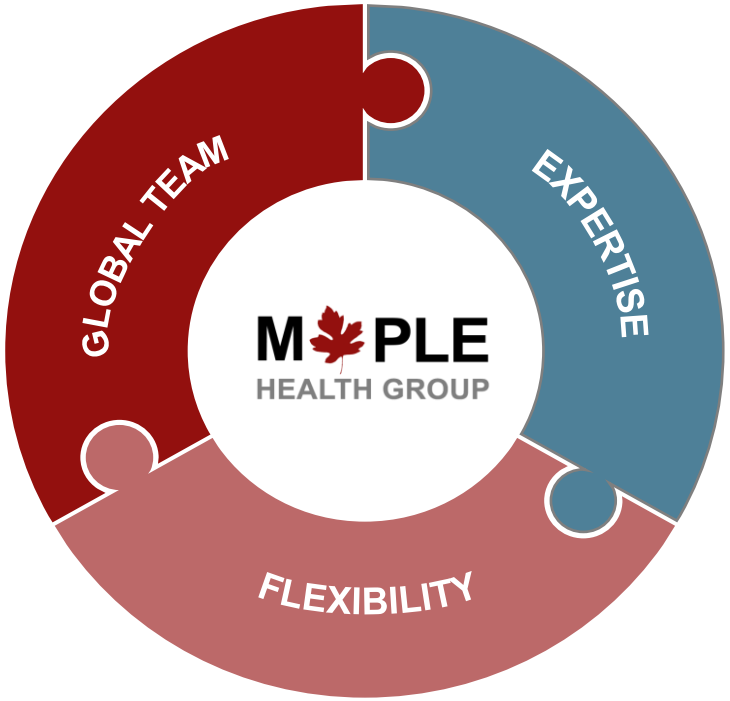
Structured Expert Elicitation for Healthcare Decision Making

2 October 2025

James Horscroft, Maple Health Group

Maple offers a strong strategic partnership underpinned by the depth and breadth of our expertise, our perfectionism and dedication.

Global team
In-house consultants in locations spanning 8 time-zones, offering expert knowledge of all major markets and time zone flexibility



Expertise
In-depth expertise across all major markets, covering 100+ engagements in the last 2 years

Flexibility in approach
Tailored project execution approach to each individual client and asset based on your requirements and our in-depth understanding of the healthcare market

- **Comprehensive** market access and HEOR **capabilities**
- **Commitment** to **exceptional quality** and striving for **continuous improvement**
- Close **integration** among teams working on different workstreams and with the client team
- **Streamlined project management** and effective communication between our clients and Maple via a dedicated point of contact

Conducting SEE to support a ZINL pharmacoeconomic dossier presents several challenges to manufacturers.

Cost

- SEE is a complex method requiring expert facilitators
- The Netherlands is a small market

Are the costs justified?

Timing

- SEE must come after data availability and model development, yet can be time consuming

Is the impact on timelines justified?

Experts

- SEE requires multiple experts and a high level of engagement
- There is often scepticism from the medical community

Is the investment of expert time justified?

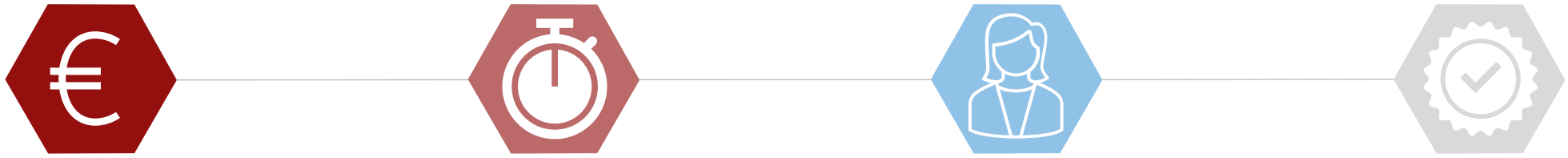
Trust

- Experience from the UK suggests a mixed level of acceptance
- The extent to which SEE will inform decision making in NL is yet to be seen

Can trust be established between manufacturers and decision makers?

SEE = structured expert elicitation; ZINL = Zorginstituut Nederland

SEE should be reserved only for parameters where the existing uncertainty has a substantial impact on cost effectiveness outcomes.



Cost

Are the costs justified?

Timing

Is the impact on timelines justified?

Experts

Is the investment of expert time justified?

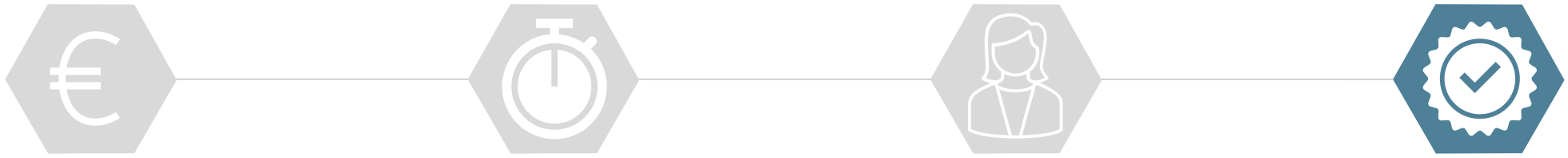
Trust

YES, if…

1. The existing uncertainty surrounding the parameters of interest substantially impact cost effectiveness outcomes
2. There is no feasible means of obtaining a higher level of evidence (e.g. RWE)
3. A pragmatic approach to SEE is taken, which minimizes time and cost, but remains sufficiently robust

SEE = structured expert elicitation; ZINL = Zorginstituut Nederland

Can trust on both sides be established through training, mutual understanding, and growing familiarity?



Can trust be established between manufacturers and decision makers?

- Could further **training** on SEE for both manufacturers and decision makers could help to improve the quality of SEE studies conducted, and the ability of decision makers to interpret them?
- Could further guidance on **pragmatic approaches** to SEE make the investment worthwhile for manufacturers and ensure that decision makers' expectations are realistic?
- Will growing **familiarity** over time naturally lead to growing trust?

SEE = structured expert elicitation

Agenda

1. Introduction to SEE
2. Case study and interactive exercise
3. Key learnings from the UK

Agenda

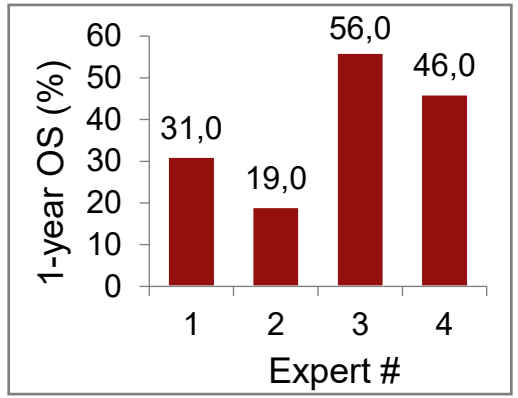
1. **Introduction to SEE**
2. Case study and interactive exercise
3. Key learnings from the UK

Structured expert elicitation obtains beliefs from experts for uncertain quantities as probability distributions in the face of uncertainty.

Traditional elicitation: example output

*"The mean estimate (range) of **1-year survival** of patients with metastatic colorectal cancer receiving best supportive care in the UK from experienced clinicians was **38% (19–56%)**."*

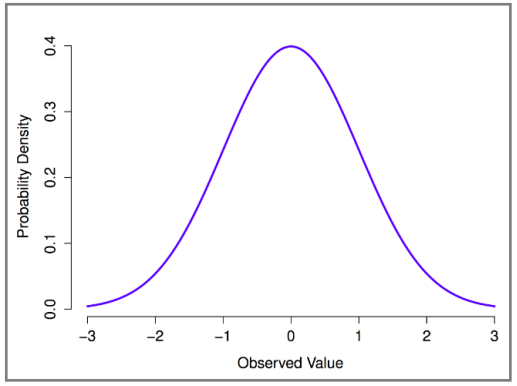
- The range tells us about the **variability** between experts' judgements, but nothing about their subjective uncertainty
- Subject to significant **bias** associated with obtaining judgements from experts



SEE: example output

*"Experienced clinicians believe there is a **99% probability** that the **1-year survival** of patients with metastatic colorectal cancer receiving best supportive care in the UK is **19–56%**, with a median estimate of **38%**."*

- Probability distribution representing the **subjective uncertainty** surrounding the quantity of interest
- Known **biases** associated with obtaining judgements from experts minimized through an intensive elicitation process



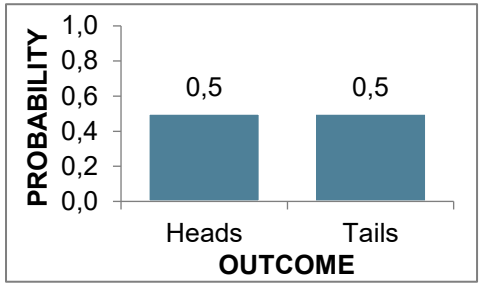
Note: SEE cannot replace empirical evidence, but is often the most robust approach where collecting empirical evidence is unfeasible.

SEE = structured expert elicitation

Probability distributions in SEE represent subjective uncertainty, not variability.

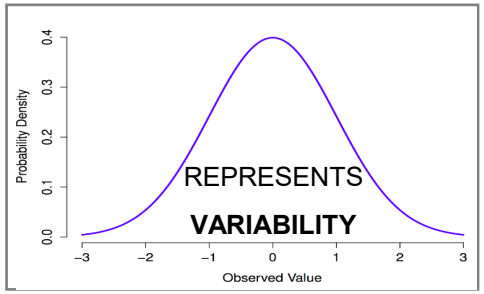
A COIN TOSS

Frequency probability based on **discrete data**



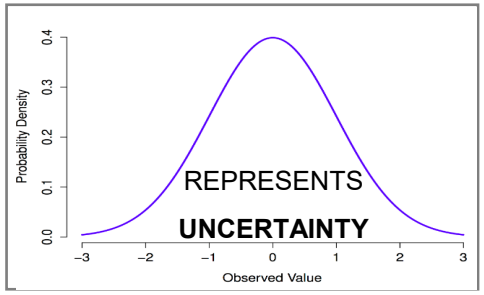
COMMUTE TIME

Frequency probability based on **continuous data**

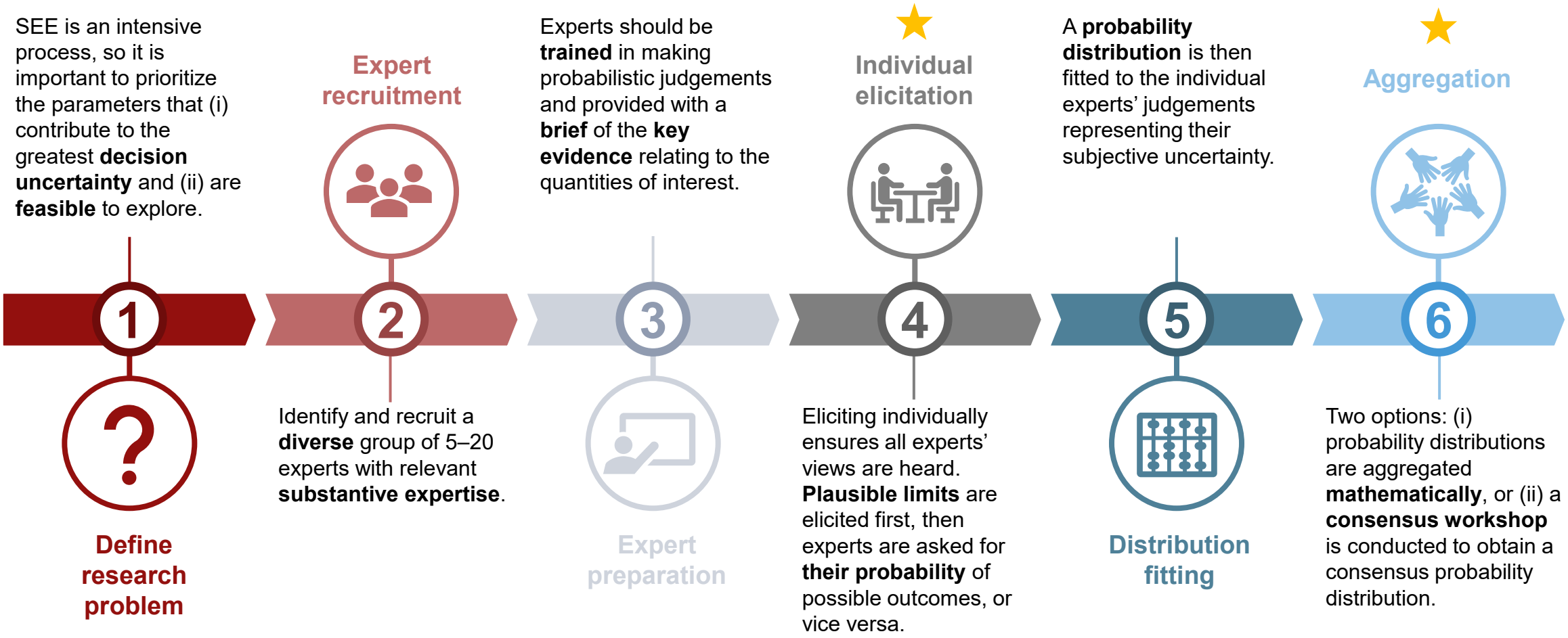


WINNING TIME IN THE 2024 PARIS OLYMPICS MEN'S 100m FINAL

Subjective probability based on **continuous data** and **prior beliefs**



There are several different frameworks for SEE, all of which follow this same broad principles.

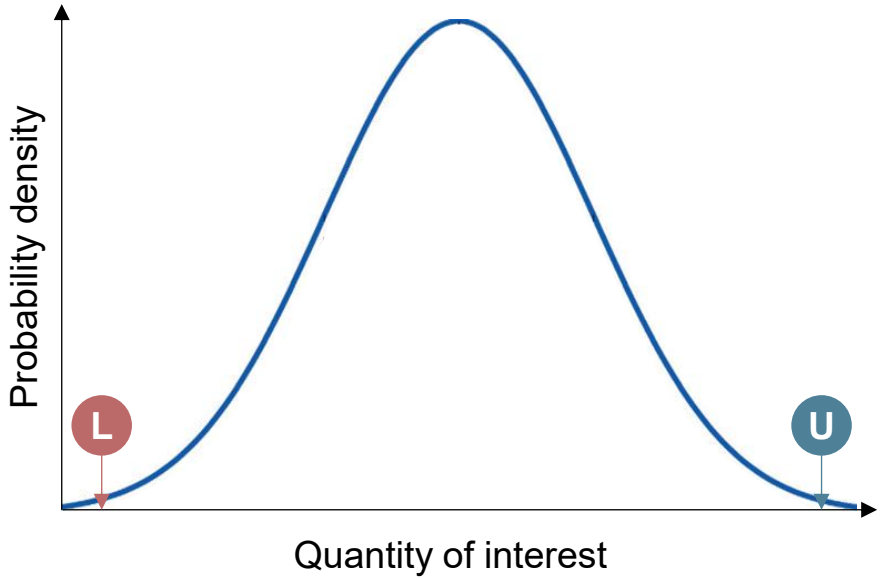


★ There are key differences between frameworks at the **individual elicitation** and **aggregation** steps.

SEE = structured expert elicitation

All encoding methods start by eliciting plausible limits.

LOWER PLAUSIBLE LIMIT (L)	A value such that you believe that there is a **1% probability** that the true value is less than L
UPPER PLAUSIBLE LIMIT (U)	A value such that you believe that there is a **1% probability** that the true value is greater than U

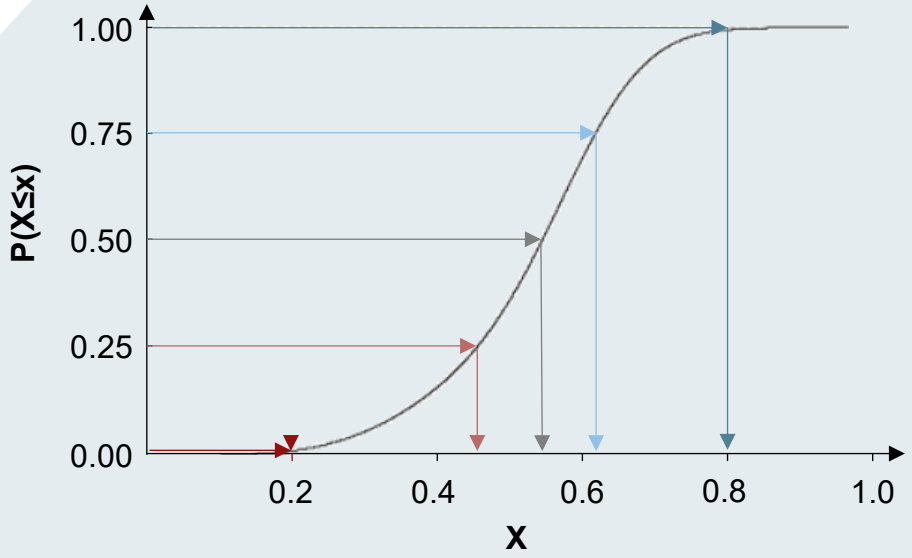


Why start with plausible limits?

- Minimizes the bias of **over-confidence** – not assigning enough probability to extreme values
- Minimizes the bias of **anchoring** – assigned too much credibility to a value already in the expert's mind
- Provides **bounds** to the uncertainty

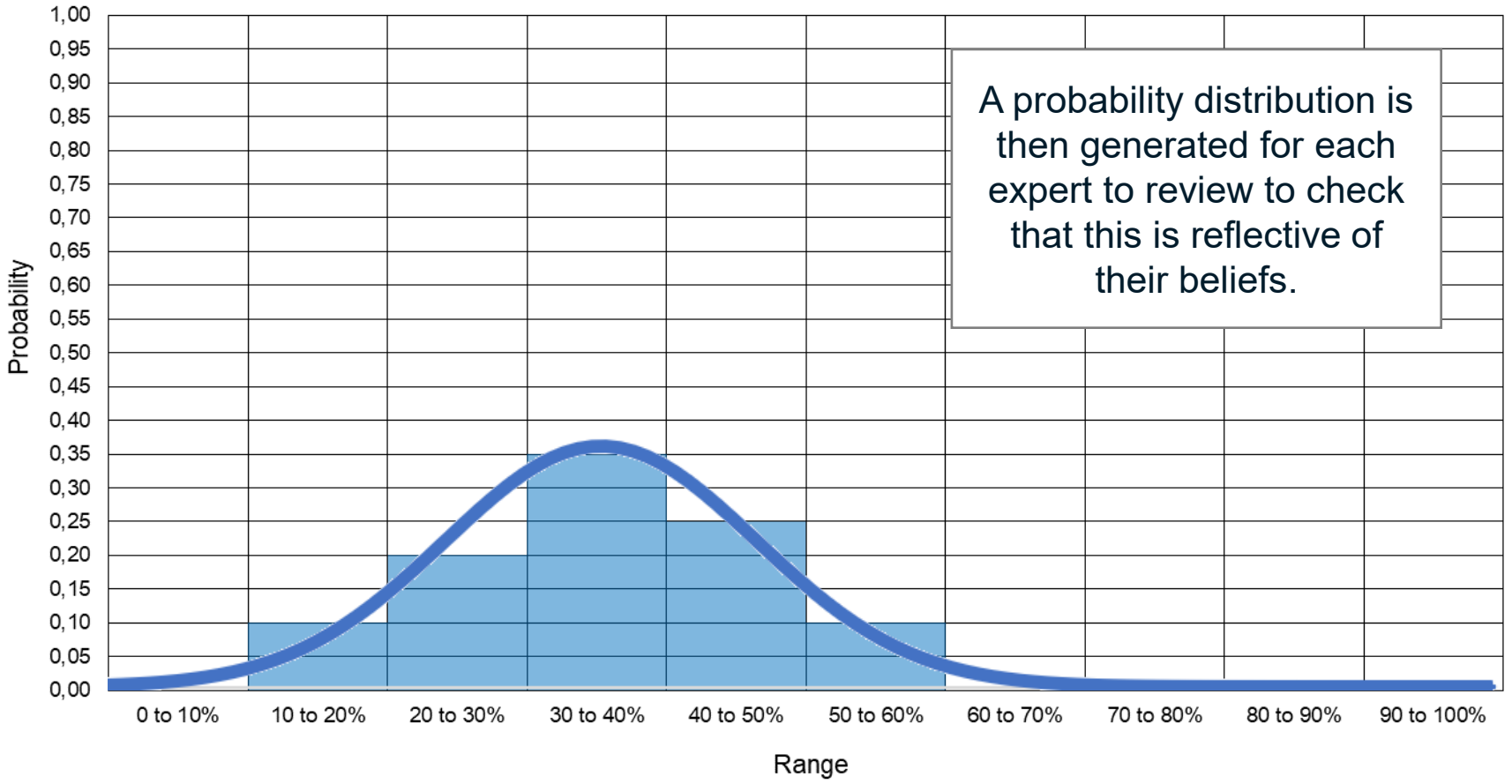
The Quartiles method is a commonly used example of a Variable Intervals Method.

LOWER PLAUSIBLE LIMIT (L)	A value such that you believe that there is a **1% probability** that the true value is less than L	0.20
UPPER PLAUSIBLE LIMIT (U)	A value such that you believe that there is a **1% probability** that the true value is greater than U	0.80
MEDIAN (M)	A value such that the true value is **equally likely** to be less than or greater than M	0.55
LOWER QUARTILE (Q_1)	A value such that you believe that the true value is **equally likely** to be less than Q_1 or between Q_1 and M	0.45
UPPER QUARTILE (Q_3)	A value such that you believe that the true value is **equally likely** to be greater than Q_3 or between M and Q_3	0.62



The Chips and Bins method is a commonly used Fixed Intervals Method.

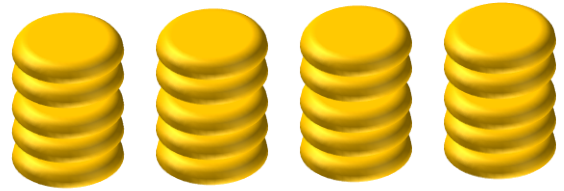
10 ‘bins’, each representing a range of possible values of the quantity



A probability distribution is then generated for each expert to review to check that this is reflective of their beliefs.

Experts are asked to place the chips in the bins to specify your knowledge and beliefs.

20 ‘chips’, each representing a probability of 0.05 (5%)



SEE is increasingly being recognized as the go-to method for addressing uncertainty where empirical evidence is lacking in healthcare decision-making.

Decision uncertainty is increasing

- RWE
- Accelerated approvals
- Orphan drugs

SEE explores and characterizes uncertainty

- Bounds uncertainty
- Informs the narrative

SEE adds robustness to assumptions

- Avoids unstructured interviews and ad boards
- Minimizes biases

SEE has been accepted at HTA

- NICE, CADTH and ZINL
- Preferred method where empirical evidence is lacking

KOL = key opinion leader; RWE = real-world evidence; SEE = structured expert elicitation

Agenda

1. Introduction to SEE
2. **Case study and interactive exercise**
3. Key learnings from the UK

In 2024, we conducted an SEE study to explore the uncertainty surrounding long-term overall survival in the pivotal trial for a new oncology intervention.

OBJECTIVE

To characterize **expert uncertainty** surrounding the **long-term overall survival (OS)** of patients in the pivotal trial, receiving either the **intervention** or **comparator**.

PARAMETERS

OS rate in the **first five years** (Years 0–5) and **second five years** (Years 5–10) of treatment for patients in the **ITT** population of the pivotal trial for patients receiving either the **intervention** or **comparator**

PURPOSE

- HTA agencies need to come to a judgement on a **relative effectiveness estimation** for an intervention vs. key comparators.
- To achieve this, there is a need to **extrapolate** OS from the clinical trial, using the most robust methods available.
- In the absence of empirical evidence, **structured expert elicitation (SEE)** can be used to characterize the uncertainty, and minimize biases associated with obtaining expert judgements.

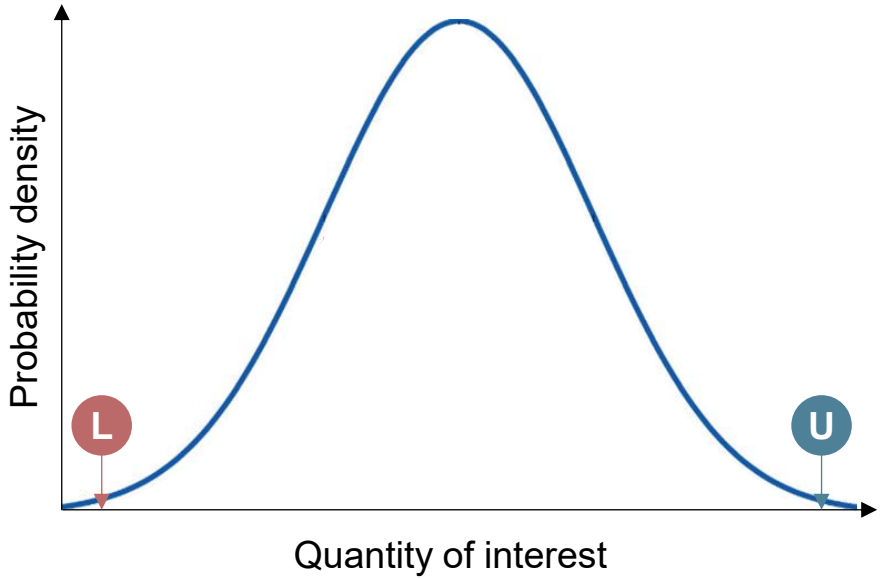
HOW WERE THE RESULTS USED?

The results were **combined** with the existing survival data from the clinical trial to inform **plausible survival extrapolations** in preparation for local HTA appraisals.

Abbreviations: HTA = Health Technology Assessment; ITT = intention-to-treat; OS = overall survival; PFS = progression-free survival; SEE = structured expert elicitation.

Experts were first asked for their plausible limits for each quantity of interest.

LOWER PLAUSIBLE LIMIT (L)	A value such that you believe that there is a **1% probability** that the true value is less than L
UPPER PLAUSIBLE LIMIT (U)	A value such that you believe that there is a **1% probability** that the true value is greater than U



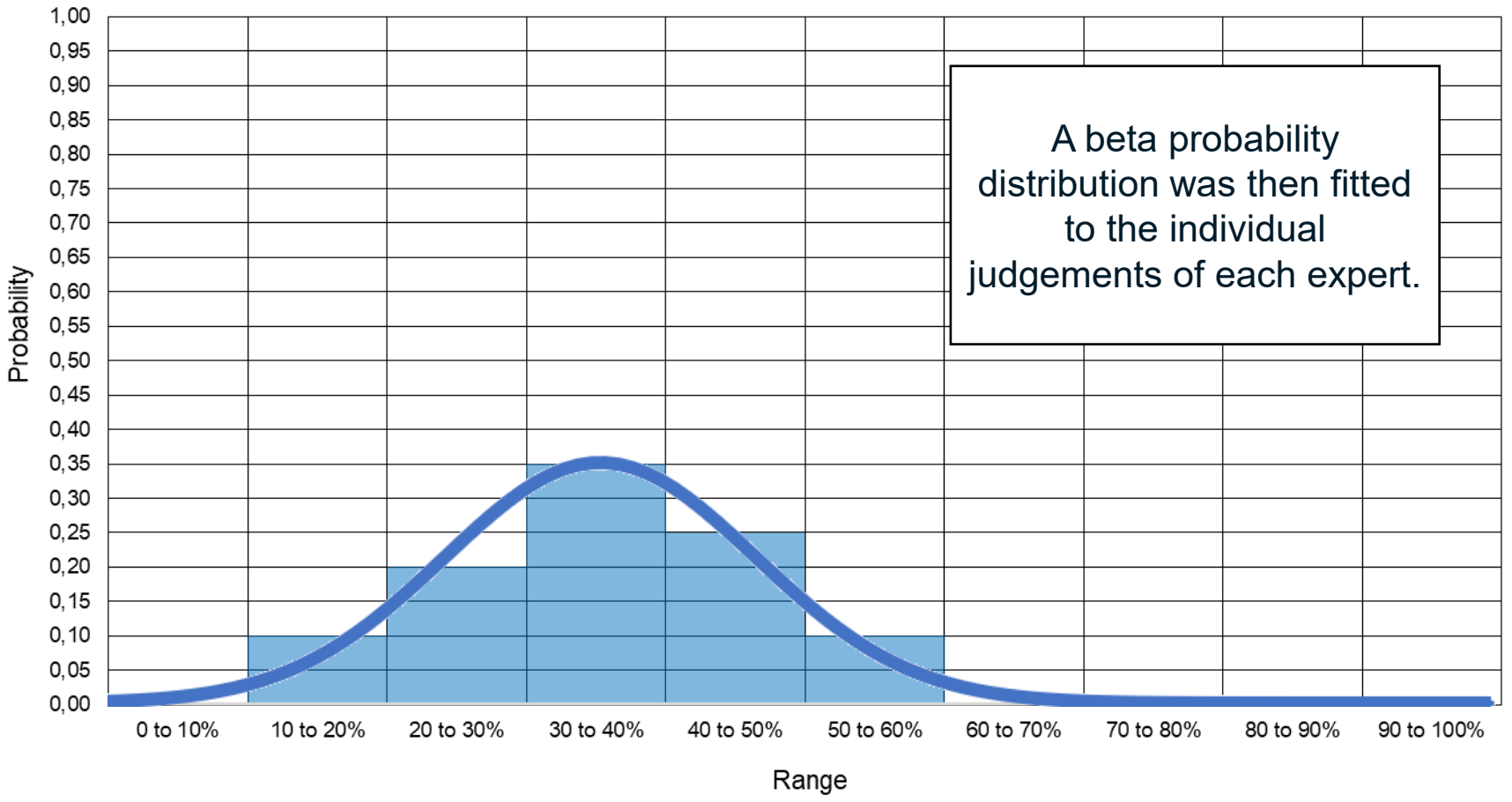
Why start with plausible limits?

- Minimizes the bias of **over-confidence** – not assigning enough probability to extreme values
- Minimizes the bias of **anchoring** – assigned too much credibility to a value already in the expert's mind
- Provides **bounds** to the uncertainty

Abbreviations: L = lower plausible limit; U = upper plausible limit

They were then asked to complete a ‘chips and bins’ exercise for each quantity of interest.

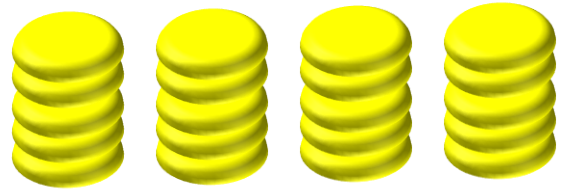
10 ‘bins’, each representing a range of possible values of the quantity



A beta probability distribution was then fitted to the individual judgements of each expert.

Experts were asked to place the chips in the bins to specify your knowledge and beliefs

20 ‘chips’, each representing a probability of 0.05 (5%)

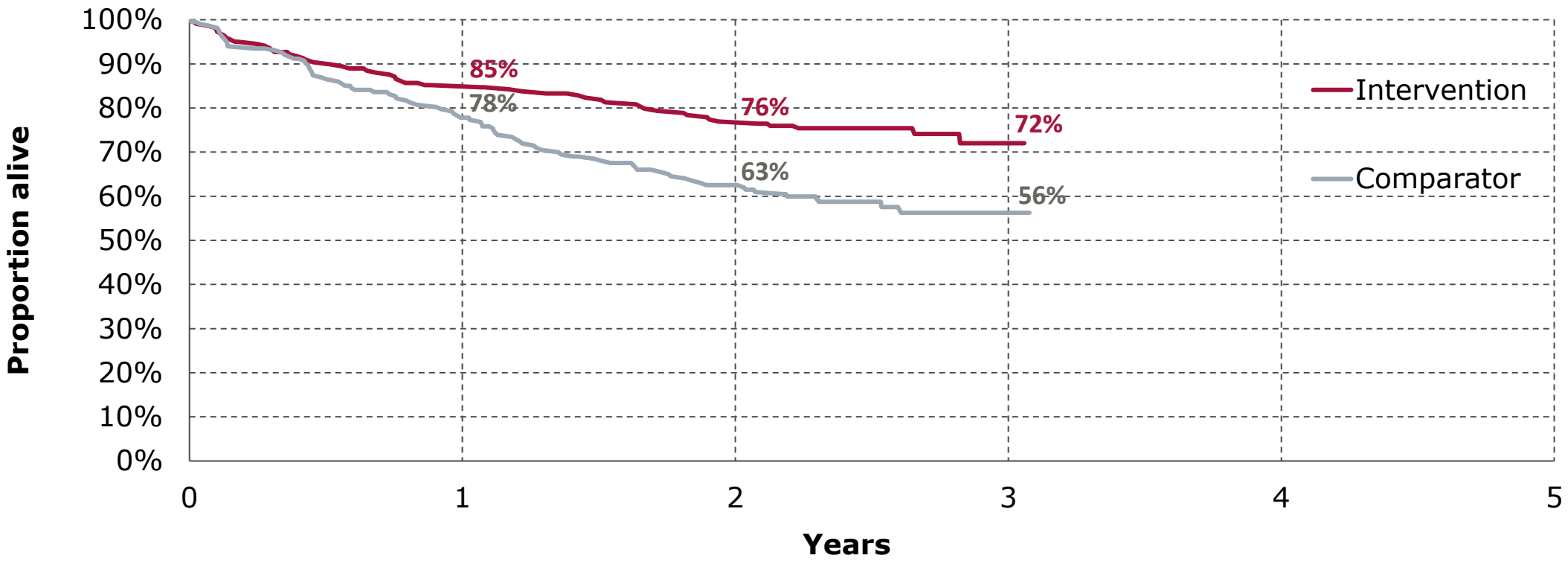


Interactive exercise: What proportion of the trial population in the intervention and comparator arms will be alive 5 years after treatment initiation?

- I'll be asking for your **lower plausible limit (L)**, **upper plausible limit (U)**, and **median (M)** based on the evidence on the following slide.
- Please make a note of your L, U and M.
- We can then discuss the rationale for your judgements and challenges with making these judgements.

LOWER PLAUSIBLE LIMIT (L)	A value such that you believe that there is a **1% probability** that the true value is less than L
UPPER PLAUSIBLE LIMIT (U)	A value such that you believe that there is a **1% probability** that the true value is greater than U
MEDIAN (M)	A value such that the true value is **equally likely** to be less than or greater than M

Key evidence to inform your judgements is shown below.



FACTORS TO CONSIDER

Survival bias due to:

- Disease indolence
- General fitness
- Treatment sensitivity

- Advancing age and comorbidities
- Progression
- Accumulation of treatment resistant clones

Discussion.

1. How did you come to your judgements?
2. Did you find the exercise easy or difficult? Why?
3. What other data could have been presented to make the exercise easier (without introducing bias)?
4. Could the question have been asked differently to make the exercise easier (without introducing bias)?
5. Do you think a consensus workshop would lead to more accurate judgements?

Agenda

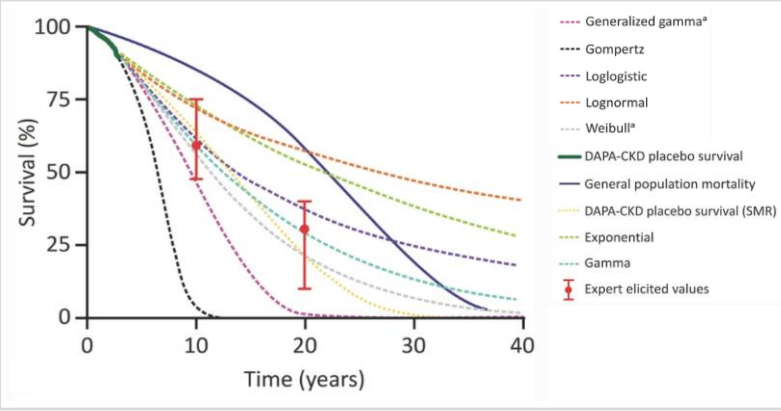
1. Introduction to SEE
2. Case study and interactive exercise
3. **Key learnings from the UK**

Historically, SEE has successfully been used to inform uncertain model parameters for NICE appraisals.

1

Long-term survival in CKD[1, 2]

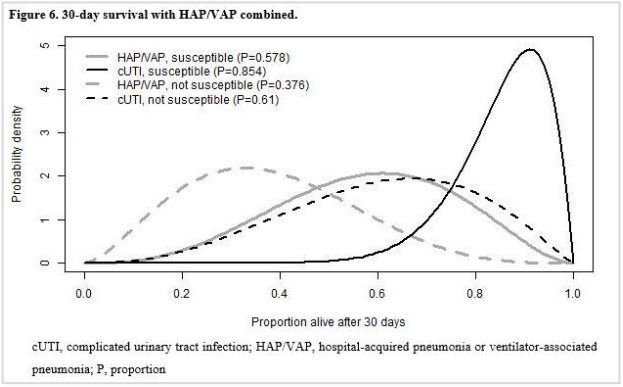
- **Problem**: uncertainty surrounding long-term survival due to immaturity of trial data
- **Quantities**: survival rates in patients with CKD receiving placebo at 10 and 20 years
- **Experts**: 6 leading disease area experts
- **Framework**: Cooke's classical method
- **Value**: confirmed that the survival function selected for the CEA has clinical validity



2

Short-term outcomes following infection in anti-microbial resistance[3]

- **Problem**: uncertainty surrounding short-term outcomes of specific infections due to lack of data
- **Quantities**: 30-day survival and hospital length of stay following treatment of specific infections
- **Experts**: 7 UK hospital consultants, microbiologists and pharmacists
- **Framework**: STEER – online survey with mathematical aggregation
- **Value**: probability distributions for 30-day survival and length of stay



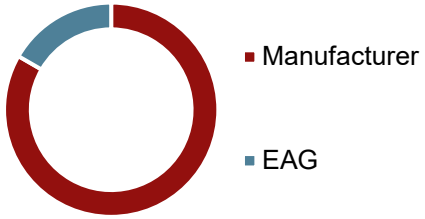
Abbreviations: CEA = cost effectiveness analysis, CKD = chronic kidney disease, NICE = National Institute for Health and Care Excellence, SEE = structured expert elicitation, STEER = Structured Expert Elicitation Resources

References: [1], Willigers et al. (2023) Advances in Therapy 40:2741–2751; [2], Dapagliflozin for treating chronic kidney disease, NICE TA775 (2022); [3], Harnan et al. (2021) Final report for the technology evaluation of ceftazidime with avibactam for treating severe aerobic Gram-negative bacterial infections.

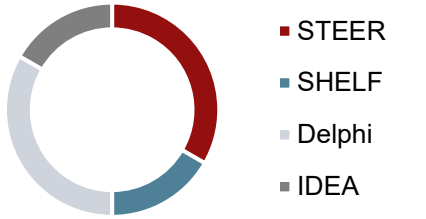
But in the first two years since SEE was included in the NICE manual, both manufacturers and decision makers appear to be adjusting to the methodology.

8% In the first two years since SEE was included in the NICE manual, only 6/73 completed appraisals have included an SEE, and only three of these involved probabilistic elicitation.

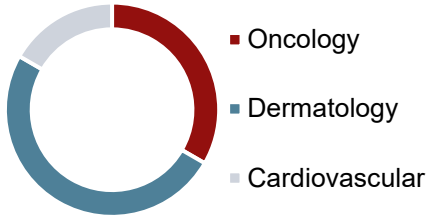
Who conducted the study?



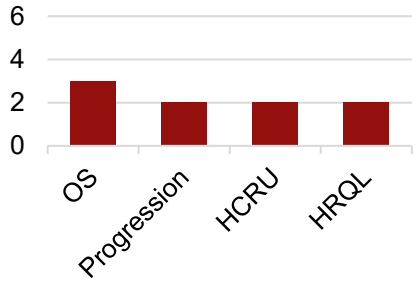
Which framework was used?



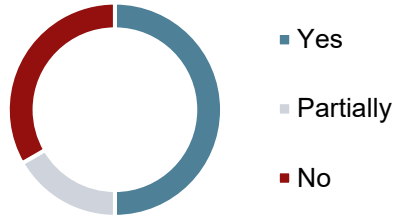
What was the indication?



What parameters were explored?



Did NICE accept the outcome?



EAG CRITIQUE OF SEE

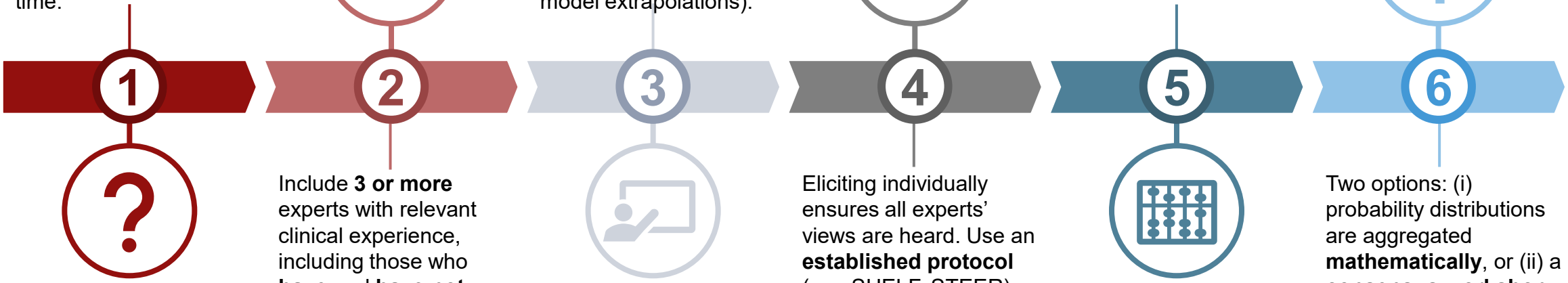
- Too few experts or exclusion of relevant expert types
- Quantities not ‘observable’ to experts
- Deviations from standard protocols
- Inaccurate representation of expert uncertainty

CONCLUSIONS

- Where SEE has been conducted to inform an economic model, it has regularly been considered **appropriate** for decision making.
- However, SEE outcomes have more often been rejected where they had a **material impact** on the model outcomes, in favour of extrapolations and simple assumptions.
- There is still work to be done to optimize methods to improve overall **trust and acceptability** for decision making.

Abbreviations: EAG = Evidence Assessment Group; HCRU = healthcare resource use; HRQL = health-related quality of life; NICE = National Institute for Health and Care Excellence; OS = overall survival; SEE = structured expert elicitation; SHELF = Sheffield Elicitation Framework; STEER = Structured Expert Elicitation Resources

NICE Technical Support Document 26 provides additional guidance for conducting SEE for long-term survival outcomes.

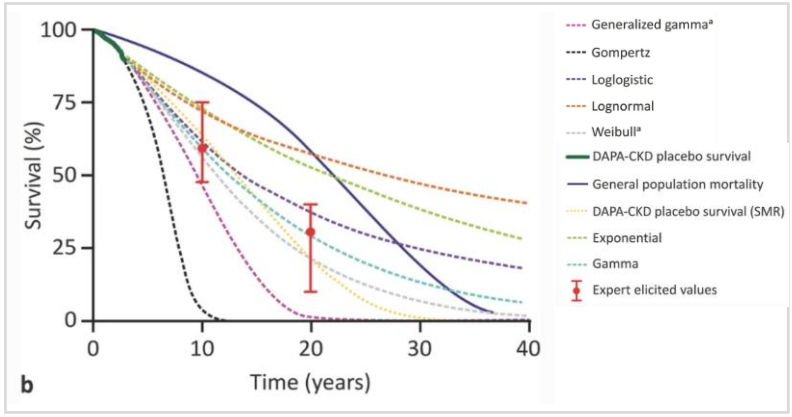


SEE = structured expert elicitation

TSD26 recommends two options for incorporating SEE results into economic models.

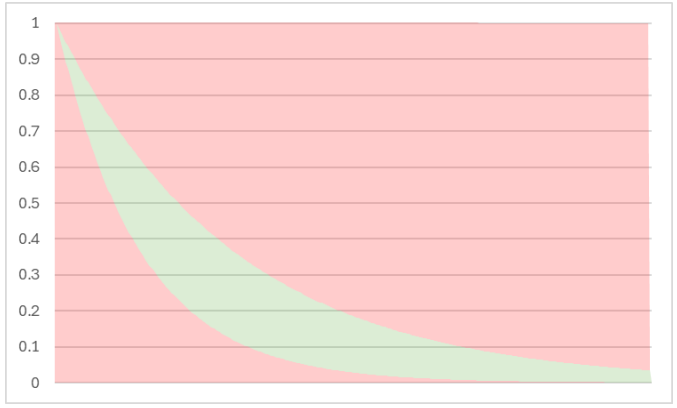
TO INFORM SELECTION OF PARAMETRIC SURVIVAL CURVES

- **Approach:** Consider which of the standard parametric survival curves are plausible based on alignment with the **median** and **90% credible interval**, and qualitative expert judgements.
- **Pros:** Simple approach, precedent for acceptance
- **Cons:** Does not use SEE output to its full potential, cannot directly inform the probabilistic sensitivity analysis



BAYESIAN MULTIPARAMETER EVIDENCE SYNTHESIS

- **Approach:** Generation of new survival extrapolations based on the combination of patient-level data from the trial and the output of the SEE exercise
- **Pros:** Incorporates expert uncertainty into the probabilistic sensitivity analysis
- **Cons:** Complex approach, lack of guidelines, no precedent for acceptance, risk of double-counting



Thanks for listening.

Discussion.

1. Do you believe SEE can help to improve healthcare decision making?
2. Under what circumstances can SEE add value?
3. What can be done to improve uptake and acceptability of SEE?
4. Would you consider using SEE?