Bespoke structured expert elicitation protocol for long-term survival outcomes

NICE DSU TSD 26

Kate Ren

Sheffield Centre for Health and Related Research (SCHARR)

University of Sheffield

09.06.2025

Disclaimer

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

- Employed (part-time) by the University of Sheffield

- Employed (part-time) by ConnectHEOR

- Member of NICE Technology Appraisal Committee C

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

1 **Critical role in determining cost-effectiveness**
- Extrapolated survival outcomes can have large variability resulting in uncertain incremental cost-effectiveness ratios
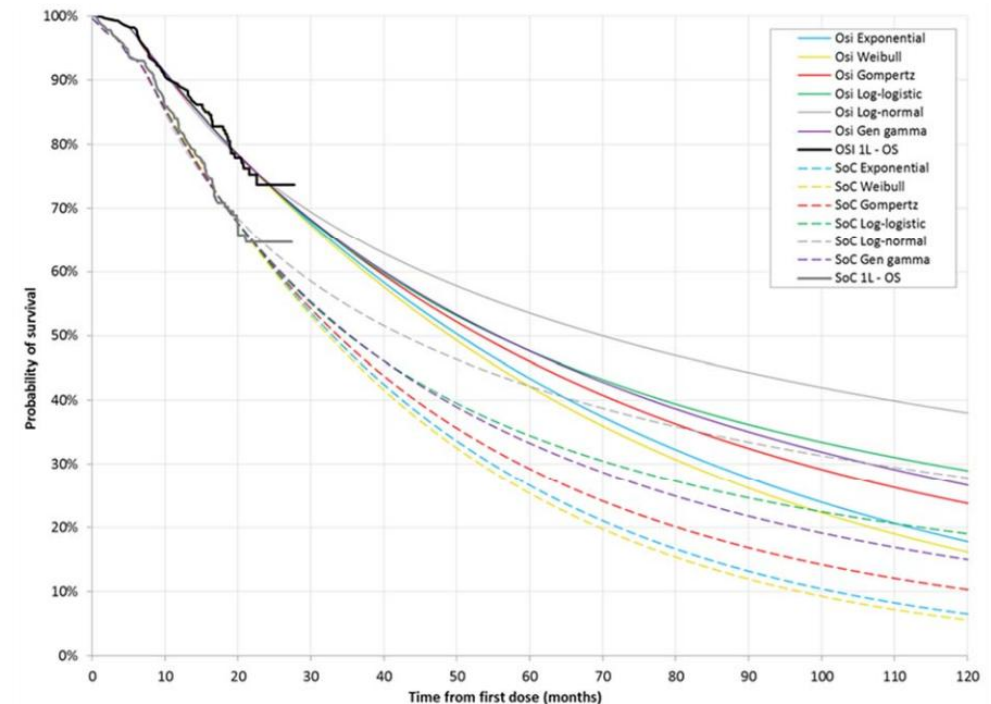
2 **Uniqueness of time-to-event data**
- Censoring
- Underlying hazard function

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

Source: https://www.nice.org.uk/guidance/ta621

3

NICE TSD 26: Expert Elicitation for Long-Term Survival Outcomes
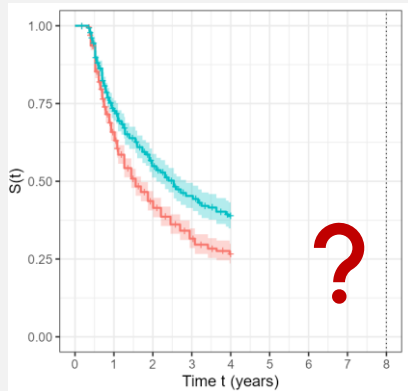
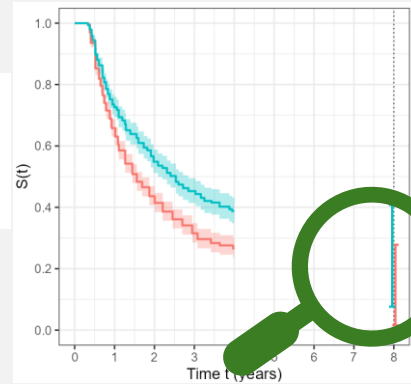**TECHNICAL SUPPORT DOCUMENT 26:** EXPERT ELICITATION FOR LONG-TERM SURVIVAL OUTCOMES

Structured Expert Elicitation

TECHNICAL SUPPORT DOCUMENT 14: SURVIVAL ANALYSIS FOR ECONOMIC EVALUATIONS ALONGSIDE CLINICAL TRIALS-EXTRAPOLATION WITH PATIENT LEVEL DATA

+

TECHNICAL SUPPORT DOCUMENT 21: FLEXIBLE METHODS FOR SURVIVAL ANALYSIS

OUTPUTS
- **Probability distribution** quantifying expert uncertainty
- **Qualitative opinion** about how the hazard may change

HOW

- **Bespoke protocol** integrating survival and hazard judgements

- **Evidence dossier** summarising relevant information to support the experts when making judgements
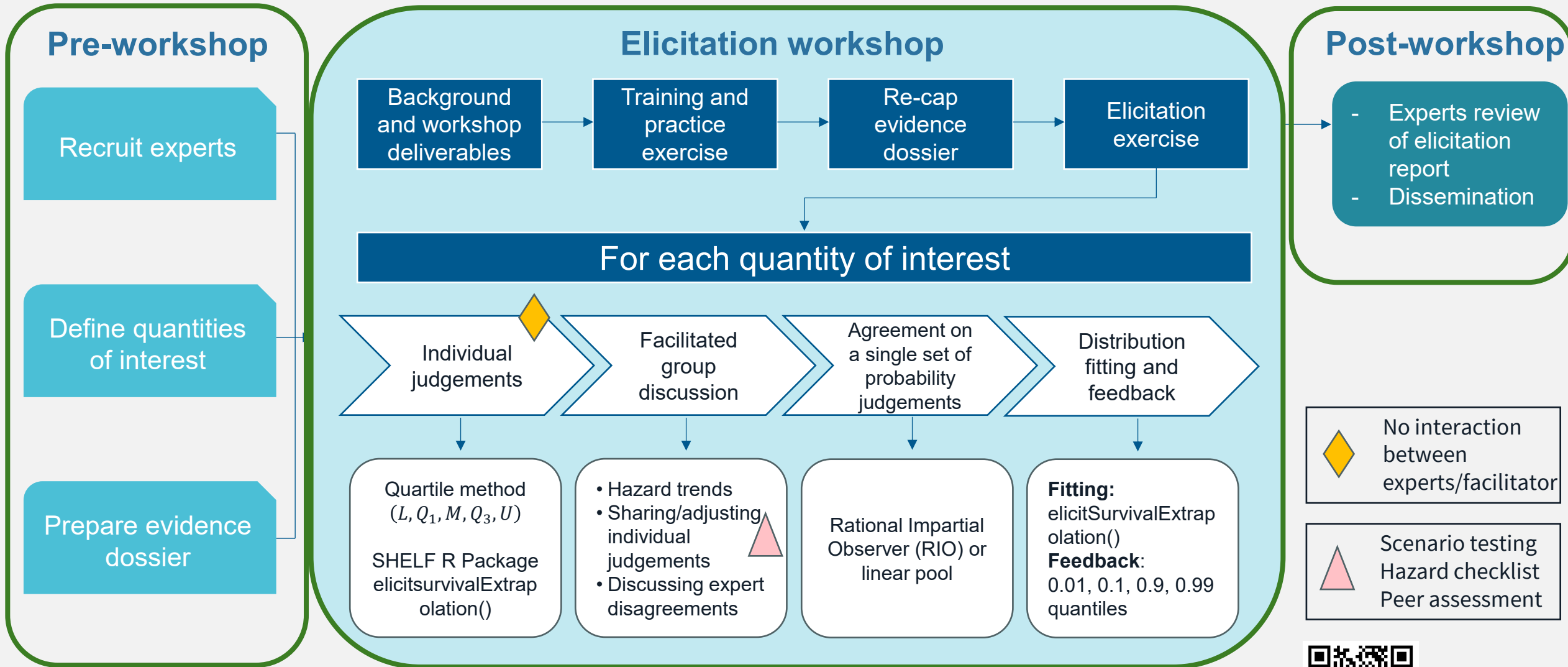
- **Training** of experts in making probability judgements; survivor and hazard functions; practice elicitation exercise

- **Panel** of experts, individual elicitation; qualitative and quantitative assessments; scenario testing; aggregation

- **Reporting** for full transparency, including the expert recruitment process; hazard discussion; individual and aggregated judgements

NICE TSD 26: bespoke SEE protocol adapted from SHELF

Pre-workshop

- Recruit experts
- Define quantities of interest
- Prepare evidence dossier

Elicitation workshop

Background and workshop deliverables → Training and practice exercise → Re-cap evidence dossier → Elicitation exercise

For each quantity of interest

Individual judgements → Facilitated group discussion → Agreement on a single set of probability judgements → Distribution fitting and feedback

Quartile method (L, Q_1, M, Q_3, U)

SHELF R Package elicitsurvivalExtrapolation()

- Hazard trends
- Sharing/adjusting individual judgements
- Discussing expert disagreements

Rational Impartial Observer (RIO) or linear pool

Fitting: elicitSurvivalExtrapolation()
Feedback: 0.01, 0.1, 0.9, 0.99 quantiles

Post-workshop

- Experts review of elicitation report
- Dissemination

◆ No interaction between experts/facilitator

▲ Scenario testing Hazard checklist Peer assessment

Software and training material: https://shelf.sites.sheffield.ac.uk/survival-extrapolation

NICE STA ID1441: tebentafusp for advanced uveal melanoma

❶ Two appraisal committee meetings

❷ Tebentafusp was not recommended

❸ Appeal hearing with upheld points

"*it is the panel's view that reasonableness requires they should **seek additional expert clinical input** on areas of important residual uncertainty, notably the most appropriate choice, and interpretation of **survival curve models** to interrogate the available data, and the most appropriate means of allocating supportive care costs in the model.*"
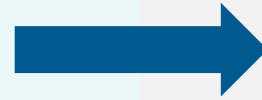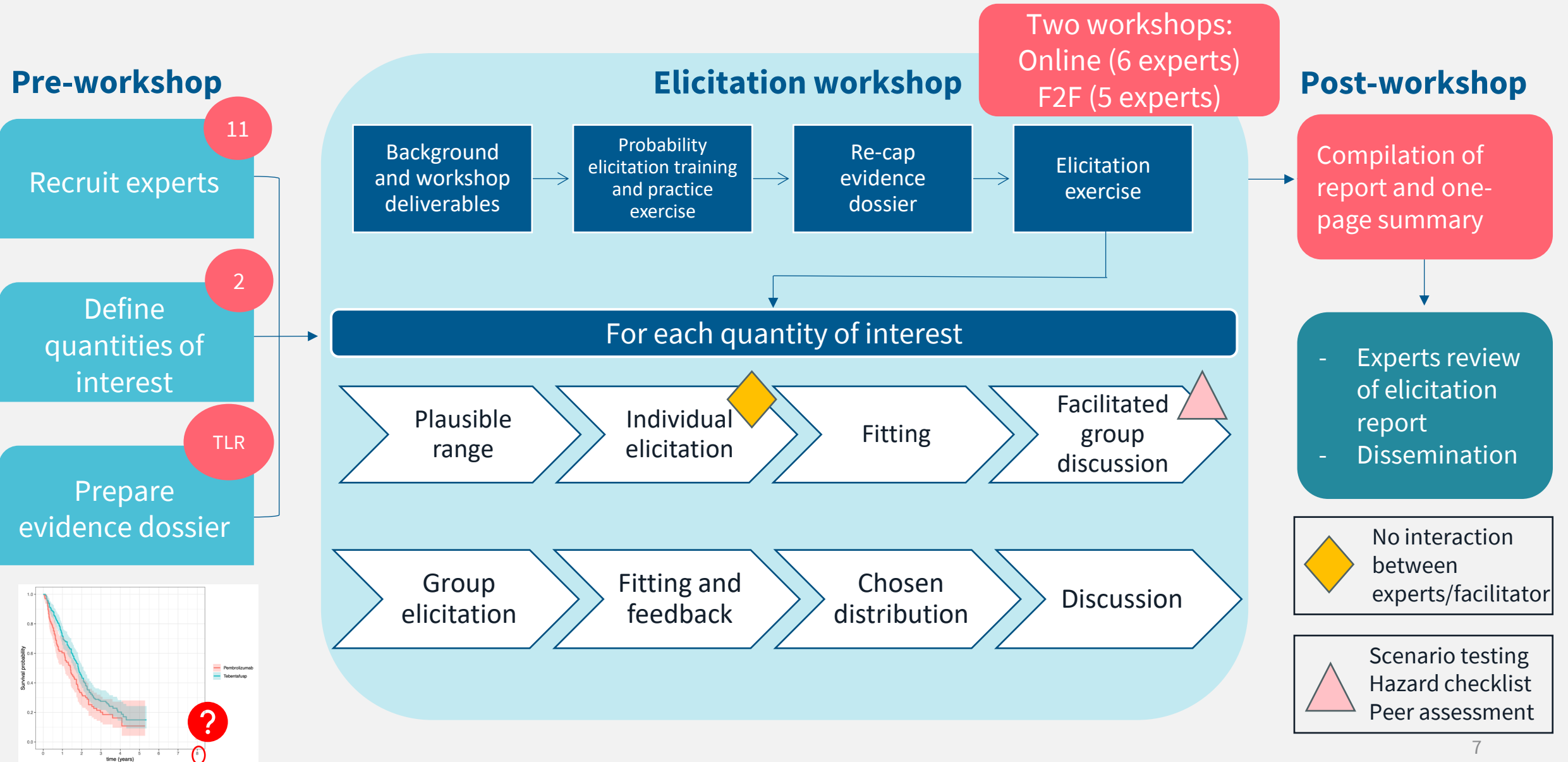
NICE DSU was instructed to

⊕ Use a structured approach to elicit expert estimates of the expected survival of people with uveal melanoma treated with pembrolizumab and those treated with tebentafusp and the uncertainty around these estimates

⊕ Also consider how the learnings from this exercise might inform the use of expert elicitation in future appraisals

6

Abbreviations: STA: single technology assessment; DSU: Decision Support Unit

NICE STA ID1441: structured expert elicitation

Pre-workshop

Elicitation workshop

Two workshops:
Online (6 experts)
F2F (5 experts)

Post-workshop

11

Recruit experts

2

Define quantities of interest

TLR

Prepare evidence dossier

?

Background and workshop deliverables → Probability elicitation training and practice exercise → Re-cap evidence dossier → Elicitation exercise

Compilation of report and one-page summary

For each quantity of interest

Plausible range → Individual elicitation ◆ → Fitting → Facilitated group discussion ▲

- Experts review of elicitation report
- Dissemination

Group elicitation → Fitting and feedback → Chosen distribution → Discussion

◆ No interaction between experts/facilitator

▲ Scenario testing Hazard checklist Peer assessment

Abbreviations: SEE, structured expert elicitation; F2F, face-to-face; TLR, targeted literature review

Obtaining individual expert judgements

- Guidance: SHELF protocol
- Quartile method (variable interval approach)
  - Lower and upper plausible limits
  - Median
  - Lower quartile (Q1)
  - Upper quartile (Q3)

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

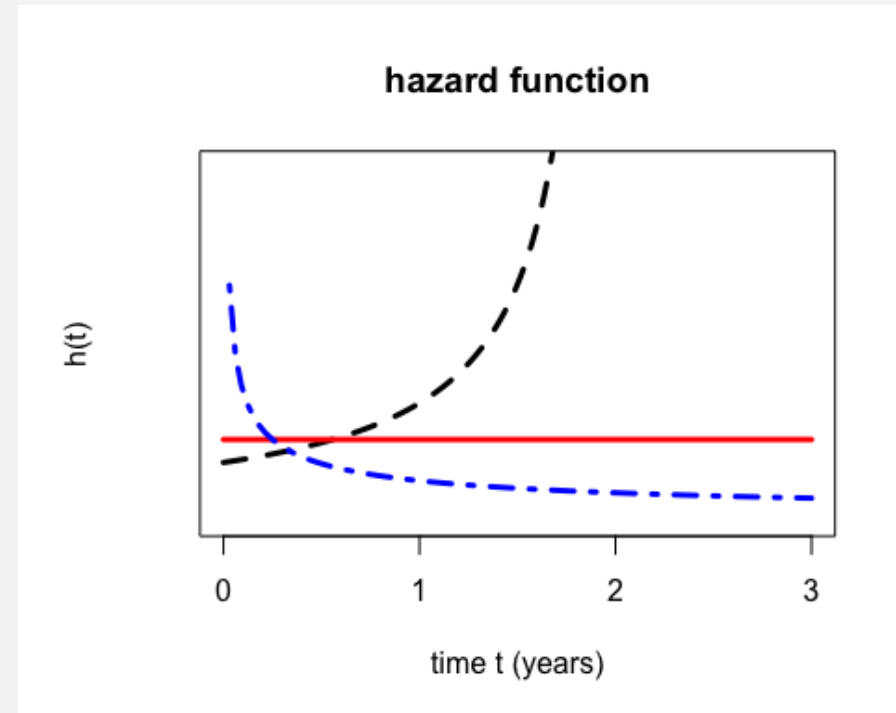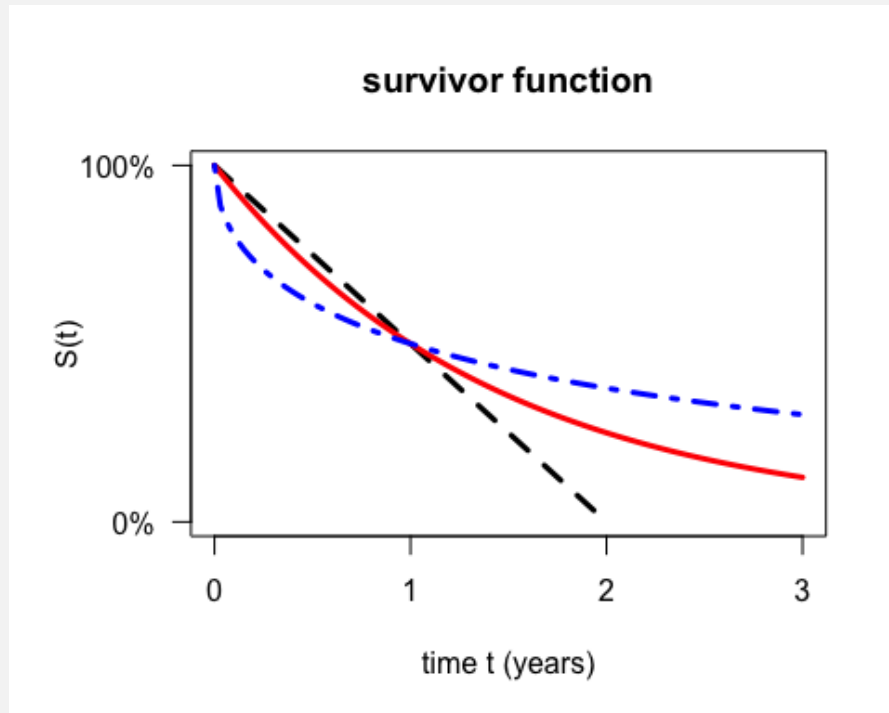**Individual judgements**

Tebentafusp

Pembrolizumab

Abbreviations: SHELF, Sheffield Elicitation Framework; Q1, lower quartile; Q3, upper quartile
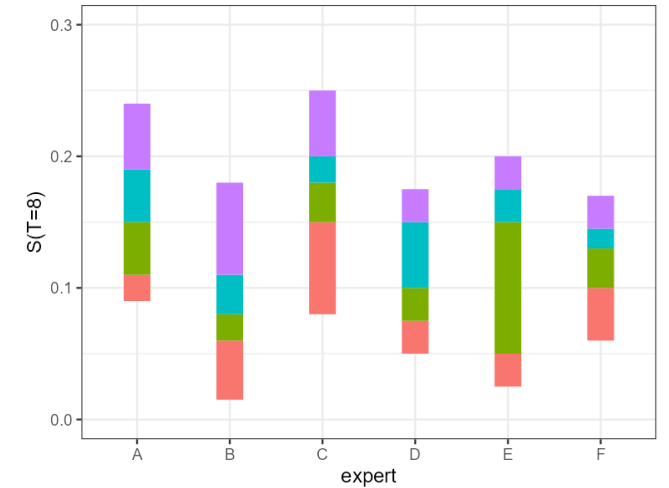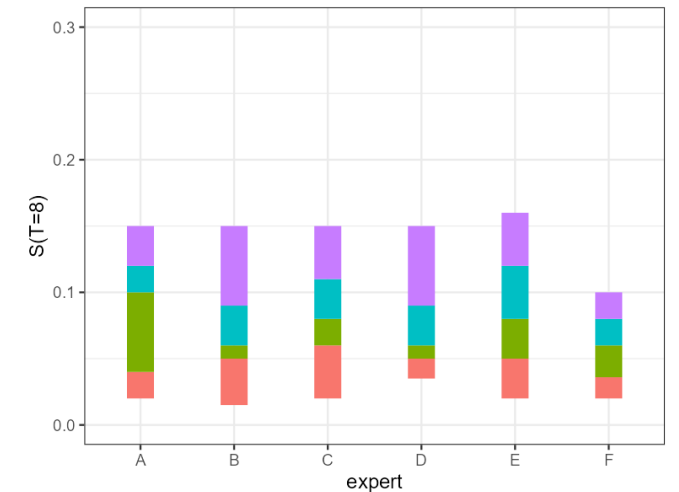
The hazard function

- Related to 'instantaneous risk of death' at time t for a patient who has survived up to time t

- Corresponds to trend in survivor function

Group discussion: 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

Aim: to give a point of reference for reflection on expert's individual judgements

Note: no claims were made by the elicitation team regarding the probability of the scenario being true.

Procedure

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

An approx. 95% credible interval reported 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
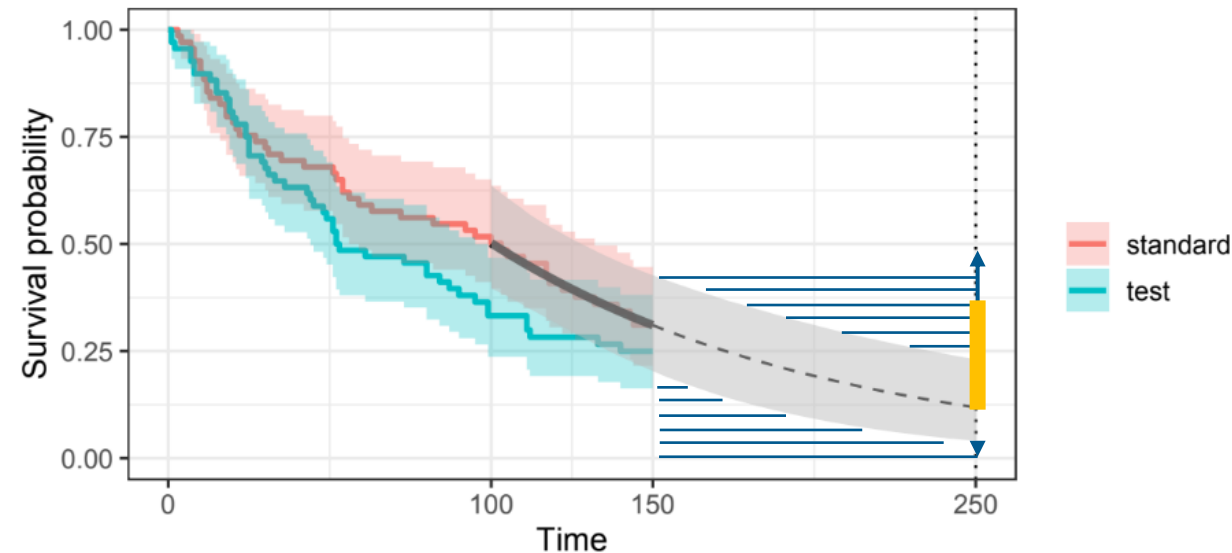
⟩ An exponential model was fitted to month 100-150 survival data and extrapolated to month 250

⟩ Scenario where hazard remains unchanged from month 100 to 250

Individual judgements and scenario testing: tebentafusp arm

- Experts provide upper and lower plausible limits, median and upper and lower quartiles
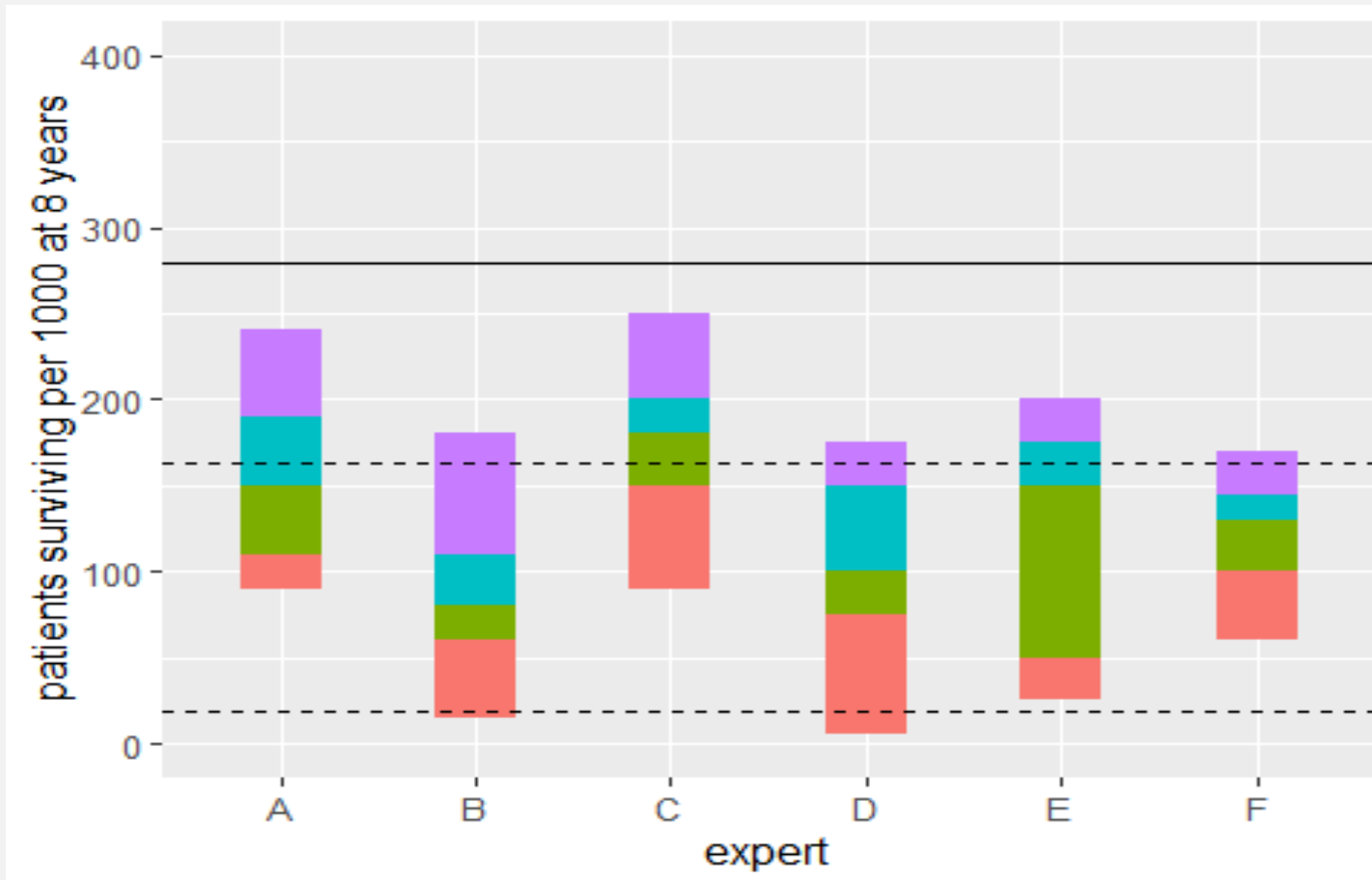
Upper Kaplan-Meier 97.5% confidence limit for survival at 4 years

Scenario testing
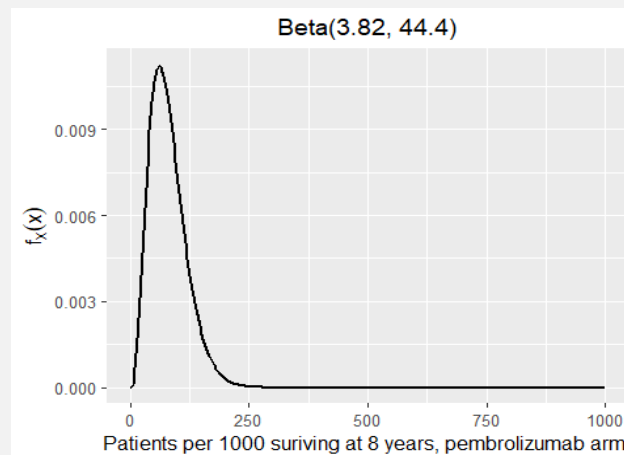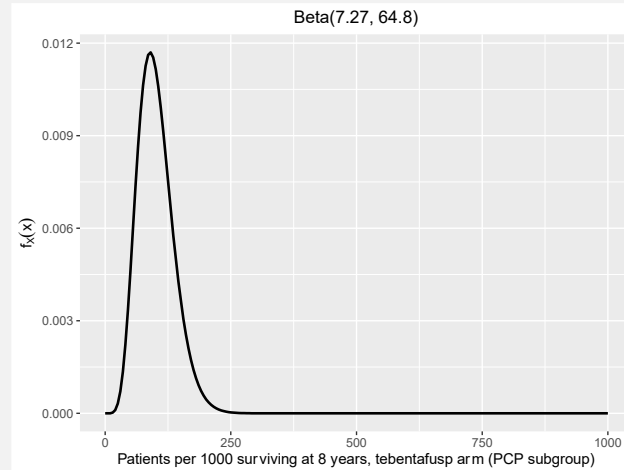
Elicited distributions

Tebentafusp

- 10% probability that survival is less than 50 per 1000
- 50% probability that survival is greater than 100 per 1000
- 5% probability that survival is greater than 150 per 1000

Pembrolizumab

As there was little disagreement between the experts, the Facilitator proposed assuming initial RIO judgements to be the median and quartiles computed as averages of the experts' individual judgements

Fitted distribution for each arm

Beta(7.27, 64.8)

Beta(3.82, 44.4)

Fitted distribution for each arm

Abbreviations: RIO, Rational Impartial Observer

13
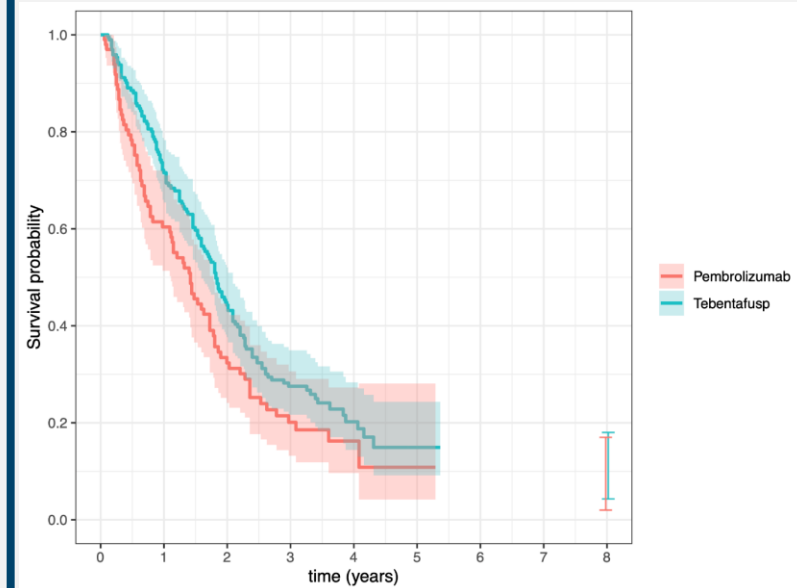
Our observations from NICE STA ID1441 case study

Expert recruitment
- Stakeholder involvement
- Initial email with motivations
- Follow-ups

Workshop format
- Online is as effective as F2F

Experts' normative skills
- Interactive training
- Include training on survival and hazard
- Practice exercise

Resources required
- 3 months
- Evidence dossier and expert recruitment preparation most time consuming

Quantity of interest
- Two quantities
- Precise definition
- Health economist input

Constraints of HCDM
- Overwhelmingly positive response from the clinical community
- Travel and accommodation were reimbursed

Evidence dossier
- Literature review
- Expert review prior to workshop

Heuristics and biases
- Overconfidence, anchoring and availability biases managed via facilitation

14

Abbreviations: F2F, face-to-face; HCDM, healthcare decision making

Summary and key recommendations from TSD 26

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 elicitation ensures maximum usefulness of the workshop outputs

Plan in advance!

References

- Oakley J. E., Ren S., Forsyth J. E., Gosling J. P., Wilson K., Latimer N., Rutherford M. J., Uttley L., Fotheringham J., NICE DSU Technical Support Document 26: Expert elicitation for long-term survival outcomes. 2025. Available from http://www.nicedsu.org.uk

- Oakley JE, O'Hagan A. SHELF: the Sheffield Elicitation Framework (version 4). https://shelf.sites.sheffield.ac.uk

- Software and training material for eliciting long-term survival: https://shelf.sites.sheffield.ac.uk/survival-extrapolation

- NICE. Tebentafusp for treating advanced (unresectable or metastatic) uveal melanoma. Technology Appraisal ID1441. Available at https://www.nice.org.uk/guidance/indevelopment/gid-ta10428, 2024