i, α)	Rate: $\mu_i = \exp\left(\beta_0 + \sum_{j=1}^{J} \beta_j x_{ij}\right)$	Shape: $\alpha \sim$ Gamma$(0.1, 0.1)$
$t_i \sim$ Gen Gamma$(\mu_i, \boldsymbol{\alpha})$	Location: $\mu_i = \beta_0 + \sum_{j=1}^{J} \beta_j x_{ij}$	$\boldsymbol{\alpha} = (\sigma, q)$ Scale: $\sigma \sim$ Gamma$(0.1, 0.1)$ Shape: $q \sim$ Normal$(0, 100)$
$t_i \sim$ Gen F$(\mu_i, \boldsymbol{\alpha})$	Location: $\mu_i = \beta_0 + \sum_{j=1}^{J} \beta_j x_{ij}$	$\boldsymbol{\alpha} = (\sigma, q, p)$ Scale: $\sigma \sim$ Gamma$(0.1, 0.1)$ Shape(1): $\log(p) \sim$ Normal$(0, 0.5)$ Shape(2): $q \sim$ Normal$(0, 2.5)$

with $\boldsymbol{\beta} = \beta_0, \ldots, \beta_J \overset{iid}{\sim}$ Normal$(0, v)$, for suitable v

See here

Bayesian Survival analysis in HTA

- We can specify "minimally informative" priors (eg like `survHE` does by default)
  – In many ways, that's the "lazy" option...
- Similarly, we can try the various models suggested in the guidelines and see what happens...



- We probably *know* something more about the likely shape of the hazard function
  – Likely to be monotonically increasing?
  – Definitely unlikely to be constant over time?...
- These considerations should drive the choice of models **over and above** testing all the options!
- What else do we know?
  – Likely average survival time
  – Chances of surviving after t^* units of time (eg >75 years old)
  – Population data to "anchor" the extrapolated survival curves
  – . . .

Example: ICD & Cardiac death

Basic idea/modelling

Use UK population data (matched by age/sex) to "**anchor**" the ICD population at risk — Benaglia et al (2015)

- Perhaps the easiest way to do this is to relate the hazard between the two populations – eg **proportional hazard** (PH) model

$$h_{\text{ICD}}(t) = e^{\beta} h_{\text{UK}}(t) \quad \Leftrightarrow \quad \text{HR} = \frac{h_{\text{ICD}}(t)}{h_{\text{UK}}(t)} = e^{\beta} = \text{Constant}$$

- Relatively easy to model – but probably very unrealistic!
  - ICD patients are at (much?) greater risk of arrhythmia death
  - If the proportion of deaths caused by arrythmia changes over time, we would induce bias, because we would be extrapolate a constant HR for all causes mortality
- Formally account for multiple mortality causes (**Poly-Weibull** model Demiris et al, 2015):

$$\begin{aligned} h_{\text{ICD}}(t) &= h^{\text{arr}}_{\text{ICD}}(t) + h^{\text{oth}}_{\text{ICD}}(t) \\ &= e^{\beta} h^{\text{arr}}_{\text{UK}}(t) + h^{\text{oth}}_{\text{UK}}(t) \\ &= e^{\beta} \alpha_1 \mu_1 t^{\alpha_1 - 1} + \alpha_2 \mu_2 t^{\alpha_2 - 1} \end{aligned}$$

- This assumes that
  - Arrhythmia hazard is **proportional** to matched UK population
  - Other causes hazard is **identical** to matched UK population

Example: ICD & Cardiac death

Turning prior *information* into a prior *distribution*

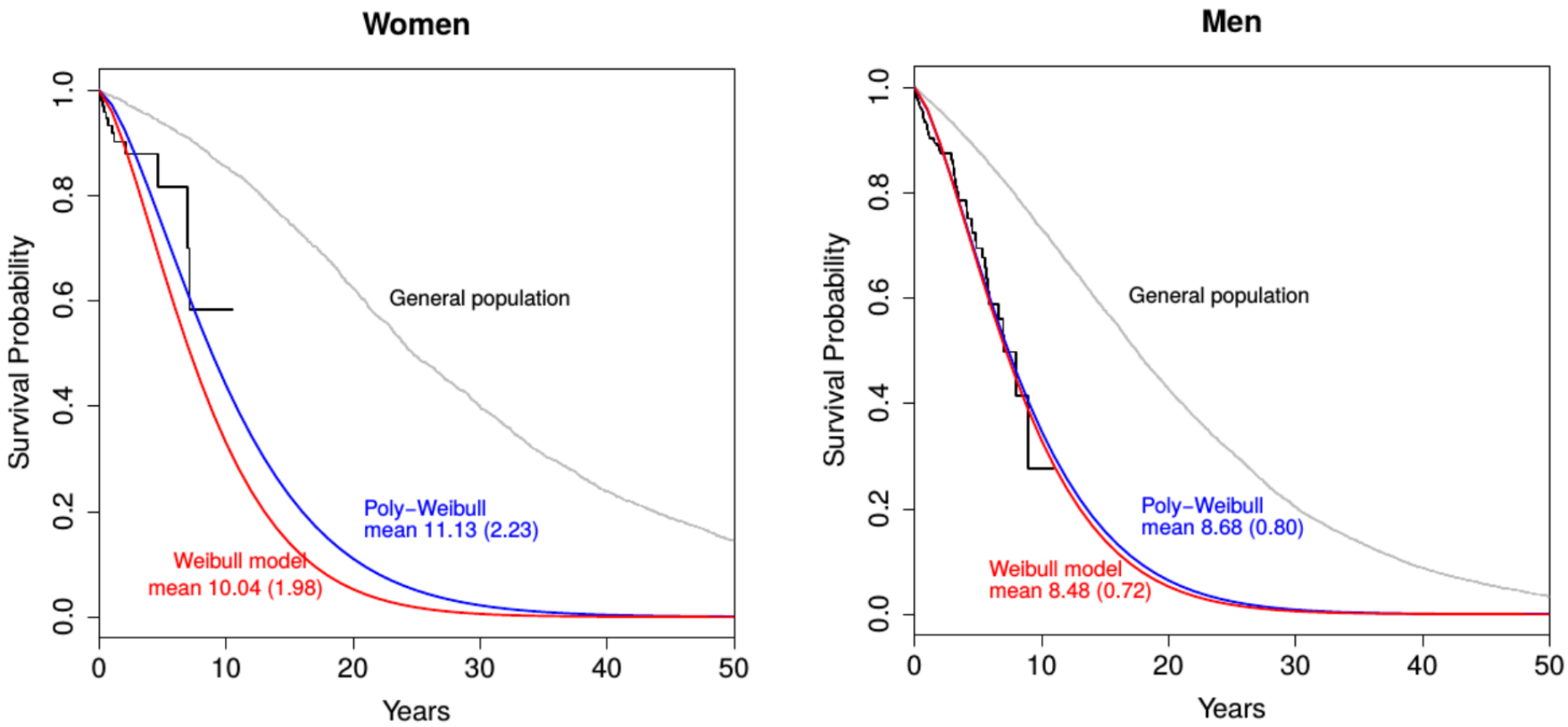
- In the ICD case, age at entry is around 60 – we **know** that people won't survive more than 60 more years
  - Setting a prior for the scale $\mu_i \sim \text{Uniform}(0, 100)$ implies that the prior mean survival of the resulting Weibull distribution is

$$\mu_i \Gamma\left(1 + \frac{1}{\alpha}\right) < 60$$

- Can also include some knowledge on the shape α and the coefficient β to limit their variations in reasonable ranges...

- This isn't necessarily easy!
  - You need to be friends with a statistician...
- Don't be lost in translation...
  - *Elicit* the actual **information** and then map it onto a possible and reasonable **distribution**
  - Mapping changes with the mathematical properties of the underlying sampling distribution selected...

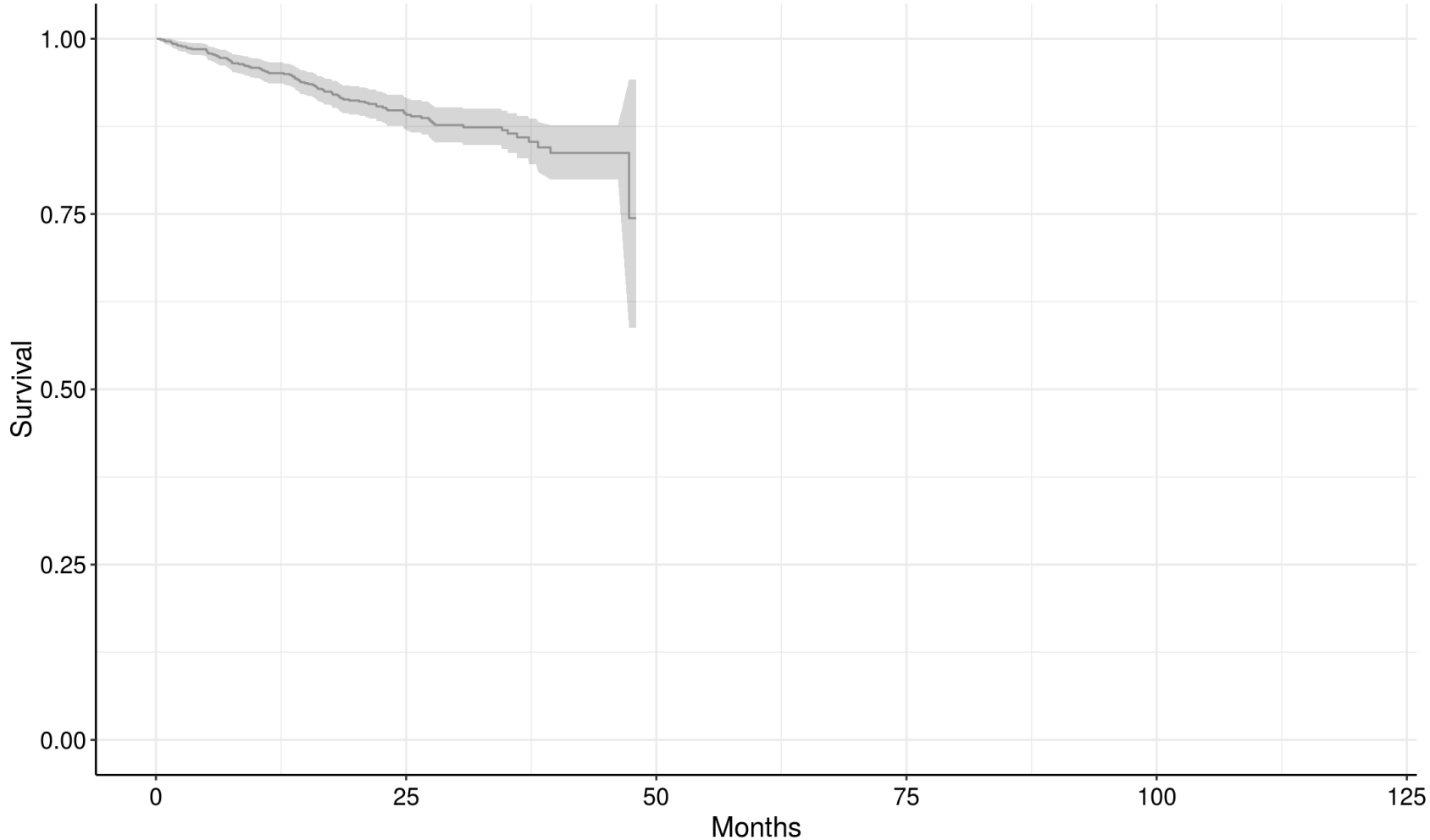
Example: ICD & Cardiac death



- Ignoring cause-specific mortality (simple Weibull model) results in larger bias, especially for females, mostly because the arrhythmia proportion of deaths does vary over time in that subgroup

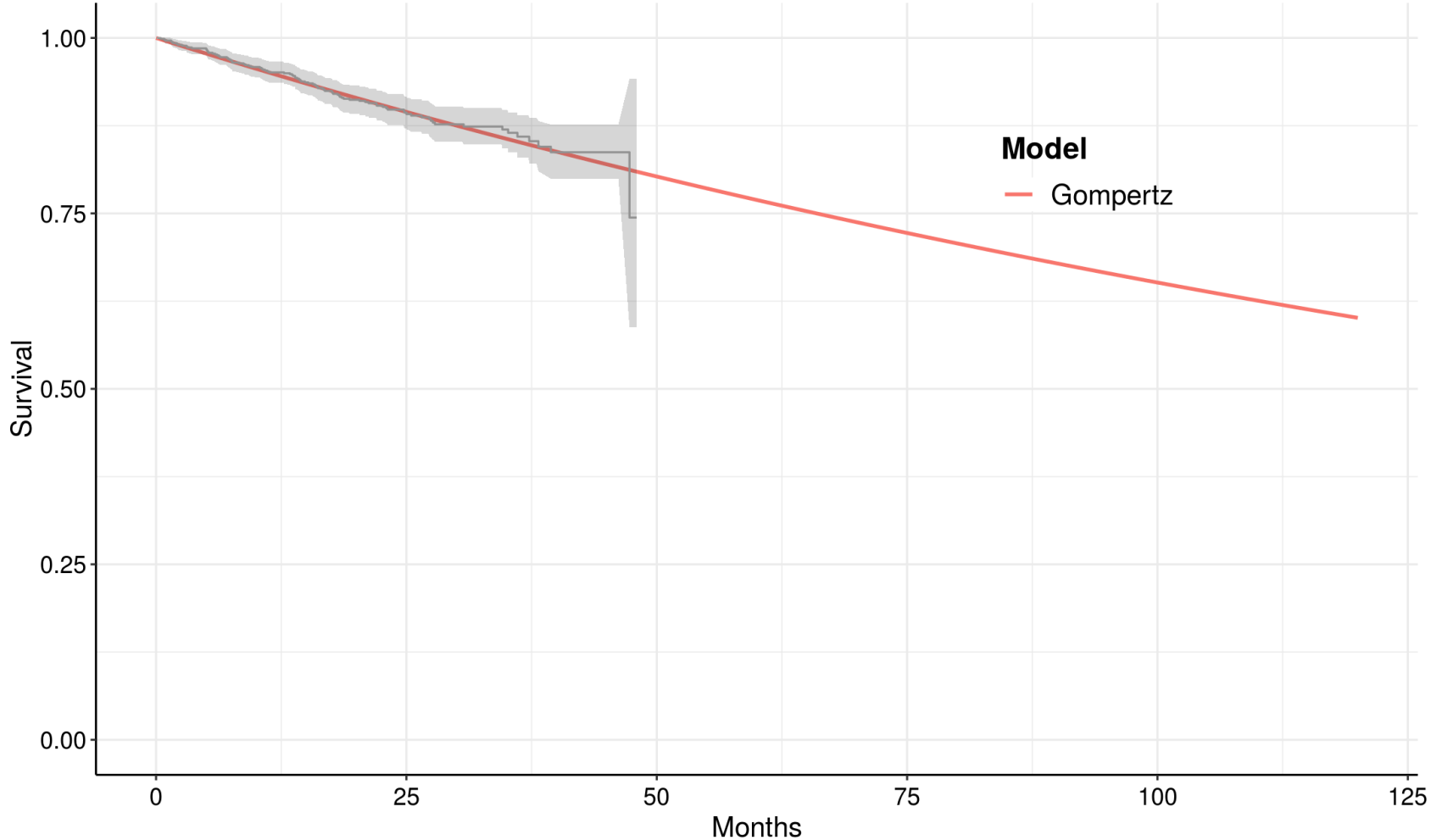
Example: constraints on $S(t)$

Observed data



Example: constraints on $S(t)$

Parametric extrapolation



Example: constraints on $S(t)$

What do we see?

- The data are **sparse** and the follow up is limited in comparison to the relevant time horizon
- The **best fitting** model responds by extrapolating a survival curves that implies $\Pr(\text{Still alive after 100 months}) > 0.5$
- This is most likely a ridiculous finding!

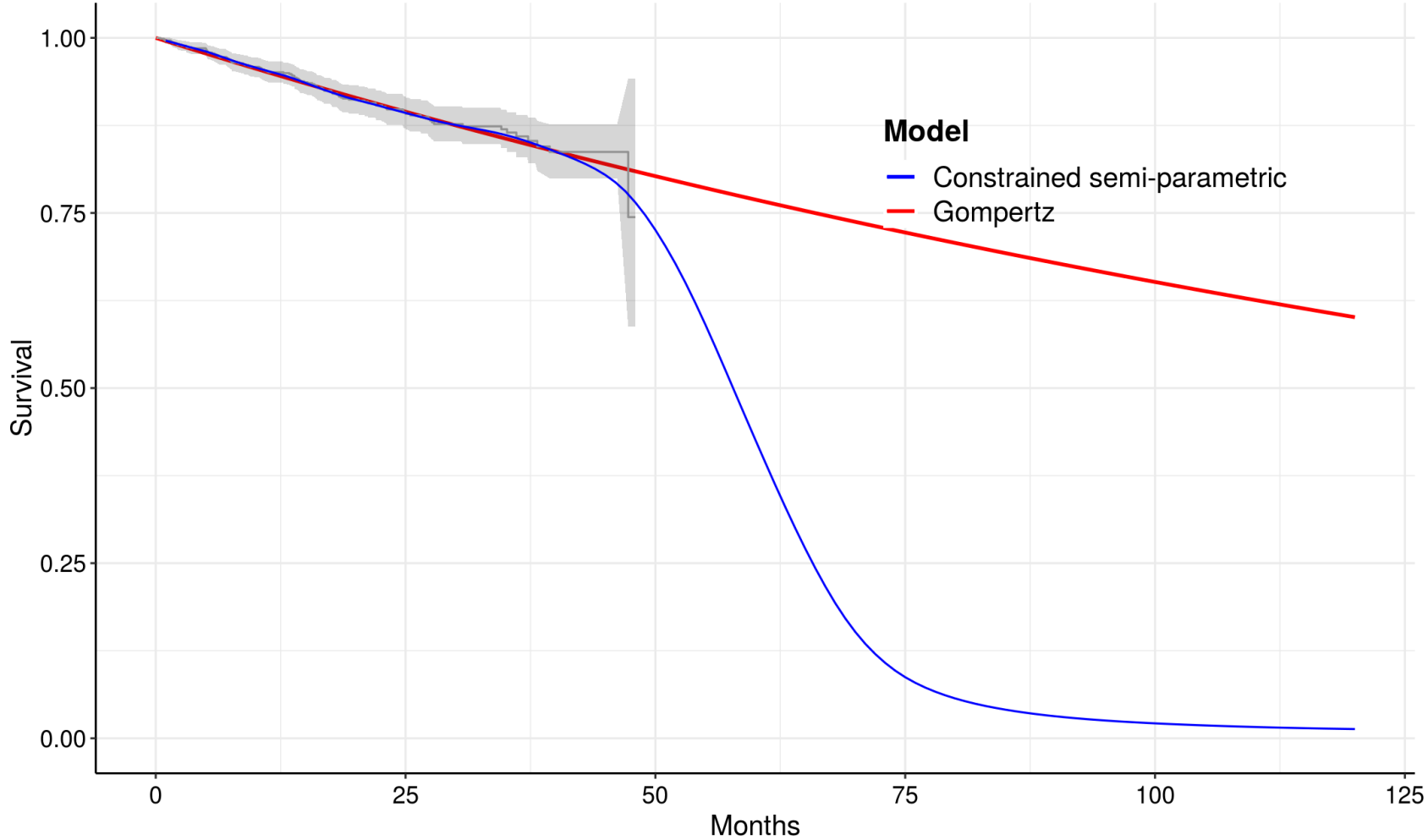
What do we know?

- Perhaps we may think a bit more carefully and figure out some kind of "constraint" or upper limit for the survival probability at a given time point in the future...
- Maybe, it's not so controversial to assume that, **before observing any data**, $\Pr(\text{Still alive after 70 months})$ should not exceed, say, 0.20
- We can use this information in our prior specification and let it be modified by the observed data
  – This is a relatively strong prior, so you would need a **really** strong signal to modify it significantly...

(*Che et al*, work in progress...)

Example: constraints on $S(t)$

Constrained semi-parametric



Mixture "cure" models (MCMs)

- Sustained treatment effect
  - Effectively, a fraction of individuals are subject to a different "data generating process"
  - The overall survival curve is a combination of two components $\Rightarrow$**mixture** model

$$S(t, \boldsymbol{x}) = S_b(t, \boldsymbol{x})[\pi(\boldsymbol{x}) + (1 - \pi(\boldsymbol{x}))\, S_c(t, \boldsymbol{x})],$$

  - $\boldsymbol{x} =$ a vector of individual level covariates
  - $S_b(t, \boldsymbol{x}) =$ (complement of) background mortality for the population with $\boldsymbol{x}$ profile
  - $S_c(t, \boldsymbol{x}) =$ (complement of) cancer-specific mortality for the population with $\boldsymbol{x}$ profile
  - $\pi(\boldsymbol{x}) =$ "cure fraction" = proportion of individuals with $\boldsymbol{x}$ profile who are "cured"
- Basically, this means that the overall survival in the population with $\boldsymbol{x}$ profile
  - Is the same as the "healthy" population for the proportion who are "cured"
  - It has an extra multiplicative, independent risk associated with the event (cancer) for those who aren't
- We may have some biological/pharmacological insight as to the potential for this phenomenon
  - Possibly plausible with immuno-oncological drugs
- **BUT**: most likely, we'll have to base our judgement on a limited follow up
  - This may suggest a plateau in one treatment arm, which is likely based on very uncertain evidence

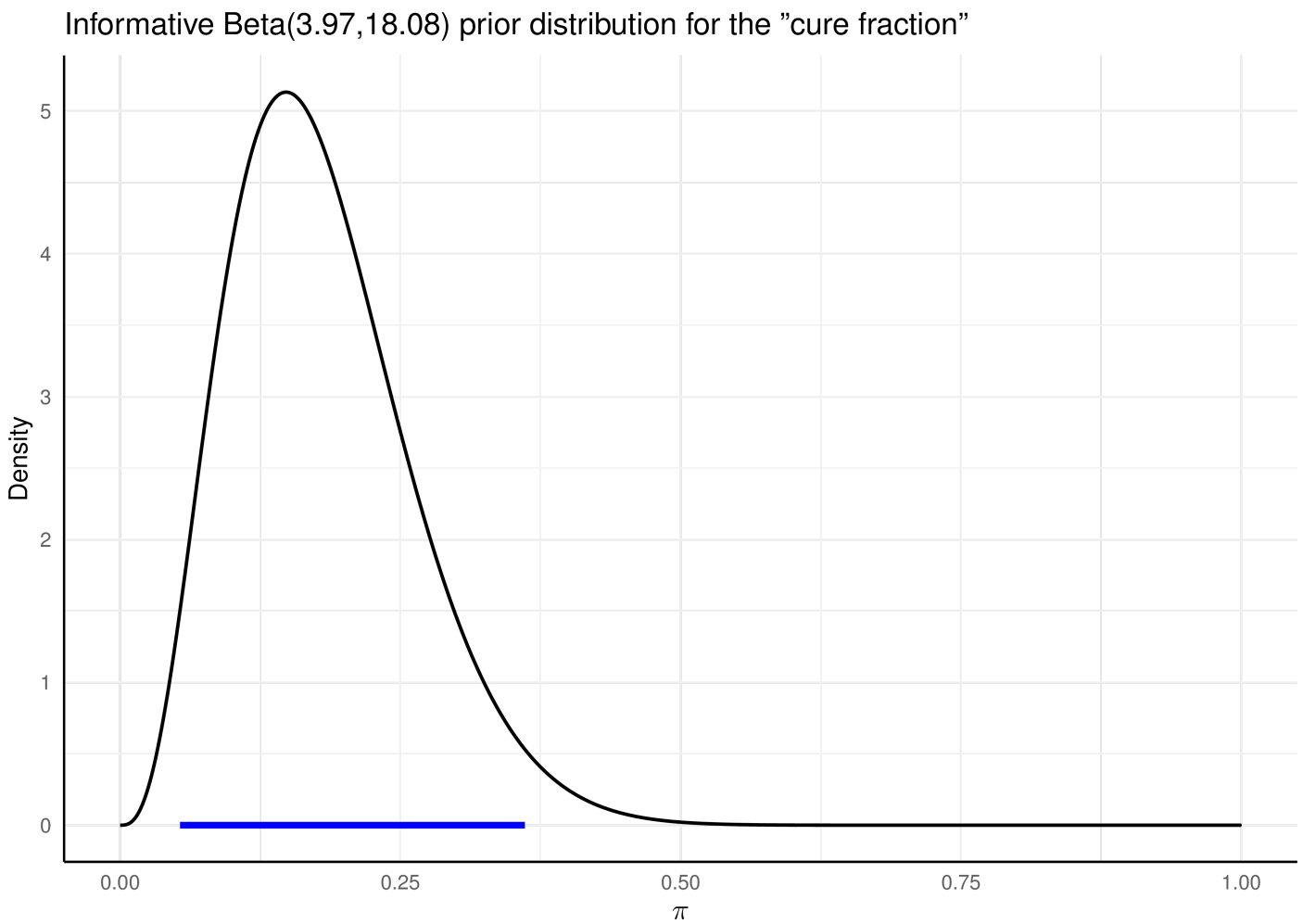
Bayesian MCMs

- "Standard" MCMs basically assume no prior knowledge about the "cure fraction" π
  - This means that the data are taken at face value
  - Weak evidence of flattening of the survival curve may lead to unrealistically large estimates for the the cure fraction
- **BUT**: assuming no prior information on π basically implies that we (implicitly) believe that it can be very large
  - Implausible in most cases
  - Can look at other, more established treatments
- If you're Bayesian about this, you may use a "regularising" prior to avoid "parachute effect"...

Typically, we can restrict the likely size of some effect, in **real** applications

- If you take 100 people on an airplane, randomise them to either get a parachute or not and measure whether they're still alive after jumping off the plane
  - Almost everybody with the parachute will be OK, almost everybody without will not!
  - You can reasonably expect a large "treatment effect" with OR $\approx$ 100, or so...
- With pharmaceutical interventions, this is not so common
  - Use *skeptical* priors (+ sensitivity analysis!) to restrict the likely range of effects
  - If the data are overwhelming pointing towards a large effect, the model will be able to pick that up
  - **BUT** you want to really see a signal before you call one...

Bayesian MCMs



- We may think that the cure fraction is unlikely to exceed 30%
- And that a reasonable expectation would be 20% "success"
- And we may want to allow for π to be actually as low as close to no "cured" at all

- We can *encode* this **information** using a Beta (3.97,18.09) distribution
  – The parameters $\alpha = 3.97$ and $\beta = 18.09$ can be "guessed" by trial and error **OR** working out some algebra
- **Before seeing any data**, this means that you're expecting
  – mean cure fraction = 0.22
  – 95% range = [0.054; 0.36]
- The observed data will modify this – but unless you have a **very** strong signal in the data, you won't go all crazy suggesting a *definite* plateau...

Conclusions

Too much, too soon?

- Tension between early introduction in the market and reimbursment decisions on the back of promising, but extremely immature data
  – Early plateau that doesn't materialise in later data cuts
- Divorce between "medical" and "economic" analysis
  – Lancet papers are OK with estimating median survival time and HRs... Economic evaluations need extrapolation to estimate mean survival time

All the help you can get

- Long-term data are ideal – if they're aligned with the population of interest and heterogeneity is manageable (and managed!)
- Often, even defining a comparator is a very complex operation and the market landscape is tricky...
  – Registry data can produce information "in real time". **But**: at the price of confounding/need for confirmation periods (conditional registration/reimbursment?)

Know what you know

- Some information ***is*** controversial and subjective and could bias the assessment. **But**: other simply isn't and we shouldn't be afraid to use it!

Science, art or arbitrariness? Evaluating the risk of treatment effect waning for novel oncology therapies

A Decision-Making Perspective

Nick Latimer, University of Sheffield, Sheffield, UK, Reader in Health Economics,
Yorkshire Cancer Research Senior Fellow

Disclosures

I am a member of NICE Appraisal Committee B

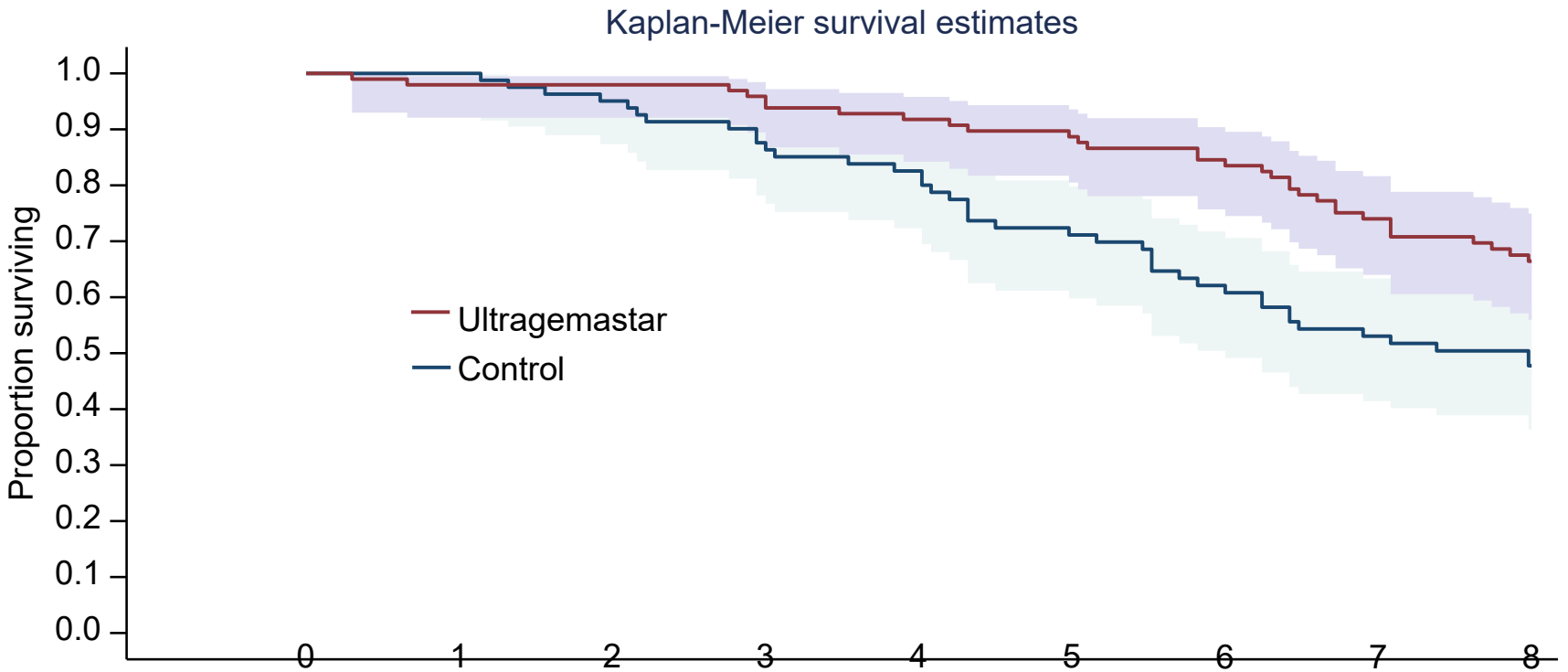
We're going to do a mock appraisal!

But all data are simulated, this example is not real! Supergemastar is fictitious. All characters are products of my imagination, and any resemblance to actual persons, or actual events, is purely coincidental

Imagine

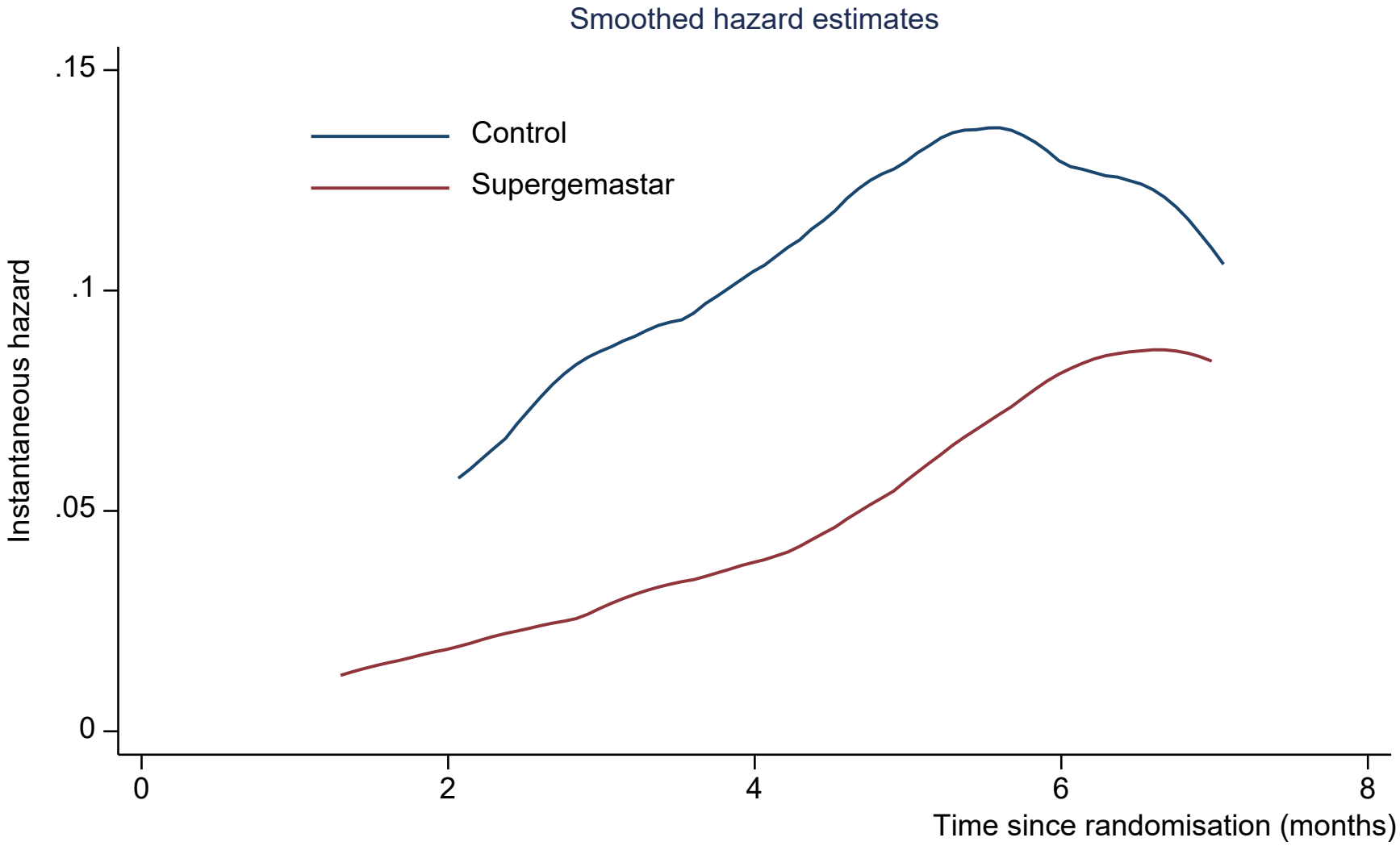
- You're a member of a NICE Appraisal Committee
- You're appraising a new cancer drug (called Supergemastar)
- You're presented with the data from the pivotal trial, and then with analyses conducted by the manufacturer and the Evidence Review Group (ERG)…

Mock Appraisal: Here are the survival curves…

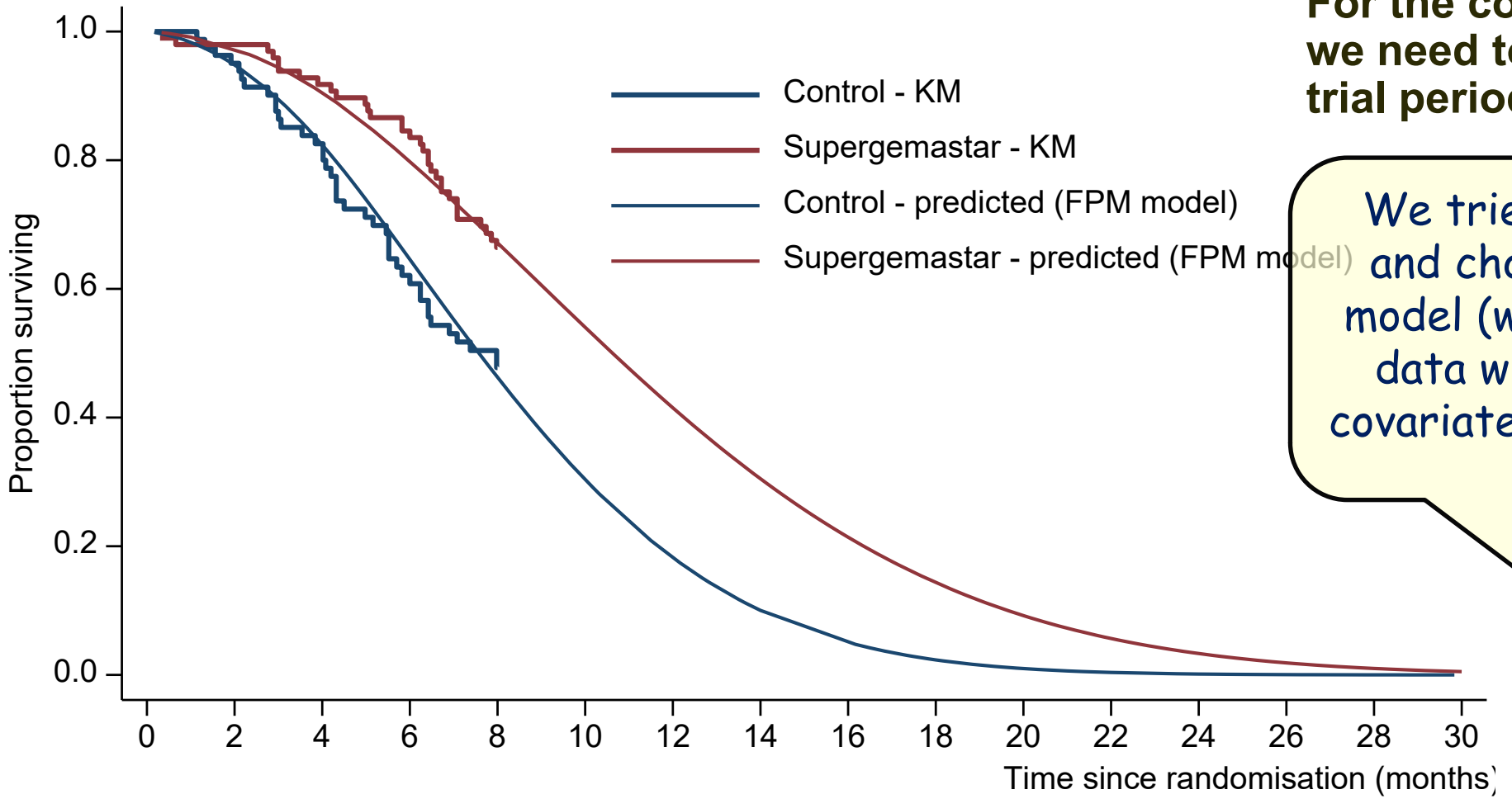


Control									
At-risk	82	81	77	70	65	56	48	41	36
Censored	0	1	1	2	3	3	4	4	5
Died	0	0	4	10	14	23	30	37	41
Ultragemastar									
At-risk	98	96	95	93	89	86	81	69	61
Censored	0	0	1	1	1	1	2	4	5
Died	0	2	2	4	8	11	15	25	32

Mock Appraisal: Here are the hazard plots…



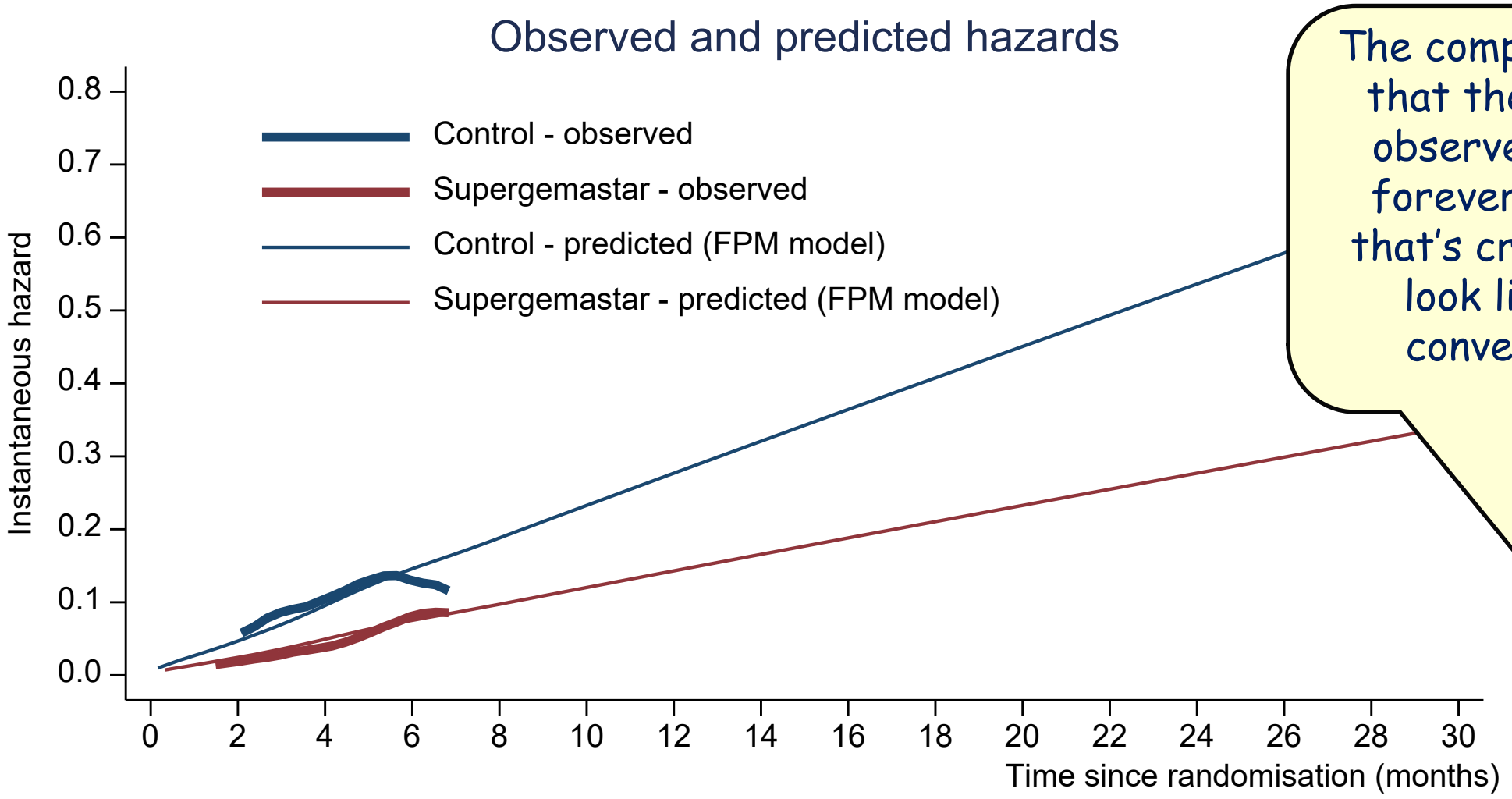
Mock Appraisal: The manufacturer's analysis…



For the cost-effectiveness analysis we need to extrapolate beyond the trial period

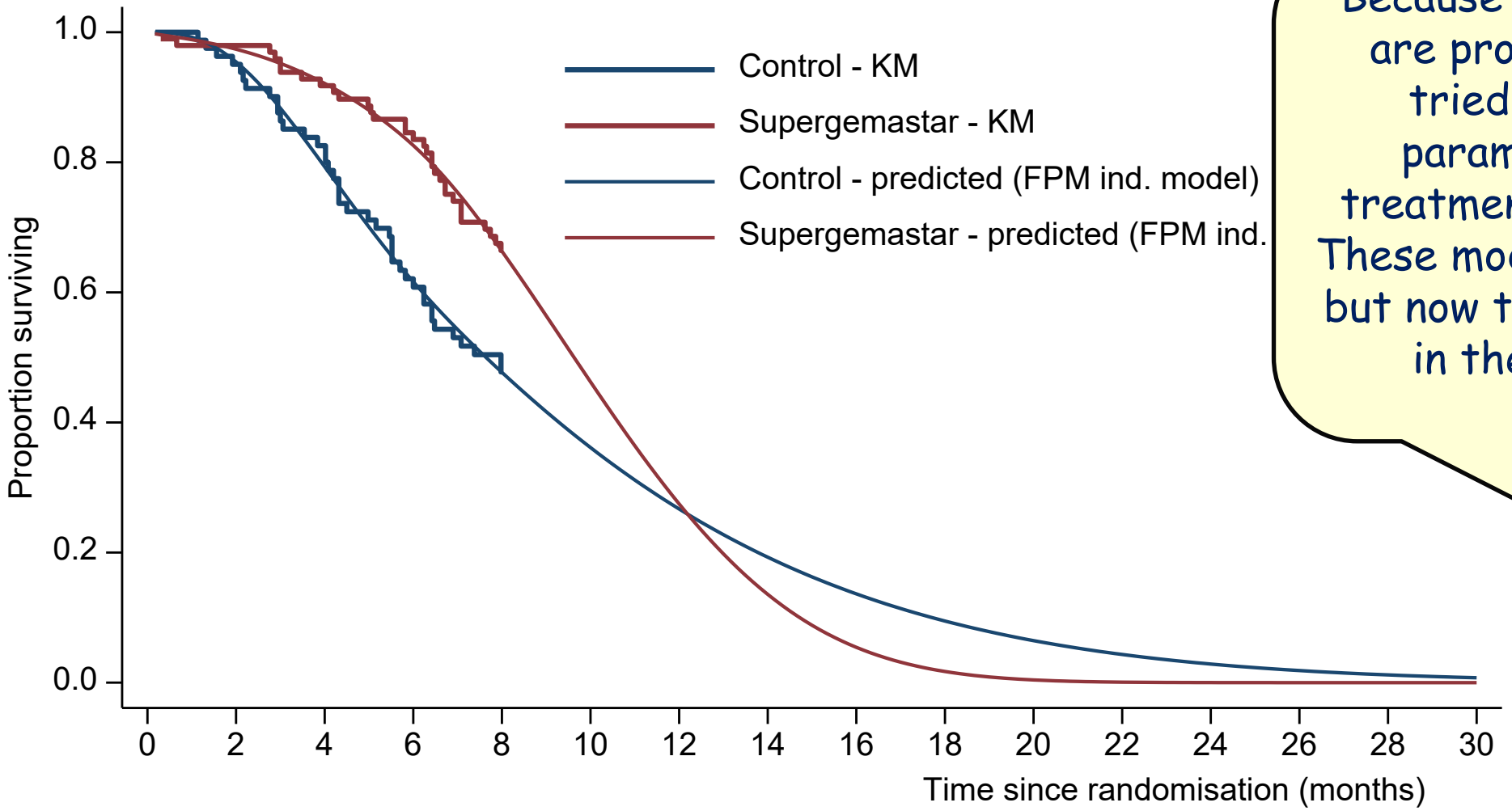
We tried several survival curves and chose a flexible parametric model (with 3 knots), fit to all the data with treatment group as a covariate. It looked like a great fit!

Mock Appraisal: The ERG comments…



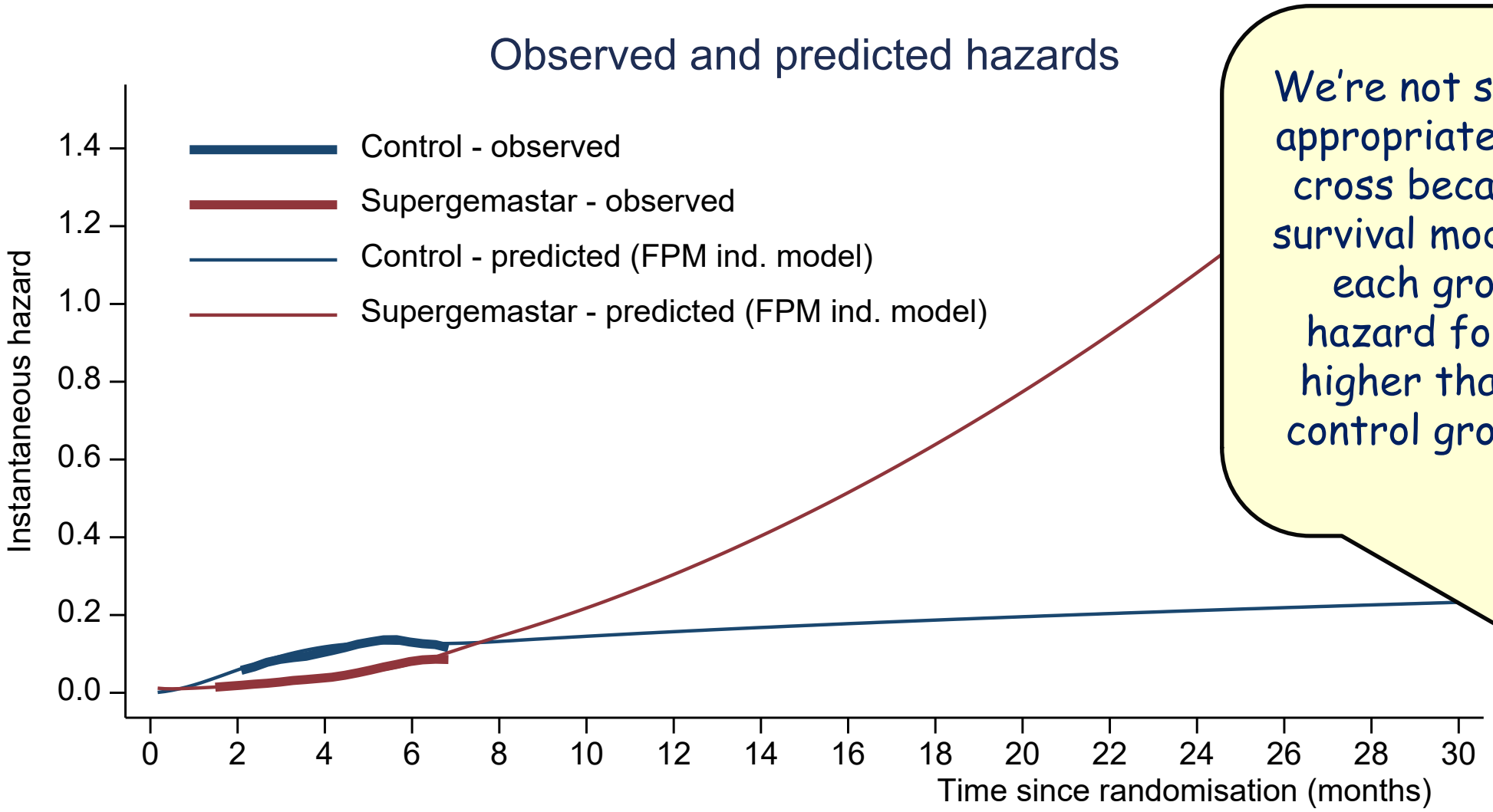
The company's model assumes that the treatment effect observed in the trial lasts forever. We're not sure if that's credible. The hazards look like they might be converging in the trial.

Mock Appraisal: The ERG tries a different analysis…



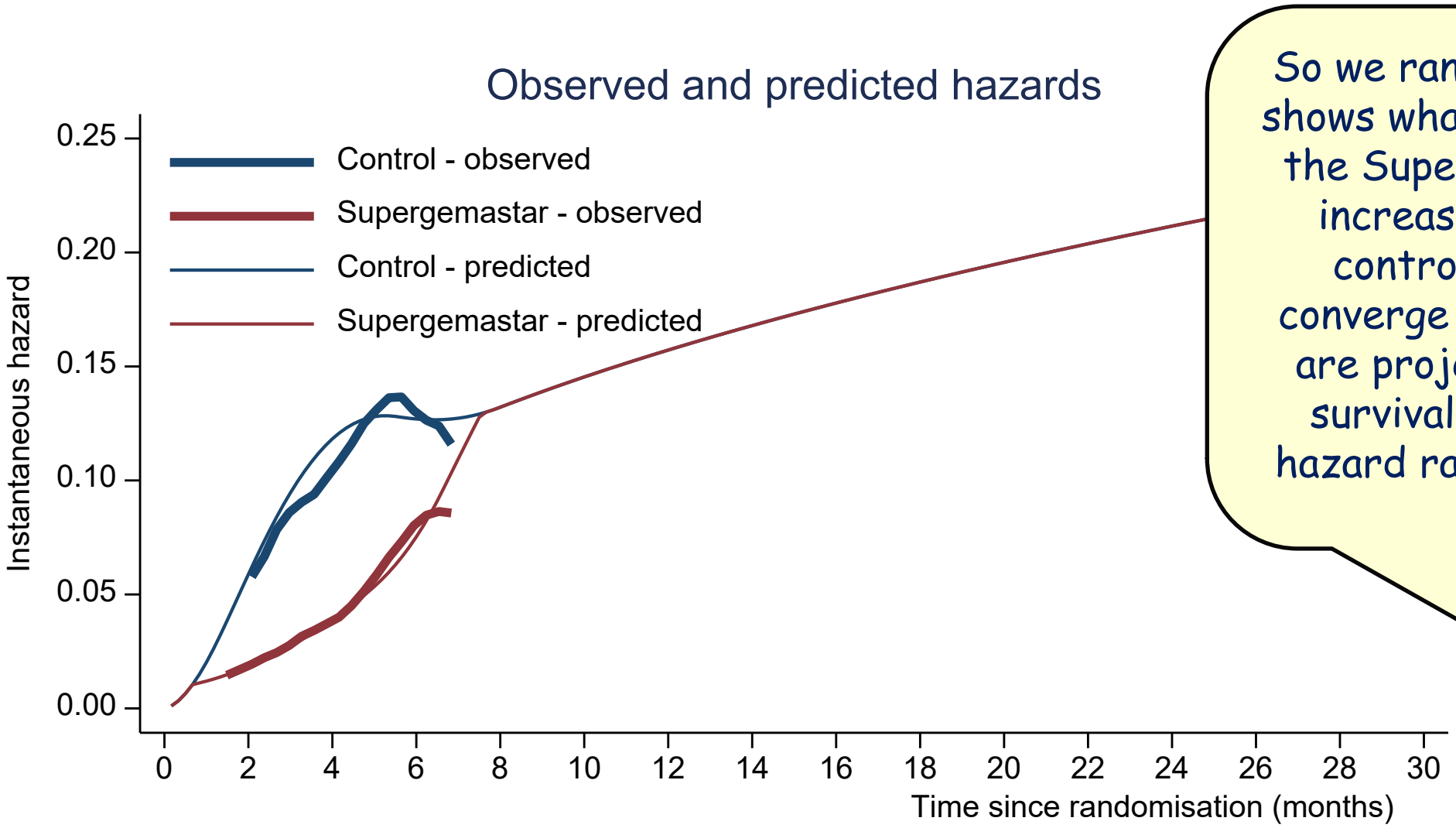
Because we're not sure if there are proportional hazards, we tried fitting the flexible parametric model to each treatment group independently. These models fit the data better, but now the survival curves cross in the extrapolated part

Mock Appraisal: But is the ERG analysis credible?...



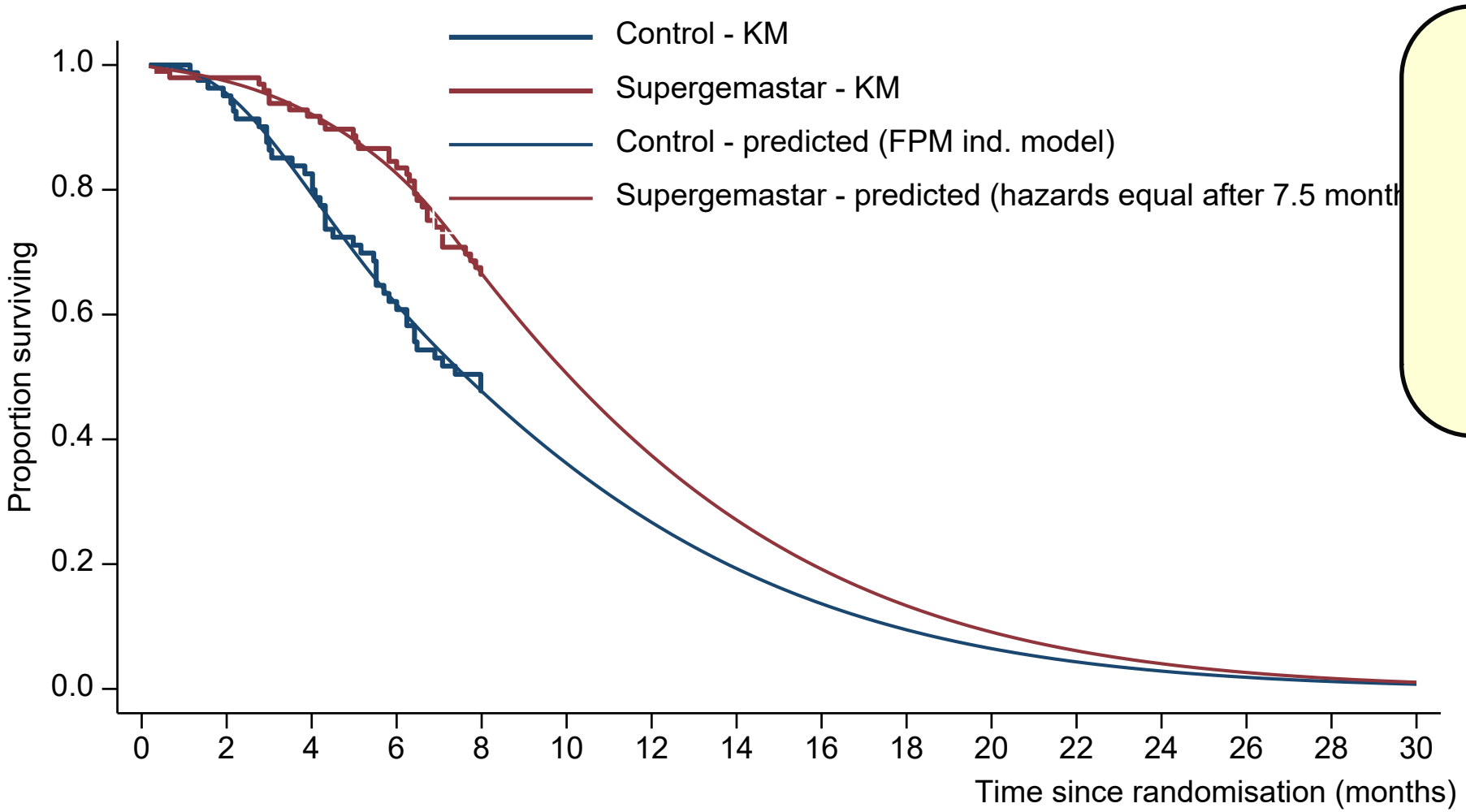
We're not sure if this analysis is appropriate. The survival curves cross because when we fit the survival models independently to each group, the estimated hazard for Supergemastar is higher than the hazard in the control group after 7.5 months

Mock Appraisal: So the ERG present another scenario…



So we ran another analysis. This shows what happens if hazards in the Supergemastar group don't increase above those in the control group, but instead converge at the point that they are projected to cross by the survival models. So here the hazard ratio = 1 from month 7.5

Mock Appraisal: So the ERG present another scenario…



In this scenario the survival curves don't cross, but they do gradually converge.

Mock Appraisal: The Committee needs to decide…

Analysis	Mean Overall Survival			Treatment effec assumptions
	Control Group	Supergemastar	Difference	
Company	8.1	11.4	3.3	HR observed in trial lasts forever
ERG	9.1	9.7	0.6	No enforced assumption, but implied HR>1 after 7.5 months
ERG scenario	9.1	11.2	2.1	HR converges to 1 at 7.5 months

Committee, you need to think. The different analyses give very different estimates of survival and very different incremental cost effectiveness ratios. They affect whether or not Supergemastar falls within accepted cost-effectiveness thresholds…

Poll #2: As a decision-maker, what would you do?

a) Accept the manufacturer's analysis
b) Accept the ERGs analysis
c) Consider a range of scenarios

My views

- Extrapolation is always going to be a problem
- There are two linked issues:
  i. **Extrapolating baseline survival** – which model to use, how to justify it…
  →But at least we have a chance – usually longer term experience with standard treatments

  ii. **Extrapolating the treatment effect** – will it increase / decrease / stay the same?
  →What do we base our assumptions on? The treatment is ***new***
  →***But we cannot avoid this question. It is fundamental to the cost-effectiveness analysis***

My views

- Extrapolation is always going to be a problem
- There are two linked issues:
  i. **Extrapolating baseline survival** – which model to use, how to justify it…
  →But at least we have a chance – usually longer term experience with standard treatments

  ii. **Extrapolating the treatment effect** – will it increase / decrease / stay the same?
  →What do we base our assumptions on? The treatment is ***new***
  →***But we cannot avoid this question. It is fundamental to the cost-effectiveness analysis***
- We have seen there are many methods that we can use, assuming various things about the treatment effect
- **Are any of them correct?**

My views

- **We don’t know!**

- Fundamentally, we are trying to predict something that is currently unknown
- We cannot know what is right and what is wrong

- **But…**

What is acceptable? My views

- We may not know the truth, but we should do our best to come up with credible estimates
- We should back these up with sensible rationale. This might include:
  - **Hazard plots** – what are their trajectories?
  - **Earlier phase studies** – have we learnt anything about the durability of the treatment effect?
  - **Nature of the disease and how the drug works** – is their biological plausibility for a durable effect?
  - **External data** – can we learn anything from other drugs for the same disease, or the same drug for different diseases?
  - **Expert opinion** – is there consensus? Can we put some sensible constraints on our analyses?
  - **Treatment pathway** – do people move onto other treatments? Does this have implications for the implied long-term treatment effect?
  - **Treatment discontinuation** – if people stop treatment, will the hazard ratio remain constant?

- **This is likely to involve a consideration of several plausible scenarios**

My views

- Understandable that there is frustration when there is a perceived lack of consistency around dealing with treatment effect waning
- But there are a lot of things to consider so it is understandable that assumptions about treatment effect waning won't always be the same
  - Case-by-case basis
  - Depends on the economic model and what is being modelled explicitly/implicitly
  - Also evidence and clinical expert advice can (and does) change over time
- Decision makers will consider the evidence put before them – make sure what you present is credible, and that you try to justify it using appropriate methods

My views

- Understandable that there is frustration when there is a perceived lack of consistency around dealing with treatment effect waning
- But there are a lot of things to consider so it is understandable that assumptions about treatment effect waning won't always be the same
  - Case-by-case basis
  - Depends on the economic model and what is being modelled explicitly/implicitly
  - Also evidence and clinical expert advice can (and does) change over time
- Decision makers will consider the evidence put before them – make sure what you present is credible, and that you try to justify it using appropriate methods

And, finally…

- Extrapolation of treatment effects is largely a data problem
  - It is the lack of data that causes the uncertainty
  - In the absence of data, it is not unreasonable to consider a range of plausible scenarios

Poll #3: From your point of view, how conservative should decision makers be when accounting for treatment effect waning?

a) Very conservative
b) Somewhat conservative
c) As realistic as possible
d) I don't know

The University Of Sheffield.

yorkshire cancer research
Saving Yorkshire Lives

Poll #4: From your point of view, should treatment effect waning be considered in each cost-effectiveness assessment of a novel oncology therapy?

a) Yes
b) No
c) Only under specific circumstances?
d) I don't know