University of Sheffield

Beyond “guesstimates" in long-term survival extrapolation:

Introducing a comprehensive step-by-step elicitation framework

Jessica E. Forsyth, Min-Hua Jen,
Jacoline Bouvy, Kate Ren

18.11.24

Speakers and Overview of the Workshop

Dr Jessica E. Forsyth
Research Associate
SCHARR,
University of Sheffield

Dr Min-Hua Jen
Senior Director: Real-World & Access analytics
Eli-Lilly

Dr Jacoline Bouvy
Programme Director - Medicines Evaluation
National Institute for Health and Care Excellence (NICE)

Dr Kate Ren
Senior Research Fellow
SCHARR, University of Sheffield
Director of Statistics
ConnectHEOR, London

Structured expert elicitation:
Definition & existing frameworks

Structured expert elicitation (SEE)

Representing subjective uncertainty about a fixed quantity using a **probability distribution**

Provides indication of uncertainty

Using a formal methodology to minimise potential bias

Leverages expert knowledge in areas of uncertainty

Provides a robust framework for integrating complex evidence and managing uncertainty

Potential biases when making probability judgements

Over confidence: not assigning enough probability to more extreme values

- Unlikely events are difficult to think about and account for when making judgements
- Facilitator helps to avoid this (along with minimisation of other biases)

Anchoring: fixing judgements on an initial value and failing to adjust after presentation of evidence

- Careful ordering of questions and establish expert opinions independently

Availability: basing judgements on evidence dependent on how easily examples come to mind

- Comprehensive evidence dossier and expert review of evidence helps reduce this

Motivational: confirmation or (un)desirability bias

- Focussing on evidence consistent with existing beliefs or preferences
- Pre-existing optimistic or pessimistic views

SHELF: The Sheffield Elicitation Framework

- A package of templates, software and guidance for conducting structured expert elicitation in person or online

Format of facilitated workshops (in person or online)
1. Training 2. Elicitation and recording of individual judgements 3. Review of individual judgements and group discussion 4. Aggregation: invoke notion of a "**rational impartial observer**" **(RIO)**

- Latest update to include a bespoke adaptation/update for long-term survival outcomes

Other protocols/frameworks

Cooke's method

- Seed/calibration questions
- Experts provide individual judgements
- Linear pooling with weights based on answers to seed questions

Delphi

- Experts provide individual judgements
- Anonymous discussion of responses
- Iterative process
- Develop group consensus

IDEA

- Combine Cooke's and Delphi
- Promoted as a less time-intensive way to conduct SEE
- Option to separate the elicitation of individual judgements and discussion periods

MRC Protocol

- Healthcare decision making (HCDM)
- Comprehensive review of SEE used within HCDM
- General guidance on conduct and reporting

General issues when using SEE

Lack of experience making probabilistic judgements

Heuristics and biases

Pessimistic views of elicitation value

Managing multiple experts' opinions

Transparency

Poll question

Have you consulted experts with regards to long-term survival outcomes?

1. Yes- but for validation purposes only (including model selection)
2. Yes- including a formal elicitation exercise for quantitative judgements
3. Yes- other
4. No- existing literature was sufficient
5. No- not necessary or relevant to my work

References

- Oakley JE, O'Hagan A. SHELF: the Sheffield Elicitation Framework (version 4). https://shelf.sites.sheffield.ac.uk/
- Cooke, R. M. (1991). *Experts in Uncertainty: Opinion and Subjective Probability in Science*. Oxford University Press.
- European Food Safety Authority (2014). Guidance on expert knowledge elicitation in food and feed safety risk assessment. EFSA Journal 2014, 12(6):3734, 278 pp. DOI: https://doi.org/10.2903/j.efsa.2014.3734
- Hemming V, Walshe TV, Hanea AM, et al. Eliciting improved quantitative judgements using the IDEA protocol: A case study in natural resource management. PLoS ONE 2018; 13: e0198468. 20180622. DOI: 10.1371/journal.pone.0198468.
- Bojke L, Soares M, Claxton K, et al. Developing a reference protocol for structured expert elicitation in health-care decision-making: a mixed-methods study. Health Technology Assessment (Winchester, England) 2021; 25: 1.

Approaches to Elicit Long-Term Survival Estimates in Healthcare

Challenges, Approaches and Industry perspectives

Min-Hua Jen
Senior Director, Eli Lilly & the Company

Chair-Elect, ISPOR Oncology SIG

18th November 2014

Agenda

- Why long-term survival matters
- How industry leverage expert insights
- The toolbox for long-term survival
  - Overview of approaches
  - Practical challenges
  - Limitations
- Industry perspective of elicited quantities
- Conclusion

Lilly

Disclaimer

The content presented is intended for informational purposes only and reflects the views of the presenter based on available data and research.

Why long-term survival matters

- Critical for assessing treatment efficacy, healthcare policy, and cost-effectiveness
- Short-term clinical data often insufficient for long-term projections
- Need for methods to estimate survival beyond available data

Lilly

How Industry leverage expert insights

- Use of expert opinion when clinical trial data is limited
- Applications in cost-effectiveness analyses, long-term extrapolation in clinical trials
- Impact on decision-making for drug approval and reimbursement

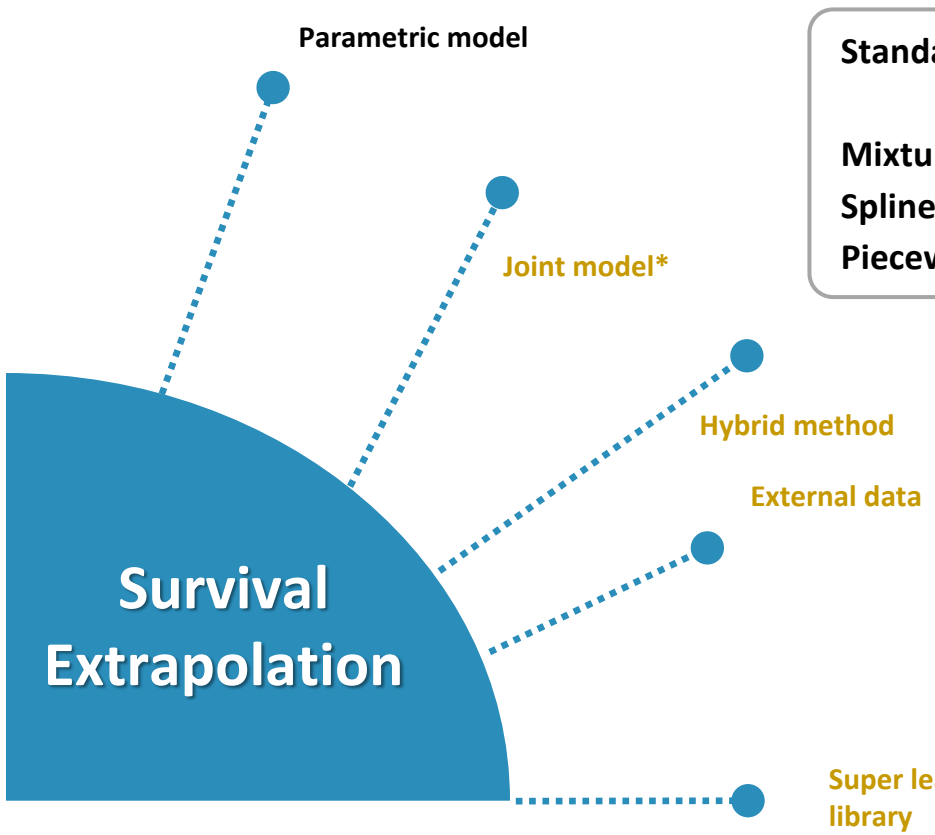
Lilly

The toolbox for long-term survival

- Use of statistical models, fitting techniques, and real-world evidence
- Use of expert elicitation opinion
  - Unstructured expert elicitation methods
- Structured expert elicitation methods

Lilly

Methods for Survival Extrapolation

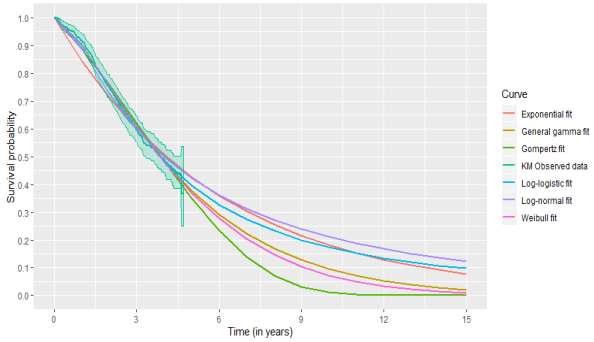


Standard: Exponential, Weibull, Gompetz, Log-normal, Log-logistic, Gamma

Mixture-based model: Cure model, Mixture exponential

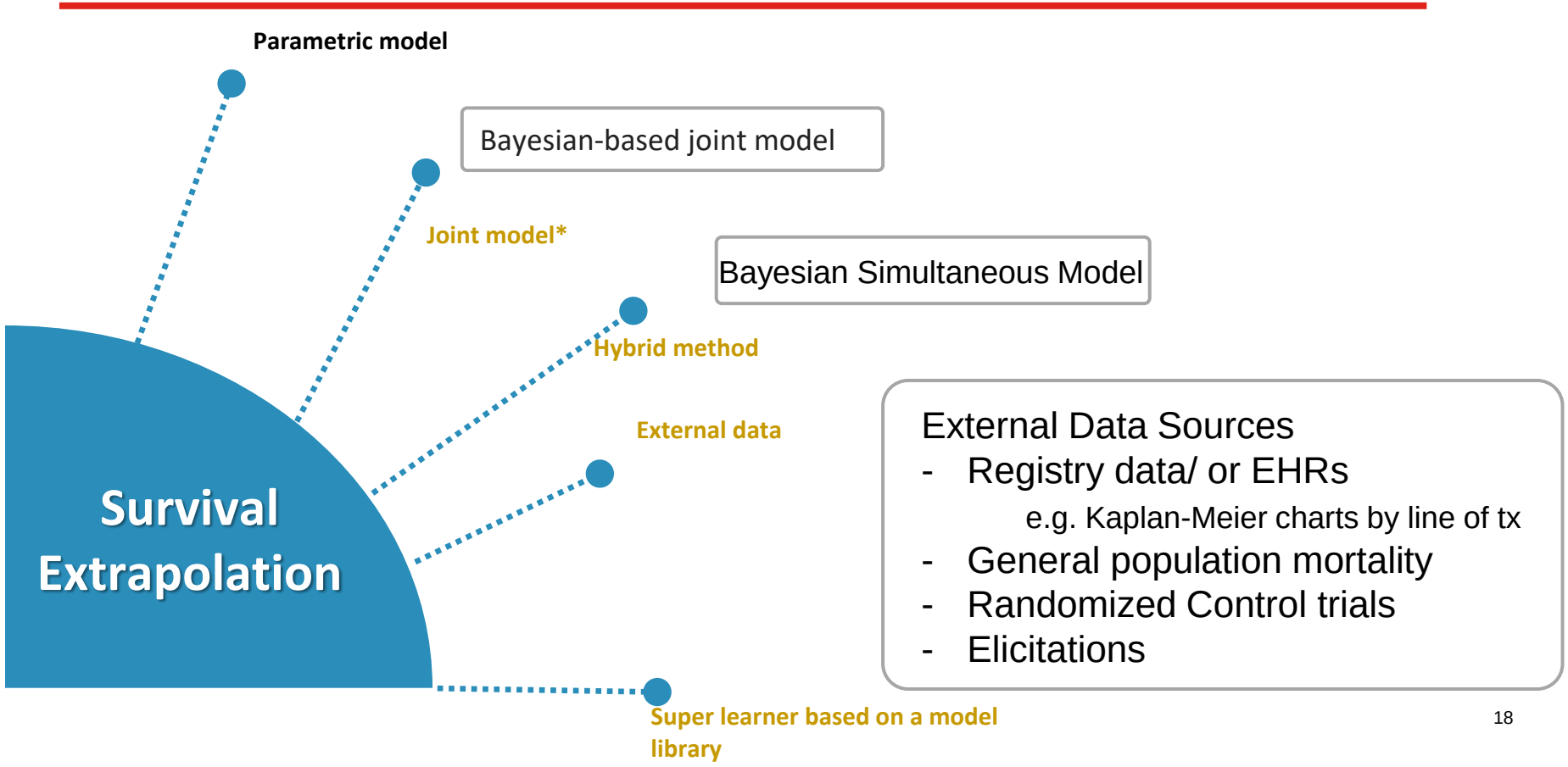
Spline-based model

Piecewise model



*Traore, S., Sashegyi, A., Winfree, K. B., Taipale, K. L., & Jen, M. H. (2023). Bayesian survival extrapolation for cost-effectiveness analysis: a case study of RELAY for ramucirumab in combination with erlotinib in the treatment of non-small-cell lung cancer. *Journal of Medical Economics*, *26*(1), 1479–1488. https://doi.org/10.1080/13696998.2023.2272534

Methods for Survival Extrapolation



*Traore, S., Sashegyi, A., Winfree, K. B., Taipale, K. L., & Jen, M. H. (2023). Bayesian survival extrapolation for cost-effectiveness analysis: a case study of RELAY for ramucirumab in combination with erlotinib in the treatment of non-small-cell lung cancer. *Journal of Medical Economics*, *26*(1), 1479–1488. https://doi.org/10.1080/13696998.2023.2272534

The toolbox for long-term survival

- Use of statistical models, fitting techniques, and real-world evidence
- Use of expert elicitation opinion
  - Unstructured expert elicitation methods
- Structured expert elicitation methods

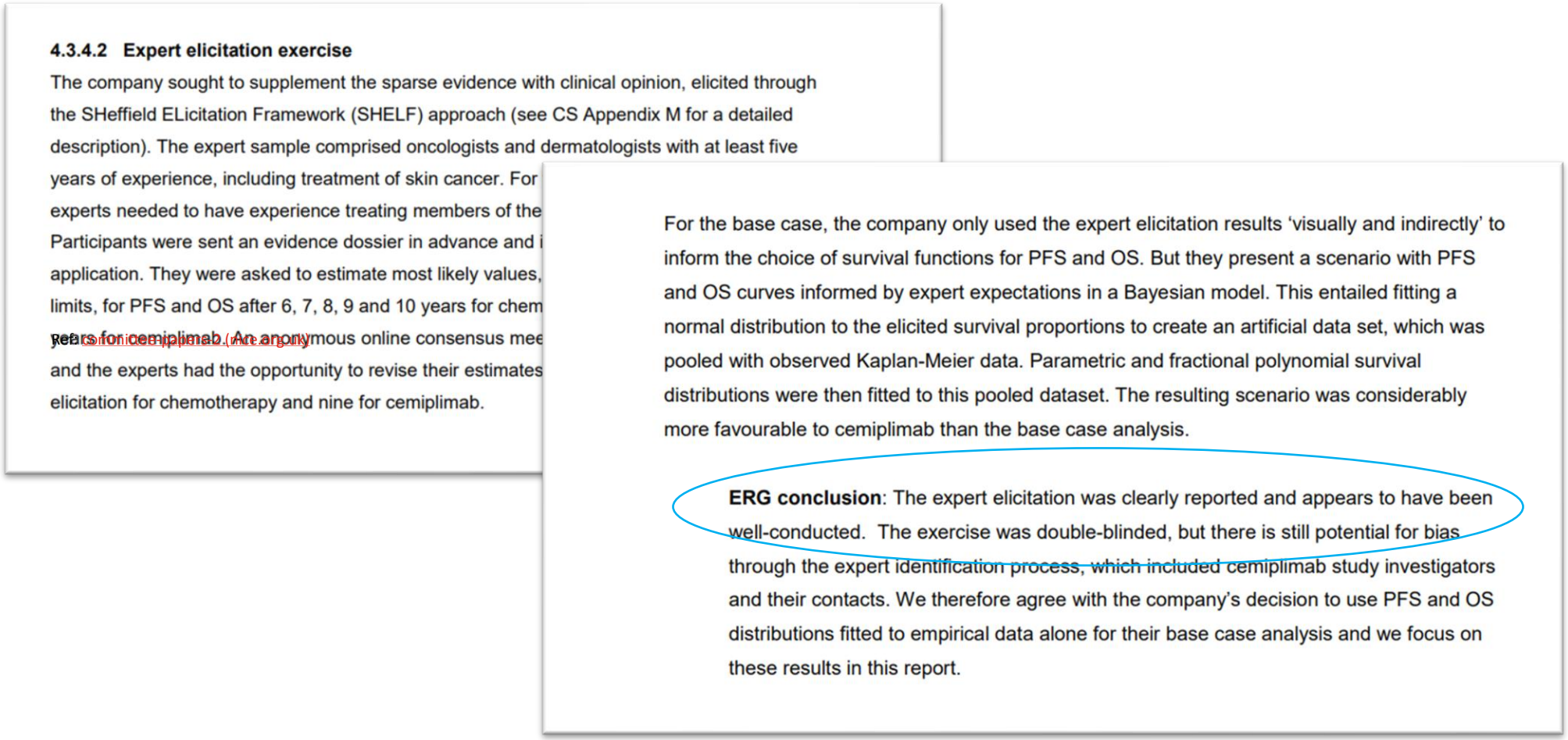
Lilly

The toolbox for long-term survival

- Use of statistical models, fitting techniques, and real-world evidence
- Use of expert elicitation opinion
  - Unstructured expert elicitation methods
- Structured expert elicitation methods

Lilly

The only SEE application for long-term survival in NICE appraisal



4.3.4.2 Expert elicitation exercise

The company sought to supplement the sparse evidence with clinical opinion, elicited through the SHeffield ELicitation Framework (SHELF) approach (see CS Appendix M for a detailed description). The expert sample comprised oncologists and dermatologists with at least five years of experience, including treatment of skin cancer. For experts needed to have experience treating members of the Participants were sent an evidence dossier in advance and i application. They were asked to estimate most likely values, limits, for PFS and OS after 6, 7, 8, 9 and 10 years for chem years for cemiplimab. An anonymous online consensus mee and the experts had the opportunity to revise their estimates elicitation for chemotherapy and nine for cemiplimab.

For the base case, the company only used the expert elicitation results 'visually and indirectly' to inform the choice of survival functions for PFS and OS. But they present a scenario with PFS and OS curves informed by expert expectations in a Bayesian model. This entailed fitting a normal distribution to the elicited survival proportions to create an artificial data set, which was pooled with observed Kaplan-Meier data. Parametric and fractional polynomial survival distributions were then fitted to this pooled dataset. The resulting scenario was considerably more favourable to cemiplimab than the base case analysis.

ERG conclusion: The expert elicitation was clearly reported and appears to have been well-conducted. The exercise was double-blinded, but there is still potential for bias through the expert identification process, which included cemiplimab study investigators and their contacts. We therefore agree with the company's decision to use PFS and OS distributions fitted to empirical data alone for their base case analysis and we focus on these results in this report.

Practical Challenges

Several practical challenges arise when eliciting long-term survival estimates:

- Limited Data: reliance on short-term trial results
- Expert variability: differences in clinical experience
- Model selection: risk of over- or underestimating survival
- Stakeholder Disagreement

Lilly

Limitation

- Data Gaps
- Model Assumptions
- Expert Knowledge
- Generalizability

Lilly

Question for the audience

When do you think SEE for long-term survival outcomes should be conducted in the HTA submission process?

1. Early within the submission preparation process
2. When choosing the most appropriate survival extrapolation
3. Whenever an advisory board or clinician meeting has been scheduled for other purposes

Industry Perspective on Elicited Quantities

- Essential for informed decision-making
- Collaboration with academia to enhance methods
- Regulatory push for transparency and evidence-based practices

Lilly

Conclusion

- Long-term survival estimates are crucial
- Diverse approaches each have unique strengths
- Ongoing improvements needed for reliability and accuracy

Lilly

University of Sheffield

Expert elicitation of long-term survival outcomes

A bespoke protocol

Kate Ren

18.11.24

Disclaimer

- This study is funded by the National Institute for Health and Care Excellence (NICE) Decision Support Unit (DSU).

- This study is used to inform the development of the NICE DSU Technical Support Document (TSD) on structured expert elicitation (SEE) for long-term survival outcomes.

Contents

- Justification for bespoke protocol for the elicitation of long-term survival outcomes

- Proposed framework
  - Preparation
  - Workshop
    - Scenario testing – consistency checking
    - Interaction between experts
  - Reporting

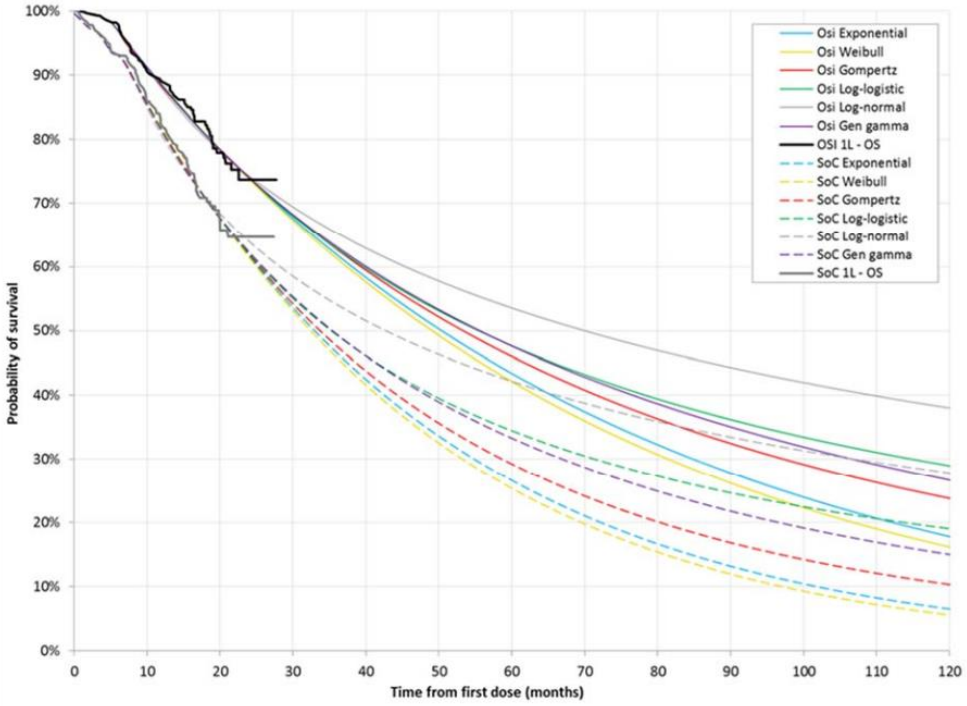
Why is a bespoke protocol needed for time-to-event data?

1. **Uniqueness of time-to-event data**
   - Censoring
   - Underlying hazard function

2. **Potential impact during reimbursement process**
   - Extrapolated survival outcomes can have large variability resulting in uncertain incremental cost effectiveness ratios

3. **Need an approach which ensures consistency with qualitative knowledge**
   - Meaningful representation of the elicited values given the qualitative knowledge

NICE TA621: Osimertinib for untreated EGFR-positive non-small-cell lung cancer



A bespoke protocol

What is required for SEE of long-term survival outcomes

Preparation for SEE of long-term survival outcomes

Define quantities of interest(s)

- Reference the PICO and decision problem

Select appropriate experts

- Consider experience, geographical location, conflicts, equality and diversity of invited experts

Select format of workshop

- Face-to-face or online
- A day workshop: 2-4 quantities of interest

Compile evidence dossier

- Comprehensive collection of all relevant information; reviewed by experts in advance

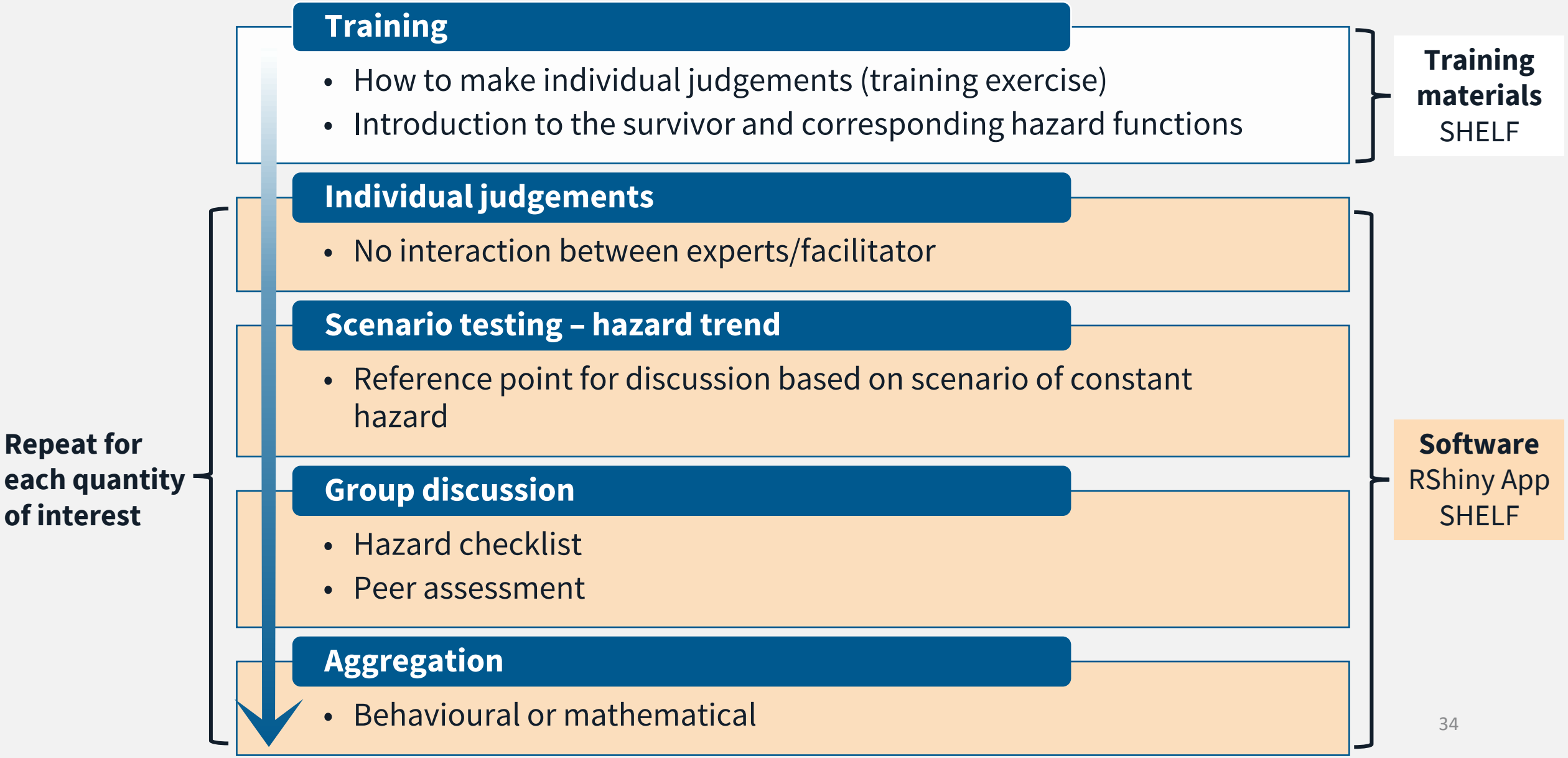
The evidence dossier – what data to present?

- ✓ Definition of PICO
- ✓ Quantities of Interest (QoIs)
- ✓ Context specific information

- ✓ Relevant Kaplan-Meier curves
  - **Without parametric model fits**

- ✓ Supporting literature
  - ✓ Direct evidence (trial publications)
  - ✓ External evidence
  - ✓ Relevant comparators

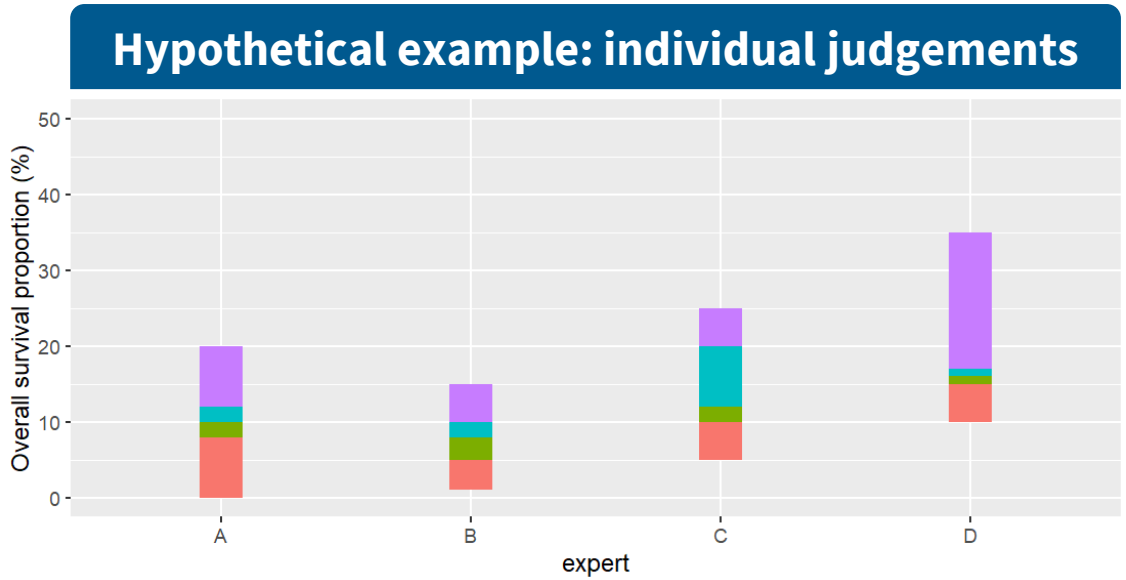
A step-by-step guide to the SEE workshop for long-term survival



Obtaining individual expert judgements

- Quartile method (variable interval method)
  - Lower and upper plausible limits
  - Lower quartile (Q1)
  - Upper quartile (Q3)
  - Median

- No discussion between experts
- Experts can clarify with the facilitator

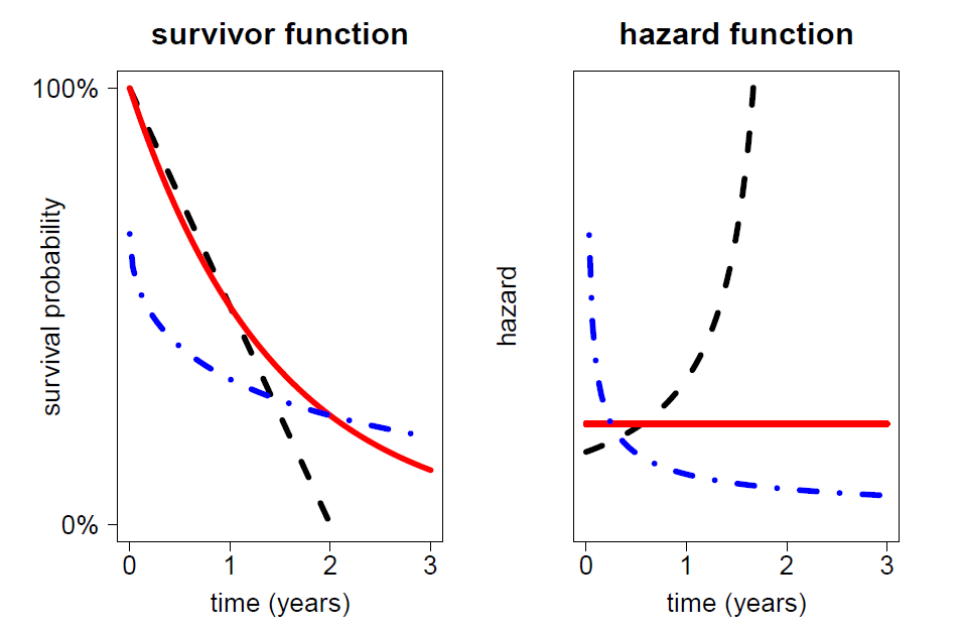


Scenario testing: 1) Importance of hazard function

Definition of hazard

Given that a person has survived to time t, what is the risk (or rate) of the event of interest occurring at that specific moment

- Different parametric models make different assumptions about the hazard



- Constant hazard
- Increasing hazard
- Decreasing hazard

Scenario testing: 2) Obtaining qualitative opinion on hazard trend

To identify whether expert believes the hazard in the target population to be changing with time and why

Increasing hazard

- **Disease**
  - Continuous deterioration
  - Aggressive disease
- **Patient group**
  - Advancing age
  - Trial inclusion/exclusion criteria
- **Treatment**
  - Not controlling the disease well
  - Delayed treatment effect

Decreasing hazard

- **Disease**
  - Improving naturally
- **Patient group**
  - Presence of subgroups within the population
- **Treatment**
  - Improvement of condition greater than deterioration due to disease
  - Cure

Scenario testing: 3) Comparing to a scenario of constant hazard

Procedure

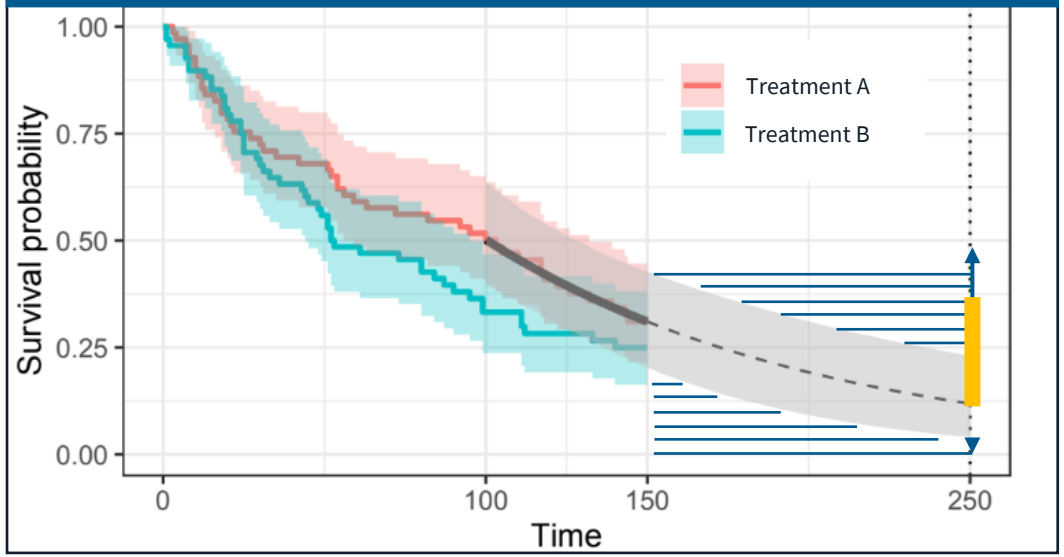
After individual judgements, the experts were presented with an extrapolation based on a particular scenario

An approximate 95% credible interval indicative of a **range of values statistically consistent** with the assumption of **no change in the hazard**

Experts are invited to reflect on whether their plausible ranges exceeded either interval limit

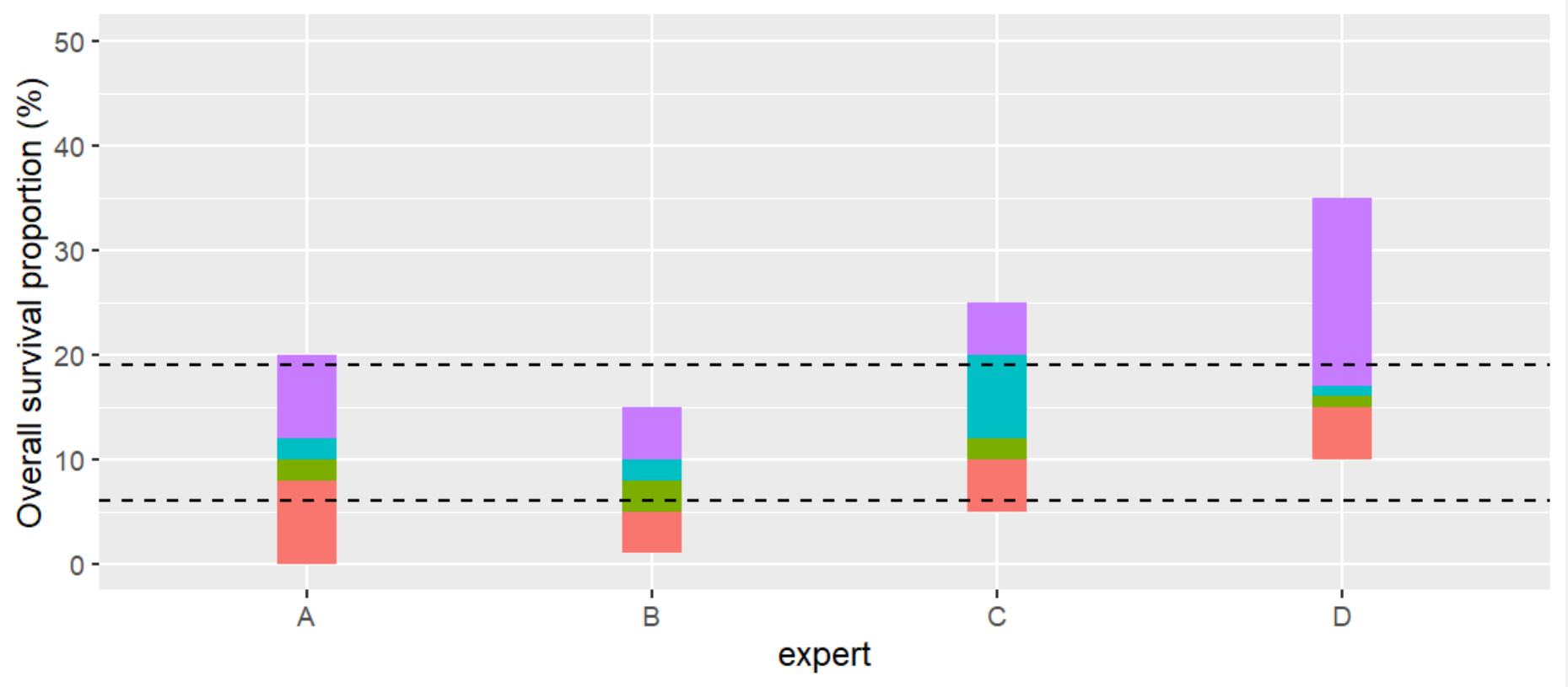
Provide a starting point for the group discussion

Hypothetical example: generation of scenario



- An exponential model was fitted to month 100-150 survival data and extrapolated to month 250
- Hazard remained unchanged from month 100 to 250

Hypothetical example: individual judgements and scenario testing



Scenario testing

Interaction between experts

Expert engagement

- Participation in a workshop ensures commitment to the exercise, and deliberations can be recorded

Expert understanding

- Training will be needed and is typically easier to conduct in person (F2F/online) rather than through self-directed learning

Validating expert judgements

- Each expert will need to provide justification for their view(s) to an audience of their peers

Bias mitigation in group discussion

1. **Training**: discuss risk of bias from an unduly influential expert
2. **Individual judgements**: obtain and document judgements independently from each expert before any interactions between the experts
3. **Facilitation**: carefully manage the discussion to ensure all experts are contributing their view(s)
4. **Final distribution**: compared the final distribution against the initial individual judgements, and check the rationale

Aggregation methods

Behavioural

Mathematical

	Behavioural	Mathematical
Definition	A single distribution that reflects the experts' consensus beliefs based on group discussion • E.g. rational impartial observer (RIO)	Experts' judgements are combined using a mathematical rule • E.g. Linear opinion pooling
Positive features	✓ Peer scrutiny ✓ Scientific debate and enquiry	✓ Easy to calculate ✓ Useful when • Not much disagreement • Different schools of thought and no obvious reason to favour one over another
Negative features	Careful facilitation is needed to mitigate potential bias (previous slide) Training 1 Individual judgement 2 Facilitation 3 Final distribution 4	May lead to inappropriate final aggregated judgment • Individual judgements without discussion: heuristic and bias • Second round of individual judgements: biased by the willingness of changing judgments

Reporting of a SEE for long-term survival outcomes

Selection of experts

Selection process

Declaration of conflicts of interest

Expertise
- Experience in disease area and treatments
- Number of patients treated

Evidence dossier

Content

Compilation

Review process

Workshop

Training: content and format

SEE format

Individual judgements and justification

Group discussion

Aggregation method and result

Expert comments on hazard trend

Summary

1. Relationship between survivor and hazard function
    - Elicited values should have internal consistency between the qualitative knowledge and quantitative estimate
    - Scenario testing
    - Qualitative discussion

2. Interaction between experts
    - Ensure commitment
    - Help expert understanding
    - Validate judgements

3. Accurate and thorough reporting of the quantitative and qualitative aspects of the SEE ensures maximum usefulness of the workshop outputs

Plan in advance!

References

- Oakley JE, O'Hagan A. SHELF: the Sheffield Elicitation Framework (version 4). https://shelf.sites.sheffield.ac.uk/
- National Institute for H, Care E. Tebentafusp for treating advanced (unresectable or metastatic) uveal melanoma. Technology Appraisal ID1441. Available at https://www.nice.org.uk/guidance/indevelopment/gid-ta10428

Q&A Session

Do you have any questions for the panel members?

Closing remarks

Issue

A persistent challenge in economic modelling for HTA: estimating long-term survival

What have we done?

Developed a bespoke protocol with software on SEE for long-term survival to improve credibility and reduces bias

Take-home messages

- Consider hazard
- Interaction between experts
- Accurate and thorough reporting: quantitative and qualitative
- **Plan in advance**

Coming soon… NICE DSU TSD