Communicating statistical concepts to a Health Technology Assessment audience

PSI Conference 2025, 11th June

Professor Nick Latimer, University of Sheffield, Petauri Evidence
n.latimer@sheffield.ac.uk

Disclosures

- I was a member of a NICE Appraisal Committee for 5 years
- I'm a member of NICE's Decision Support Unit
- I work part-time for Petauri Evidence (previously known as Delta Hat), a consultancy company
- **These are my own opinions, not necessarily those of NICE, the DSU, or Petauri Evidence**

Abbreviations: NICE, National Institute for Health and Care Excellence; DSU, Decision Support Unit

Introduction

- We need to make our work understandable to whoever our audience is
- A company submission to an HTA agency forms a critical part of their “communication”
- The general “story” of a submission to a HTA agency must be easily understandable by everyone

→ But my focus today is going to be on the complex statistical analyses that often form an important part of submissions made to HTA agencies

Abbreviations: HTA, Health Technology Assessment

Introduction

Two key points to bear in mind:

1. **For complex statistical analyses, methods experts on Appraisal Committees are your audience**

- For these aspects, committees look to their methods experts for guidance

2. **Transparency is key**

- A crucial component of communication is transparency
- An Appraisal Committee will be reluctant to make a positive recommendation if important information is not provided

Abbreviations: HTA, Health Technology Assessment

Introduction

This talk

- Two examples of complex statistical analyses submitted to HTA agencies where transparency is usually lacking
- From the perspective of a Committee methodologist
- Inspired by real submissions, but the examples are not real

1. **Analyses to adjust for treatment switching in an RCT**
2. **Extrapolation of treatment effects**

Abbreviations: HTA, Health Technology Assessment; RCT, Randomised Controlled Trial

Treatment switching

- Imagine that you are the methods expert on an Appraisal Committee
- I’m going to present some treatment switching adjustment analyses for you to consider
- But first, some background on treatment switching

Treatment switching – background

- Treatment switching is when patients in an RCT switch onto subsequent treatments during the trial follow-up period. Often an issue in cancer trials, and affects survival outcomes

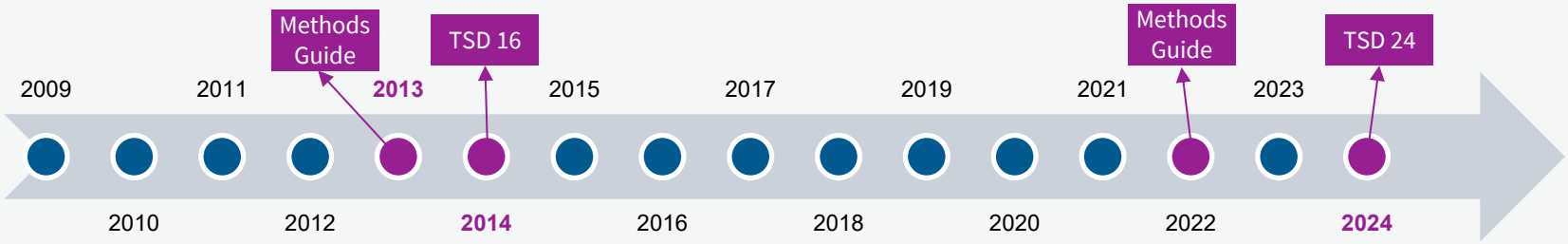
HTA decision problem: What is the effectiveness and cost-effectiveness of inserting the new treatment into the treatment pathway at the specified line of therapy, compared to retaining the current standard treatment pathway?

- If patients switch onto treatments that are not used in clinical practice, a standard intention-to-treat (ITT) analysis does not directly address the decision problem
- An analysis that estimates counterfactual outcomes that would have been observed in the absence of switching is more relevant – a hypothetical estimand

Abbreviations: HTA, Health Technology Assessment; ITT, Intention-to-treat; RCT, Randomised Controlled Trial

Treatment switching – background

- Adjusting for treatment switching is not a new thing in HTA
- NICE addresses switching in its Methods Guide, and the NICE Decision Support Unit has published two technical support documents on the topic



- Unfortunately, that does not mean that analyses are always done well!

Abbreviations: HTA, Health Technology Assessment; NICE, National Institute for Health and Care Excellence; TA, Technology Appraisal; TSD, Technical support document

Methods for adjusting for switching

1. **Inverse Probability of Censoring Weights (IPCW)** ["no unmeasured confounding" assumption and prone to error when switch proportions are high due to high weights]
2. **Two-Stage Estimation (TSE)** ["no unmeasured confounding" assumption and switching soon after a "secondary baseline"]
3. **Rank Preserving Structural Failure Time Model (RPSFTM)** ["common treatment effect" assumption]

- Each method can produce unbiased results when assumptions hold
- But often the methods (and different specifications of the same method) give (importantly) different results

- Need to consider plausibility of assumptions, test sensitivity of results, assess model outputs
- These things are covered in detail in NICE Decision Support Unit TSDs 16 and 24

Latimer NR, Abrams KR. NICE DSU Technical Support Document 16: Adjusting survival time estimates in the presence of treatment switching. (2014); Bell Gorrod, H., Latimer, N. R., Abrams K.R. NICE DSU Technical Support Document 24: Adjusting survival time estimates in the presence of treatment switching: An update to TSD 16. 2024. https://www.sheffield.ac.uk/nice-dsu/tsds/survival-analysis

Abbreviations: IPCW, Inverse Probability of Censoring Weights

Ultramab Vs Standalabine

Population: Previously untreated patients with metastatic cancer X

Intervention: Ultramab

Comparator: Standalabine

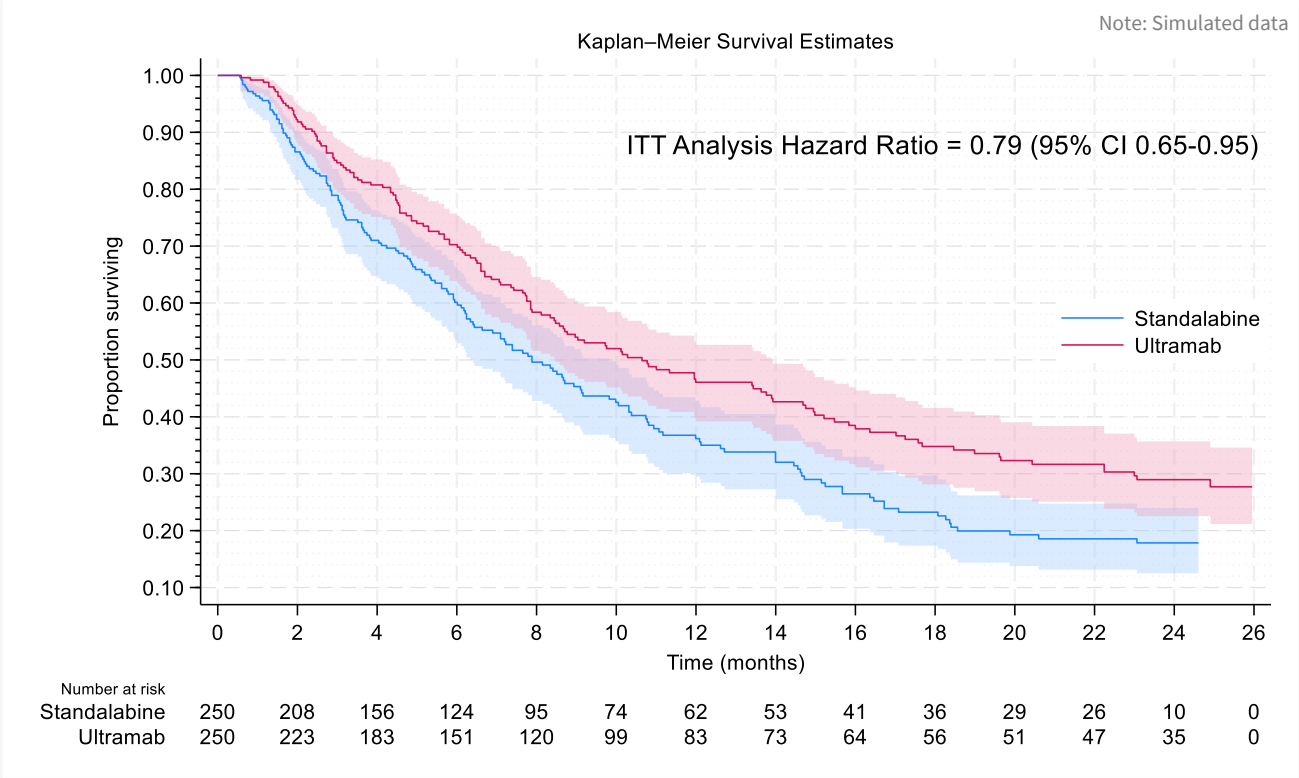
NICE Decision Problem: Assess the effectiveness and cost-effectiveness of inserting Ultramab into the treatment pathway for patients with cancer X

Imagine you are the methods expert on the NICE Appraisal Committee

Abbreviations: NICE, National Institute for Health and Care Excellence

Ultramab Vs Standalabine

Trial Results



Ultramab Vs Standalabine

University of Sheffield

Trial Characteristics

- Many patients randomised to the control group switched onto ultramab after disease progression, but ultramab is not available as a next-line treatment in the NHS
    - 230 patients in the control group experienced disease progression
    - 200 (87%) of these switched onto ultramab
    - 30 (13%) switched onto standard NHS treatments

Abbreviations: RCT, Randomised Controlled Trial; NHS, National Health Service

Ultramab Vs Standalabine

University of Sheffield

Company Submission

- The company report that it is relevant to adjust for switching from the control group onto ultramab
- The company considered RPSFTM, TSE and IPCW adjustment methods
  - They conclude that RPSFTM is not appropriate because they do not believe that the common treatment effect assumption is plausible
  - They conclude that TSE is not appropriate because some of the switchers did not switch immediately upon disease progression
  - Therefore, they applied the IPCW method
  - → They censored patients who switched from standalabine to ultramab, then used weighting to adjust for potential selection bias

Abbreviations: RPSFTM, Rank Preserving Structural Failure Time Model; TSE, Two-Stage Estimation; IPCW, Inverse Probability of Censoring Weights

Ultramab Vs Standalabine

University of Sheffield

Company Submission

- The company considered two model specifications for their IPCW analysis
  - Model 1 included all potentially prognostic variables collected in the trial, but the company do not describe how these variables were identified
  - Model 2 included a smaller selection of covariates, but approach for variable selection is not described

Analysis	Hazard Ratio	Incremental Cost-Effectiveness Ratio
Intention-to-treat	0.79 (95% CI: 0.65-0.95)	£75,000 per QALY gained
IPCW Model 1	0.43 (95% CI: 0.31-0.65)	£28,000 per QALY gained
IPCW Model 2	0.53 (95% CI: 0.40-0.71)	£38,000 per QALY gained

The company chose to use IPCW Model 1 in their base case because “it includes all potentially prognostic variables, and the IPCW method relies upon the ‘no unmeasured confounding’ assumption”. Model coefficients and the distribution of the weights are not presented

Abbreviations: IPCW, Inverse Probability of Censoring Weights; QALY, Quality Adjusted Life Year

Ultramab Vs Standalabine

As the methods expert on your NICE Appraisal Committee, are you content with this analysis?

Abbreviations: NICE, National Institute for Health and Care Excellence

Ultramab Vs Standalabine

University of Sheffield

Problems identified by the Committee methods expert

- The rationale for ruling out RPSFTM and TSE is not sufficient
  - Sensitivity analysis around the common treatment effect assumption could have been done (RPSFTM)
  - It is not reported how many switchers had switching delayed sometime after disease progression, or how long the delay was (TSE)
- IPCW may be unreliable given the high switching proportion
  - Why do the two different model specifications give such different results?
  - Confounding in the smaller model, or extreme weights / breakdown of positivity in the “full” model?
  - Detail on the weighting models and the distribution of the weights is needed

Abbreviations: RPSFTM, Rank preserving structural failure time mode; NUC, No unmeasured confounding; TSE, Two-stage estimation; IPCW, Inverse Probability of Censoring Weights; TSD, Technical Support Document

Ultramab Vs Standalabine

University of Sheffield

Problems identified by the Committee methods expert

- The rationale for ruling out RPSFTM and TSE is not sufficient
  - **Sensitivity analysis** around the common treatment effect assumption could have been done (RPSFTM)
  - **It is not reported** how many switchers had switching delayed sometime after disease progression, or how long the delay was (TSE)
- IPCW may be unreliable given the high switching proportion
  - Why do the two different model specifications give such different results?
  - Confounding in the smaller model, or extreme weights / breakdown of positivity in the "full" model?
  - **Detail on the weighting models and the distribution of the weights is needed**

These things are all recommended in TSDs 16 and 24:

- **Sensitivity analysis** (including around the common treatment effect assumption)
- **Reporting** on the nature of the switching observed
- **Detail** on model outputs (including weighting models and distributions of the weights)

→ **But are rarely done!**

Abbreviations: RPSFTM, Rank preserving structural failure time mode; NUC, No unmeasured confounding; TSE, Two-stage estimation; IPCW, Inverse Probability of Censoring Weights; TSD, Technical Support Document

Ultramab Vs Standalabine

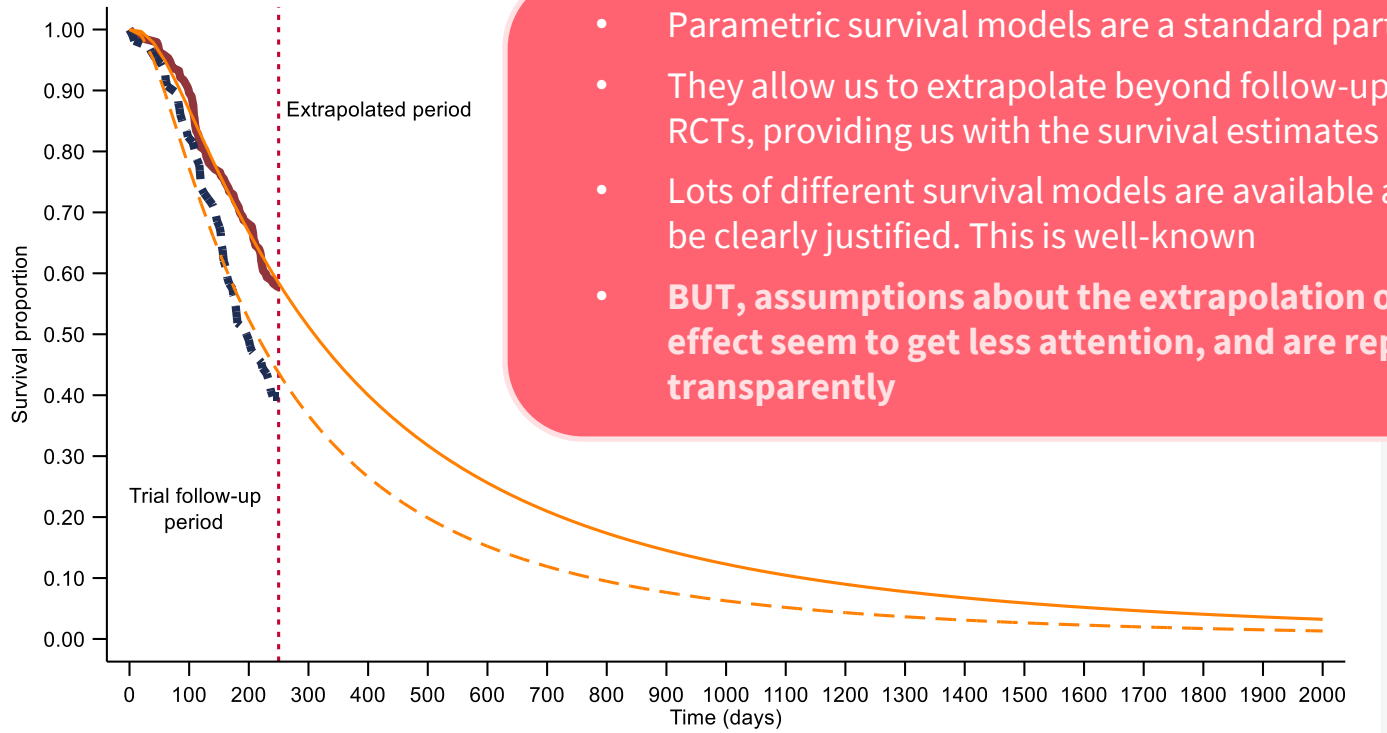
- I would not be satisfied with this analysis
- I would be suspicious that the company had cherry-picked the analysis that made them look best – a lack of clarity and transparency leads to a lack of trust

- The analysis would be rejected, and the company would need to do lots more work
- The decision-making process would be slowed down considerably
- **This happens A LOT!**

For effective, efficient communication with the Appraisal Committee, and to build trust:

- ➔ **Methodological choices need to be justified clearly**
- ➔ **Complex analyses must be reported transparently**
- ➔ **Guidance documents need to be followed!**

Modelling the treatment effect



- Parametric survival models are a standard part of HTA
- They allow us to extrapolate beyond follow-up periods observed in RCTs, providing us with the survival estimates we need for HTA
- Lots of different survival models are available and choices have to be clearly justified. This is well-known
- **BUT, assumptions about the extrapolation of the treatment effect seem to get less attention, and are reported much less transparently**

Abbreviations: HTA, Health Technology Assessment; RCT, Randomised Controlled Trial

Modelling the treatment effect

- Two approaches for modelling the treatment effect in a survival context

1. Fit one survival model to all the trial data, with treatment group as a covariate
2. Fit two independent survival models – one to each treatment group

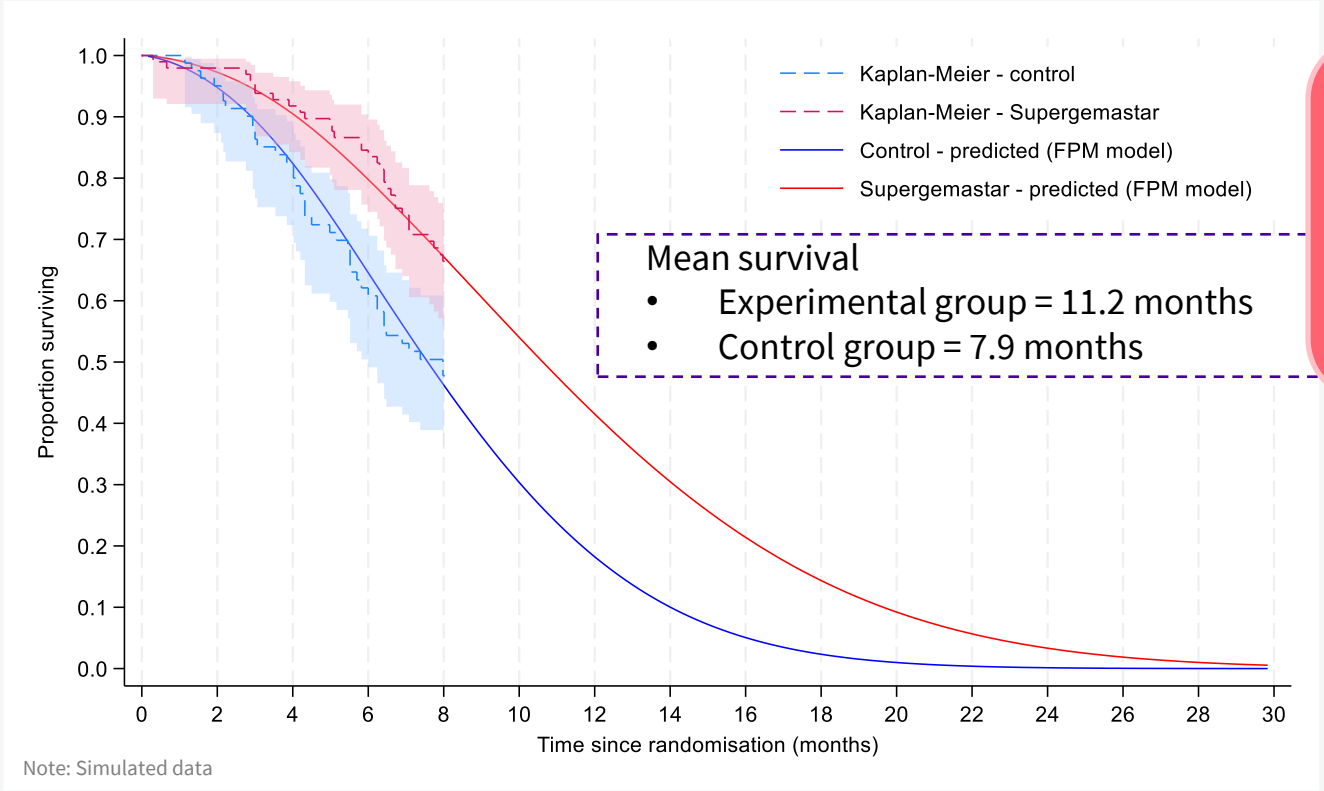
- If we choose (1), we assume that the treatment effect is constant over time
- If we choose (2) we can allow for a treatment effect that changes over time

- Often (2) is chosen, to avoid assuming a constant treatment effect…

… but this means that what *is* being assumed about the treatment effect is unclear

Abbreviations: HTA, Health Technology Assessment; RCT, Randomised Controlled Trial

Imagine – Company analysis



Mean survival

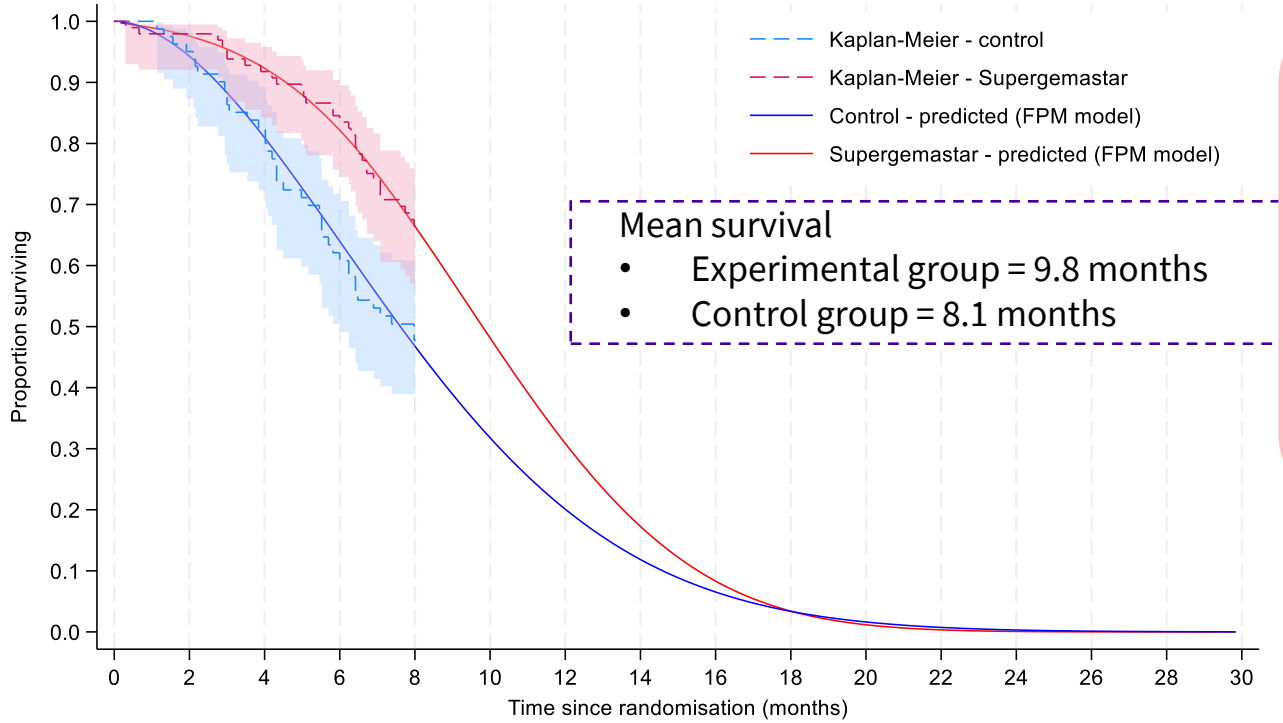
- Experimental group = 11.2 months
- Control group = 7.9 months

The company choose option (1) – fit one model with treatment group as a covariate, assuming a constant treatment effect

Note: Simulated data

Abbreviations: FPM, Flexible Parametric Model

Imagine – EAG analysis



Mean survival

- Experimental group = 9.8 months
- Control group = 8.1 months

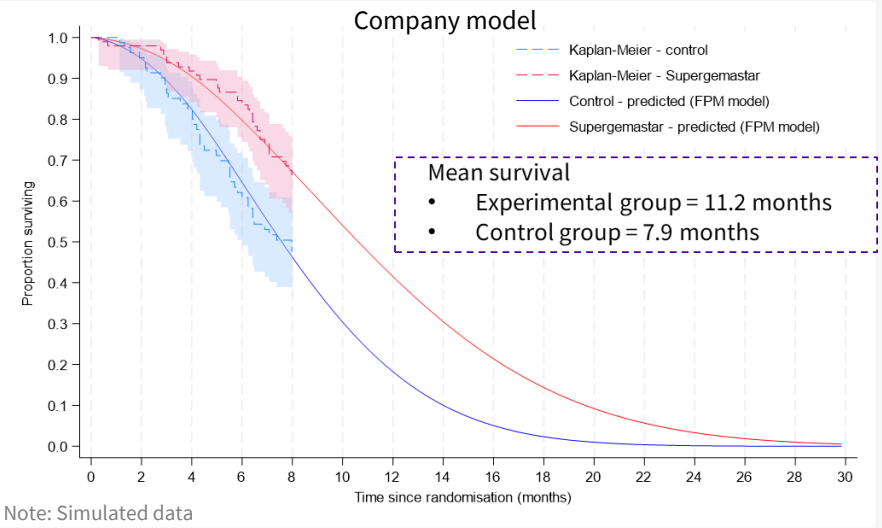
Note: Simulated data

The External Assessment Group choose option (2) – fit an independent model to each treatment group, making no assumption about the treatment effect

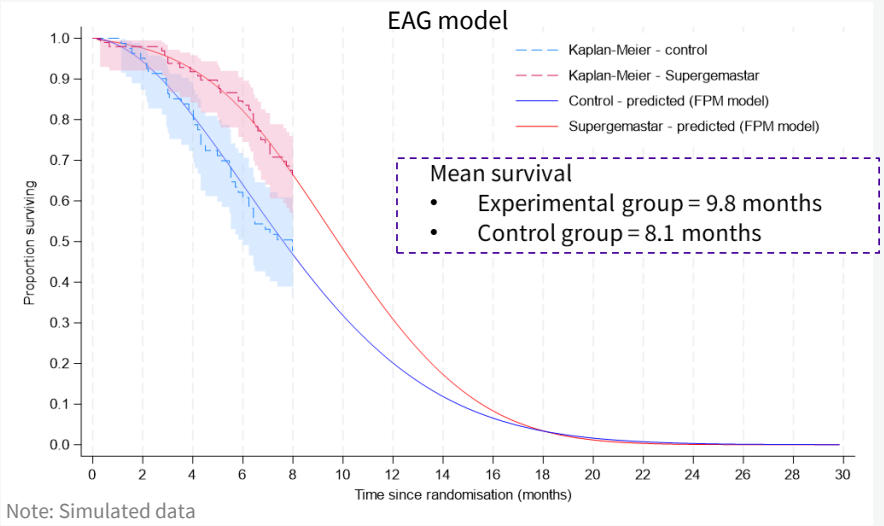
Abbreviations: FPM, Flexible Parametric Model; EAG, External Assessment Group

Imagine – EAG analysis

- Which analysis do you prefer?
- Which would you base a decision on, if you were the methods expert on the Appraisal Committee?



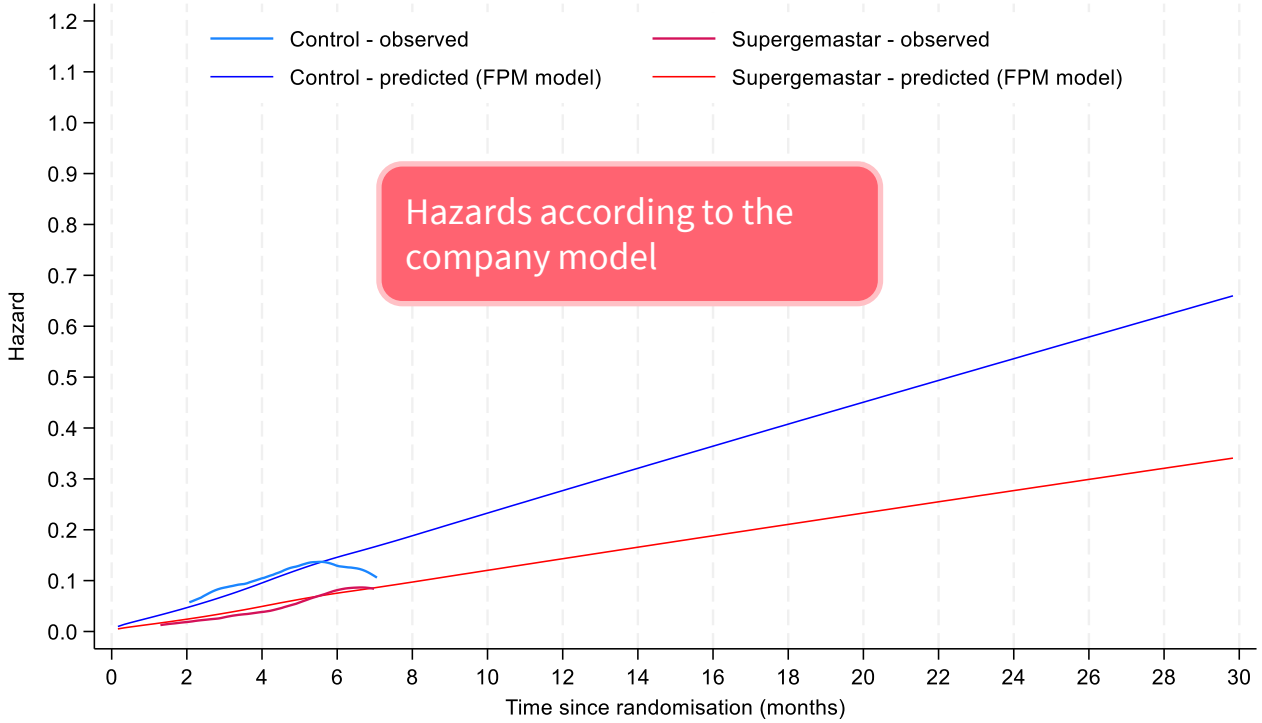
Note: Simulated data



Note: Simulated data

Abbreviations: FPM, Flexible Parametric Model; EAG, External Assessment Group

Present the hazards!



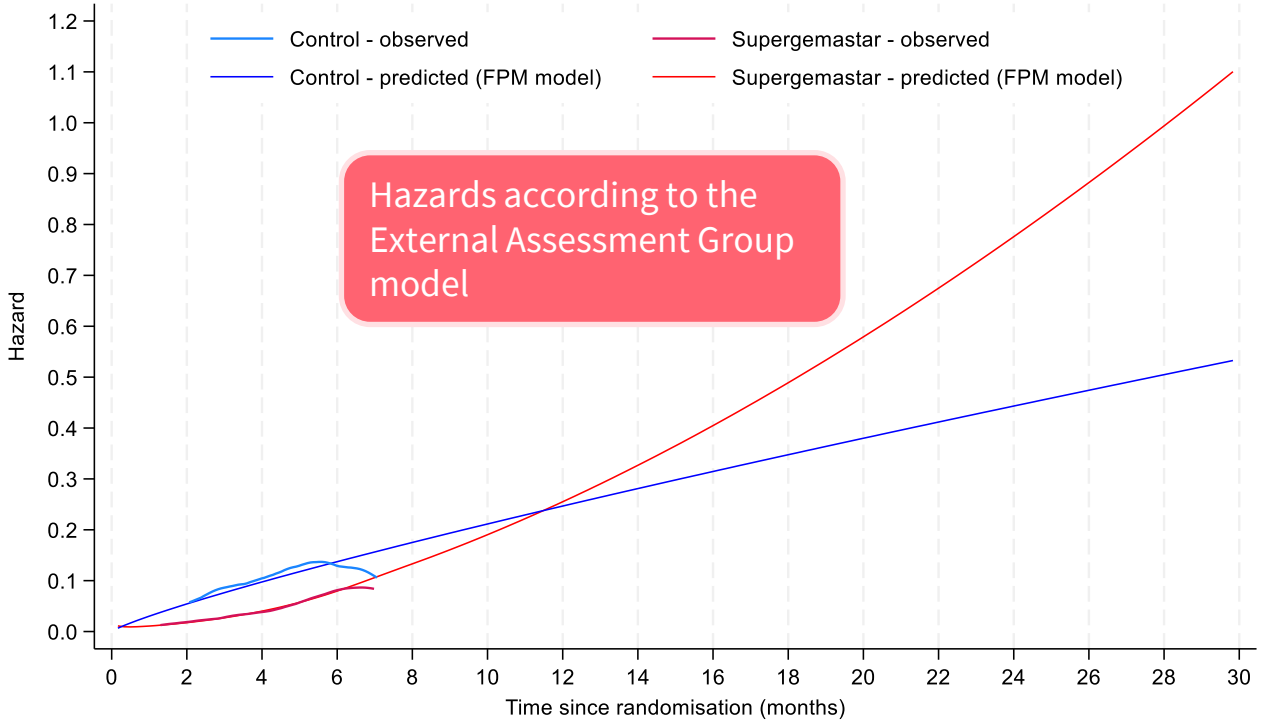
Before making a decision, we should assess the hazards, and the treatment effect implied by the models.

This may seem obvious! But these plots are very rarely presented (but are recommended by TSD 21)

Abbreviations: FPM, Flexible Parametric Model; TSD, Technical Support Document

TSD 21: Rutherford, MJ., Lambert,PC., Sweeting, MJ., Pennington, R., Crowther, MJ., Abrams, KR., Latimer, NR. NICE DSU Technical Support Document 21. Flexible Methods for Survival Analysis. 2020. https://www.sheffield.ac.uk/nice-dsu/tsds/flexible-methods-survival-analysis

Present the hazards!



Hazards according to the External Assessment Group model

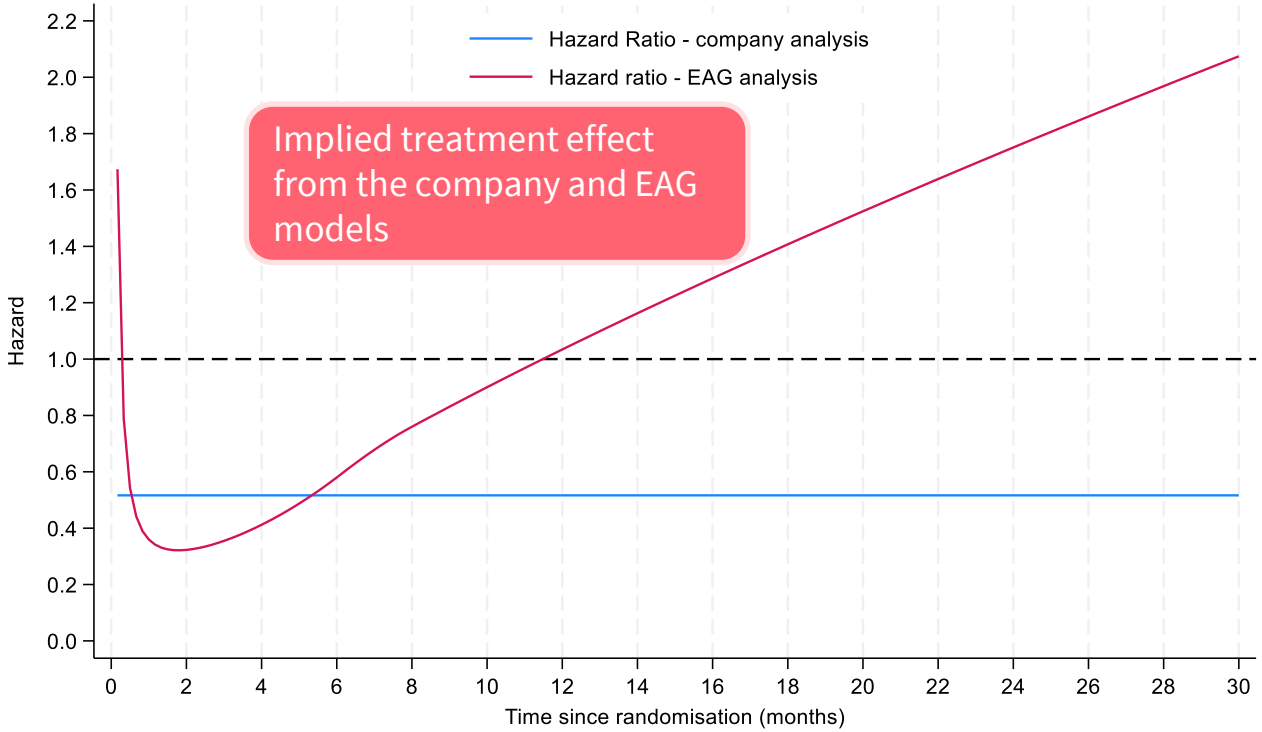
Before making a decision, we should assess the hazards, and the treatment effect implied by the models.

This may seem obvious! But these plots are very rarely presented (but are recommended by TSD 21)

Abbreviations: FPM, Flexible Parametric Model; TSD, Technical Support Document

TSD 21: Rutherford, MJ., Lambert,PC., Sweeting, MJ., Pennington, R., Crowther, MJ., Abrams, KR., Latimer, NR. NICE DSU Technical Support Document 21. Flexible Methods for Survival Analysis. 2020. https://www.sheffield.ac.uk/nice-dsu/tsds/flexible-methods-survival-analysis

And the implied treatment effect!



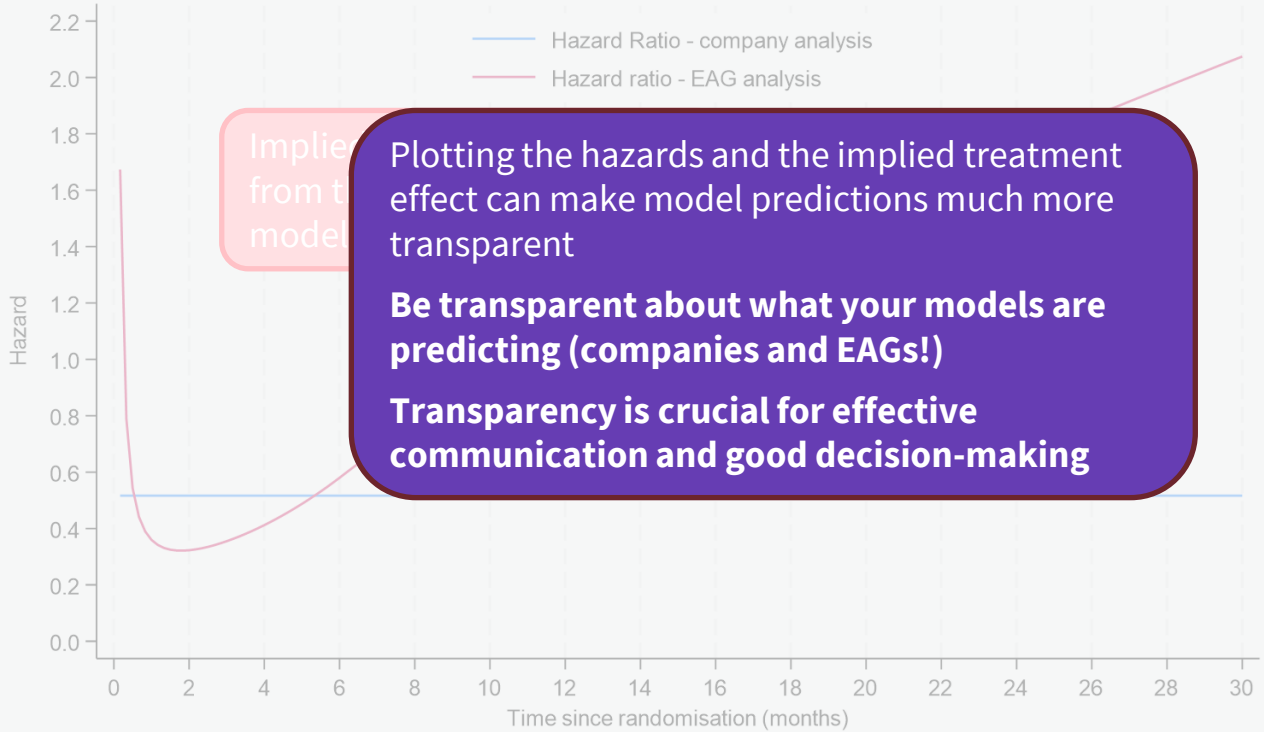
Before making a decision, we should assess the hazards, and the treatment effect implied by the models.

This may seem obvious! But these plots are very rarely presented (but are recommended by TSD 21)

Abbreviations: FPM, Flexible Parametric Model; EAG, External Assessment Group

TSD 21: Rutherford, MJ., Lambert,PC., Sweeting, MJ., Pennington, R., Crowther, MJ., Abrams, KR., Latimer, NR. NICE DSU Technical Support Document 21. Flexible Methods for Survival Analysis. 2020. https://www.sheffield.ac.uk/nice-dsu/tsds/flexible-methods-survival-analysis

And the implied treatment effect!



Plotting the hazards and the implied treatment effect can make model predictions much more transparent

Be transparent about what your models are predicting (companies and EAGs!)

Transparency is crucial for effective communication and good decision-making

Before making a decision, we should assess the hazards, and the treatment effect implied by the models.

This may seem obvious! But these plots are very rarely presented (but are recommended by TSD 21)

Abbreviations: FPM, Flexible Parametric Model; EAG, External Assessment Group

TSD 21: Rutherford, MJ., Lambert,PC., Sweeting, MJ., Pennington, R., Crowther, MJ., Abrams, KR., Latimer, NR. NICE DSU Technical Support Document 21. Flexible Methods for Survival Analysis. 2020. https://www.sheffield.ac.uk/nice-dsu/tsds/flexible-methods-survival-analysis

Summary

- When you do complex statistical analyses for HTA submissions, your audience is primarily the methods experts on the appraisal committee. Communication is through your submission
- **Clarity and transparency are key components of effective communication and trust**
  - Methodological choices need to be justified clearly
  - Complex analyses must be reported transparently

  } Tell the story of what you're doing and why
- **Keep up with methodological guidance documents**
  - Cancer submissions to NICE routinely refer to DSU TSD 14 on survival analysis (published in 2011)
  - Very few plot hazards and implied treatment effects (recommended in TSD 21, published in 2020)
  - TSD 24 (2024) presents detailed reporting guidelines for treatment switching – which are rarely followed
- **This is relevant for many types of complex statistical analyses**
  - Reporting lacks clarity and transparency even though guidance is available (another example is the NICE Real World Evidence Framework, which is rarely cited in submissions)
  - This harms communication and trust, and makes decision-making inefficient

Abbreviations: HTA, Health Technology Assessment; NICE, National Institute for Health and Care Excellence; DSU, Decision Support Unit; TSD, Technical Support Document

Thanks for listening