Experiences of Structured Elicitation for Cost-Effectiveness Analyses

Marta Soares, Centre for Health Economics, University of York

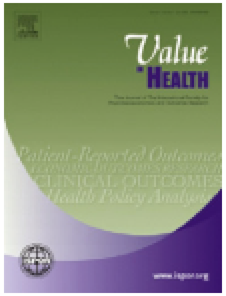
LSHTM, 29 June 2018

VALUE IN HEALTH 21 (2018) 715–723

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval



Experiences of Structured Elicitation for Model-Based Cost-Effectiveness Analyses

Marta O. Soares, PhD[1,*], Linda Sharples, PhD[2], Alec Morton, PhD[3], Karl Claxton, PhD[1,4], Laura Bojke, PhD[1]

[1]Centre for Health Economics, University of York, York, UK; [2]Medical Statistics Department, London School of Hygiene and Tropical Medicine, London, UK; [3]Management Science, University of Strathclyde, Glasgow, UK; [4]Department of Economics, University of York, York, UK

What is expert elicitation?

"systematic process of formalizing and quantifying, typically in probabilistic terms, expert judgments about ***uncertain*** *quantities"* [White paper on elicitation]

- Bayesian inference (subjective priors).
- Used formally in support of decision making in other areas of science: e.g. food safety. [EFSA]
- Uncertainty reflects degree of belief over an uncertain quantity, i.e. imperfect knowledge (epistemic uncertainty)
- Good elicitation (structured expert elicitation, SEE) should minimise bias and heuristics

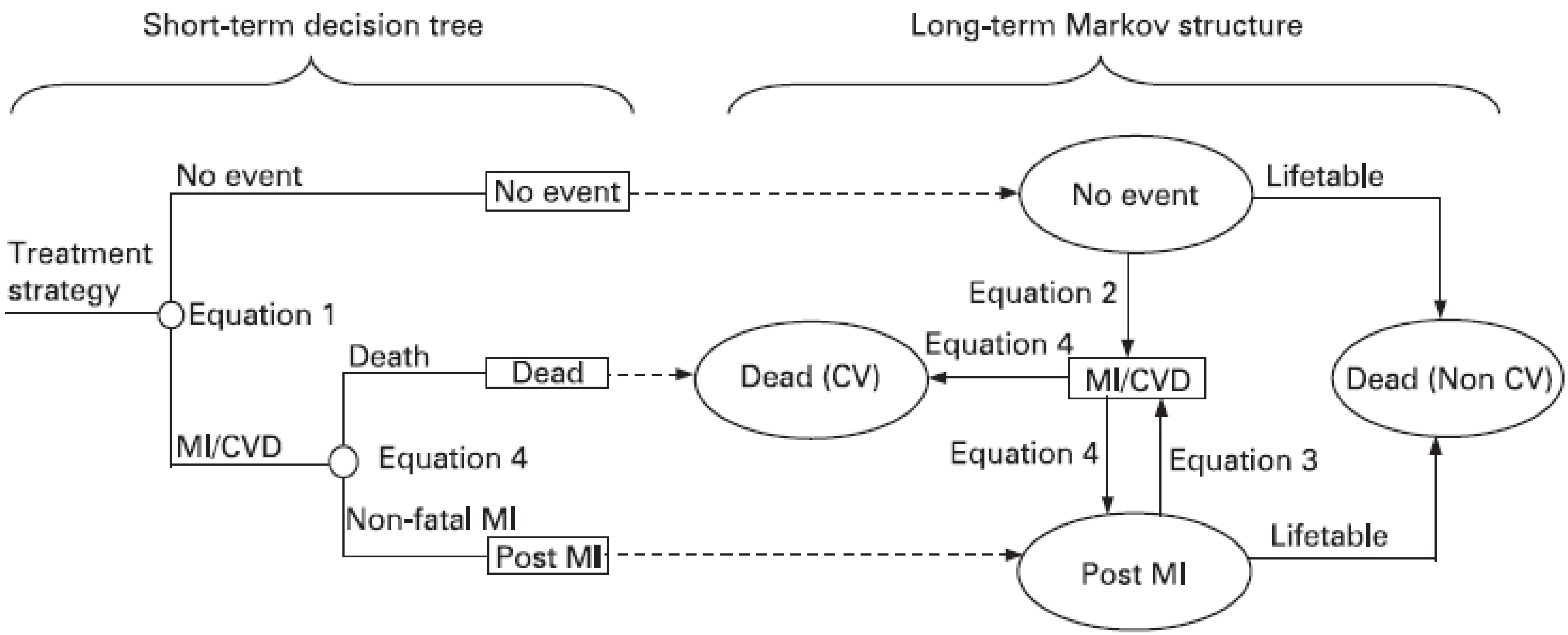
But inevitably, the probabilities elicited are personal

What is decision modelling for health care decision making?

- Health systems need to make a number of decisions on health care resources: which, whom, when, where…
  - Typical example is reimbursement/access (NICE)
  - Other examples: individual funding requests (NHS England)
- Decisions informed by HTA including cost-effectiveness, the latter comparing interventions in terms of:
  - long-term effects on population health (typically measured in QALYs), and
  - overall cost implications for relevant stakeholders or individuals.
- Evaluations are often model-based, where evidence from multiple sources and on multiple aspects of disease and treatment are considered together, and the model used to extrapolate to the long term.
- Uncertainty in the evidence can result in uncertain cost-effectiveness estimates.

Decision modelling

State transition models



Decision modelling

UNIVERSITY of York

CHE
Centre For Health Economics

Parameter space

- Transitions
- Relative effectiveness
- Health state costs
- Health state HRQoL weights
- Structural uncertainties

Evidence space

- Empirical evidence
- Mechanistic evidence
- Judgements

Breadth and quality of evidence

- Evidential uncertainty
- Less representative
- Sparse
- Non high quality
- Indirectly relevant evidence
- Inconsistent evidence

Decision modelling and decision making

- Which intervention is worthwhile using
  - Based on expected values
  - Rejecting an intervention that is expected to be beneficial on the grounds of uncertainty solely is not ethically justified

- Under uncertainty, further timely research may be needed
  - To reverse initial guidance is proven wrong
  - Avoid the consequences of uncertainty

- Approval/rejection can be conditioned on research being conducted

Uses for elicitation in CE modelling

- Uncertainty is pervasive in CE modelling
  - SEE can help characterise it
- Judgements are ALWAYS required for a decision to be reached
  - SEE can contribute to accountability
- Increased interest in SEE with decisions closer to launch and increased use of 'early modelling'.
- Possible uses of elicitation in HTA:
  - Initial estimates where empirical data is absent [refs]
  - weights for alternative structural assumptions (model averaging)
  - weights for bias correction (observational evidence)
  - between trial heterogeneity in (network) meta analysis
  - sample size calculations
  - ….

How to conduct an elicitation and motivation for further work

- An elicitation exercise requires a number of decisions on aspects of its design, conduct and analysis.
- Little research to support the choices that need to be made.
  - Accuracy cannot be measured as expert beliefs are inherently unobservable
  - Existing methodological research: noncommittal, inconsistent, poor quality…
- Methodological uncertainties may be main reason for the limited use of SEE in the context of HTA.
- There are no protocols (guides to good practice) in HTA, but there are generic ones, e.g.:
  - Sheffield elicitation framework (SHELF)
  - Cooke's classical method (Cooke R.M 1991)
- Not clear if any can appropriately be used to inform health care decisions

MRC work

Developing a reference protocol for expert elicitation in health care decision making

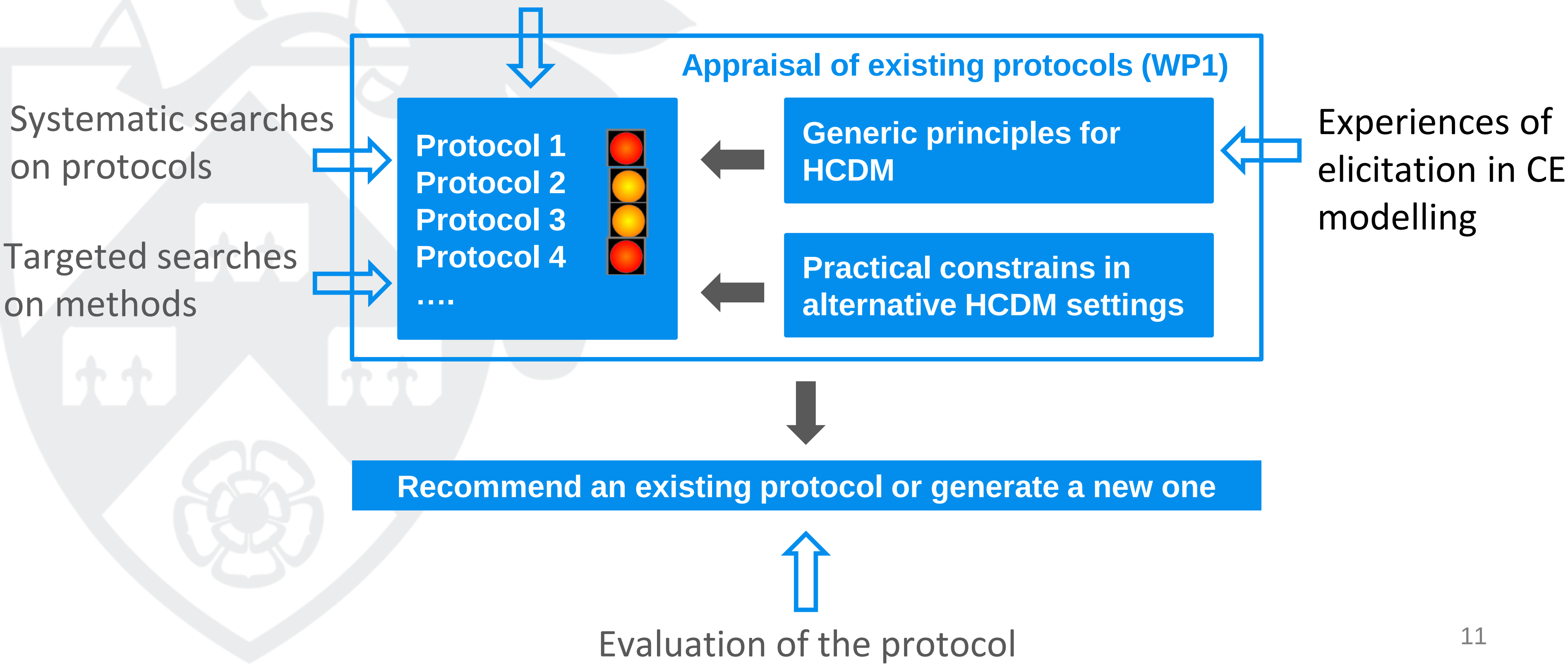
Bojke L, Claxton K, Fox A, Jankovic D, Soares M (U of York),
Taylor A (U of Leeds),
Sharples L (LSHTM),
Jackson C (U of Cambridge),
Morton A, Colson A (U of Strathclyde)

Timeline: 11.2017 – 03.2019

Workplan for MRC work

Experiments (WP2,3)

- Comparison of two methods to elicit distributions (parameter uncertainty)
- Can individuals elicit accurately when their knowledge-base is different to target quantity?
- How do individuals revise their estimates in Delphi-type processes?



HCDM: Health Care Decision Making

Experiences of elicitation in CE

- An update of a previous review [Grigore 2013] was conducted to identify applied SEEs (eliciting uncertainty) in cost-effectiveness modelling

- Main aims:
  - summarise the basis for methodological choices made in each application
  - record the difficulties and challenges reported by the authors in the design, conduct, and analyses.

Experiences of elicitation in CE

Study	Type of strategy under investigation
Garthwaite et al. [14]	Treatment
Leal et al. [10]	Diagnostic/screening
Girling et al. [15]	Treatment
Stevenson et al. [16]	Prevents transmission
Meads et al. [12]	Diagnostic/screening
McKenna et al. [19]	Treatment
Haakma et al. [13]	Diagnostic/screening
Stevenson et al. [17]	Treatment
Speight et al. [25]	Diagnostic/screening
Sperber et al. [22]	Treatment
Brodtkorb [26]	Several exercises conducted

Study	Type of strategy under investigation
Colbourn et al. [28]	Diagnostic/screening
Soares et al. [9]	Treatment
Bojke et al. [18]	Treatment
Cao et al. [11]	Diagnostic/screening
Fischer et al. [23]	Treatment
Poncet et al. [27]	Diagnostic/screening
Grigore et al. [24]	Treatment
Wilson et al. [20]	Treatment
Meeyai et al. [21]	Vaccine
Grimm et al. [35]	Diagnostic/screening

R&D, research and development.
† Rate of implementation in clinical practice over time.

Experiences of elicitation in CE

The paper focussed on:

- Aspects related to the design
  - Specification of the quantities to elicit
  - Selection of experts
  - Elicitation method
  - Consensus vs. mathematical aggregation, weighting of experts
- Experiences with the conduct of the exercise
- Experiences with the analyses and interpretation
  - Considerations on validity
  - Synthesis of multiple beliefs in mathematical aggregation
  - Deriving smooth prior distribution functions
  - Further use of elicited evidence in decision modelling
- Considerations on bias

Experiences of elicitation in CE

Today, I will focus on:

- Aspects related to the design
    - Specification of the quantities to elicit
    - Selection of experts
    - Elicitation method
    - Consensus vs. mathematical aggregation, weighting of experts
- Experiences with the conduct of the exercise
- Experiences with the analyses and interpretation
    - Considerations on validity
    - Synthesis of multiple beliefs in mathematical aggregation
    - Deriving smooth prior distribution functions
    - Further use of elicited evidence in decision modelling
- Considerations on bias

Experiences of elicitation in CE:

Specification of the quantities to elicit

- Several quantities can often be used to inform a parameter.
- For a probability-related parameter, alternative quantities elicited **may** be:

Alternative quantities to inform probability-related parameters	Mathematical specification*
Probability, or conditional probability	P[D=1] or P[D=1|A]
Odds	P[D=1]/(1-P[D=1])
Transition probability	$P[D_{c+1}=1 \mid D_c=0] = P[c<T<c+1]$
Survival probability	$S(t)= P[D_t=1] = P[T>t]$
Time to event summaries	e.g. mean or median time to event
Rates or hazards	h(t)
Parameters of the hazard function	e.g. Weibull(shape, scale)

Experiences of elicitation in CE:

Specification of the quantities to elicit

- Several quantities can often be used to inform a parameter.
- For a probability-related parameter, alternative quantities elicited **have** been:

	Alternative quantities to inform probability-related parameters	Mathematical specification*
✓	Probability, or conditional probability	P[D=1] or P[D=1|A]
✗	Odds	P[D=1]/(1-P[D=1])
✓	Transition probability	$P[D_{c+1}=1 \mid D_c=0] = P[c<T<c+1]$
✓	Survival probability	$S(t)= P[D_t=1] = P[T>t]$
✓	Time to event summaries	e.g. mean or median time to event
✗	Rates or hazards	h(t)
✗	Parameters of the hazard function	e.g. Weibull(shape, scale)

Experiences of elicitation in CE:
Specification of the quantities to elicit

Considerations when choosing quantities to elicit

- Appropriateness for experts
  - observable rather than unobservable quantities [Kadane 1998]
    - Relative effectiveness parameters expressed as probabilities [9–12]
    - Sensitivities and specificities into probabilities of the true disease status of the patients conditional on the test results [12].
  - Different experts elicit different quantities
    - geneticists elicited accuracy of a genetic test and cardiologists elicited parameters related to disease progression [10]
  - Heterogeneity
    - elicit separately for population subgroups [12,13]

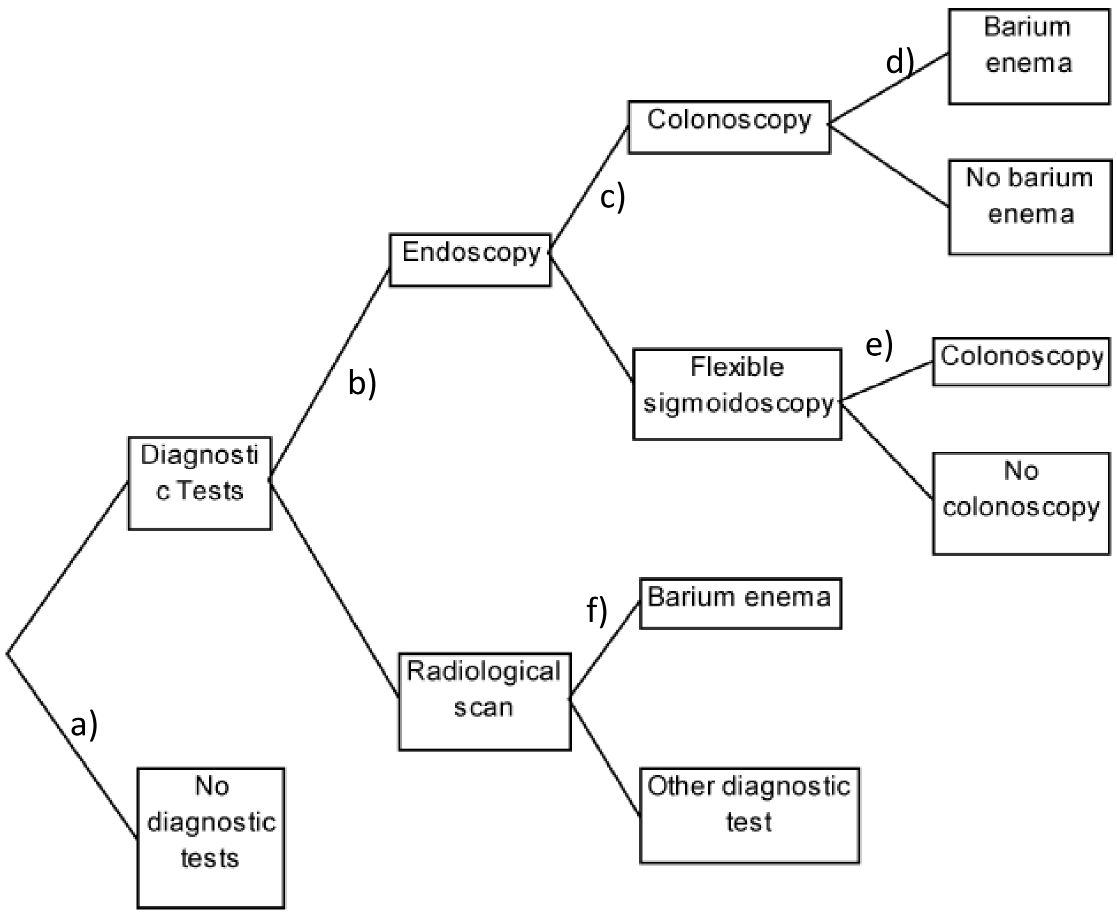
Experiences of elicitation in CE:
Specification of the quantities to elicit

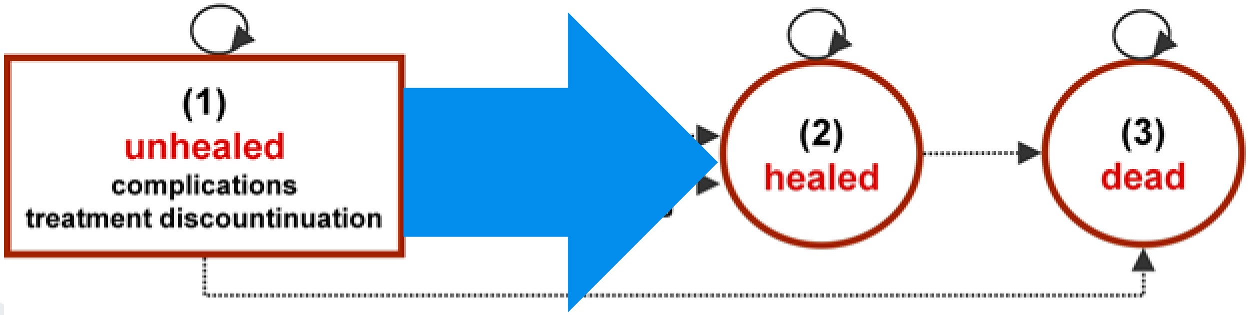
- Statistical concerns
  - Fit for purpose and allow combining with existing empirical evidence [9].
    - e.g. time dependency
  - Coherence between quantities
    - e.g. elicit multinomial events using conditional binomials
  - Dependencies
    - Avoid dependencies between quantities elicited, e.g. re-expressing parameters using conditional independence [9,13].
    - Two studies explicitly elicited dependency [14,18]

Experiences of elicitation in CE:
Specification of the quantities to elicit

- Burden to experts
    - Limit number of target parameters to elicit
    - Elicit homogeneous quantities throughout (e.g. probabilities)
    - Use filter questions (e.g. do you think X differs from Y?)
    - Avoid dependency elicitation
        - Elicit only for covariate indicated as relevant by the expert

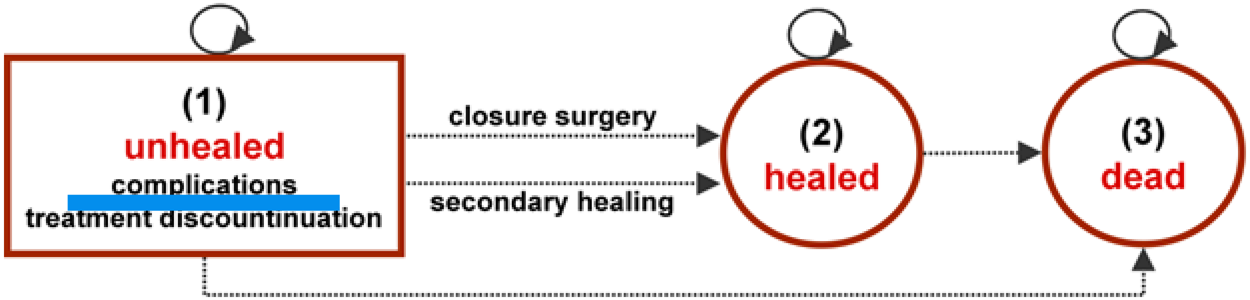
Garthwaite et al

Questions elicited	Re-structured tree
(a) What proportion of patients referred for investigation of symptoms of bowel cancer do not undergo diagnostic testing (i.e., go straight to surgical intervention)? (b) What proportion of the patients referred undergo endoscopy (flexible sigmoidoscopy or colonoscopy) as their first investigation rather than a radiological scan (barium enema, colonography)? (c) Of those patients undergoing endoscopy, what proportion undergo colonoscopy as their first investigation? (d) Of those patients undergoing colonoscopy, what proportion then have a barium enema? (e) Of those patients who undergo flexible sigmoidoscopy, what proportion then have a colonoscopy? (f) Of those patients undergoing a radiological scan, what proportion undergo a barium enema as their first investigation?	



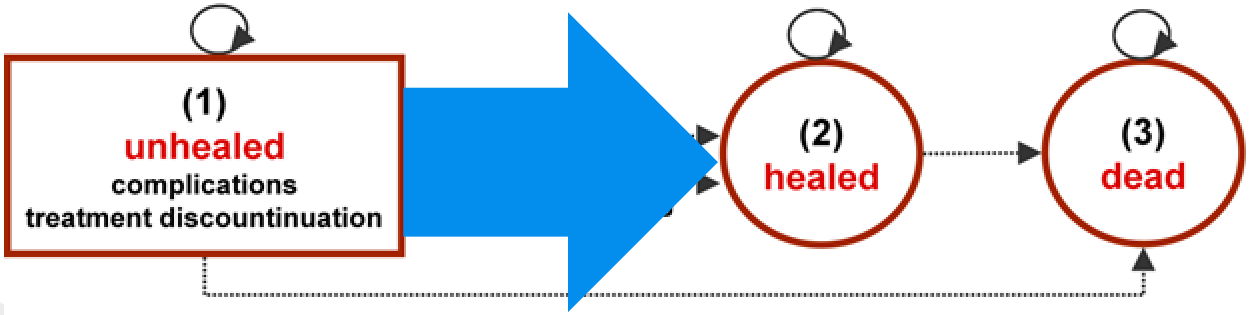
Soares et al

Q1: Think of unhealed patients in the UK. 6 months after starting treatment with *X* what % of patients who are alive do you think would have a healed reference ulcer?	P[S_6=healed | S_0=unhealed] (survival)



Soares et al

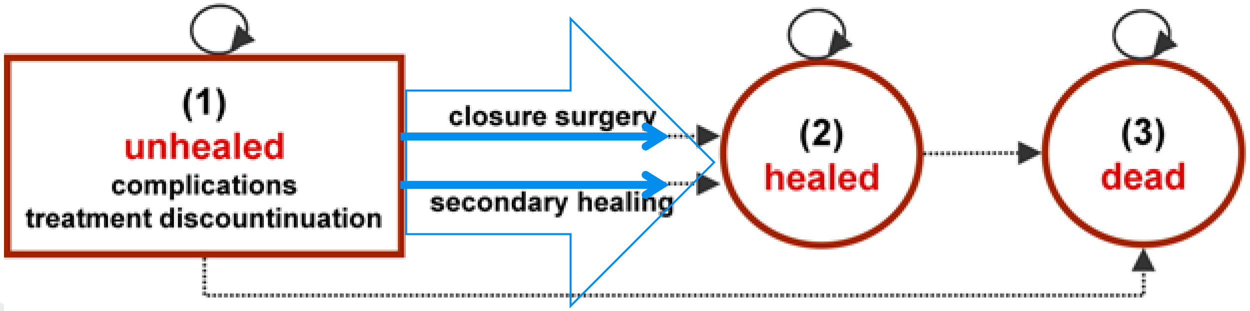
Q1: Think of unhealed patients in the UK. 6 months after starting treatment with *X* what % of patients who are alive do you think would have a healed reference ulcer?	P[S_6=healed | S_0=unhealed] (survival)
Q2: Now think of those patients who are still unhealed after 6 months. At this 6 month point, what % of these unhealed patients do you think would have complications?	P[S_6=complications | S_6=not healed, S_0=unhealed] (conditional probability)



Soares et al

Time dependency

Q1: Think of unhealed patients in the UK. 6 months after starting treatment with *X* what % of patients who are alive do you think would have a healed reference ulcer?	P[S_6=healed | S_0=unhealed] (survival)
Q2: Now think of those patients who are still unhealed after 6 months. At this 6 month point, what % of these unhealed patients do you think would have complications?	P[S_6=complications | S_6=not healed, S_0=unhealed] (conditional probability)
Q3: Think of those patients who are still unhealed after 6 months. What % do you think would heal their reference ulcer between 6 and 12 months?	P[S_{12}=healed | S_6=unhealed, S_0=unhealed] (survival, conditional)



UNIVERSITY of York

CHE Centre For Health Economics

Soares et al

Q1: Think of unhealed patients in the UK. 6 months after starting treatment with *X* what % of patients who are alive do you think would have a healed reference ulcer?	P[H] (survival)
Q2: Now think of those patients who are still unhealed after 6 months. At this 6 month point, what % of these unhealed patients do you think would have complications?	P[S_6=complications | S_6=not healed, S_0=unhealed] (conditional probability)
Q3: Think of those patients who are still unhealed after 6 months. What % do you think would heal their reference ulcer between 6 and 12 months?	P[S_{12}=healed | S_6=unhealed, S_0=unhealed] (conditional survival)
Q4: Think of patients that had their pressure ulcer healed completely at some point. What % of these patients achieved healing as a direct result of closure surgery?	P[S|H] (conditional probability)
Q5: Consider patients with an ulcer on which closure surgery has been performed. What % of patients who received this surgery do you think will have healed 1 month post-surgery?	P[H|S] (conditional probability)

Law of total probability $$P[S] = \frac{P[S|H] \cdot P[H]}{P[H|S]}$$

Experiences of elicitation in CE:
Selection of experts

- Not much known about who constitutes a good expert
- Desirable skills: substantive, normative, adaptive

- All applications recruited health professionals. Criteria were:
  - recognition by peers [10],
  - specialist knowledge or clinical experience [9,10,13,18,19,22,23],
  - based in the relevant jurisdiction [9,10,18,19],
  - research experience [10,22,23], and
  - lack of involvement in product development [13].
  - In early technology assessment, applications have also looked for other factors such as interaction with colleagues, seen as indicative of the adaptive skills required in this context.
- health care professionals unlikely to have knowledge of elicitation and may have only sparse quantitative skills [9,14,24]

Experiences of elicitation in CE:
Selection of experts

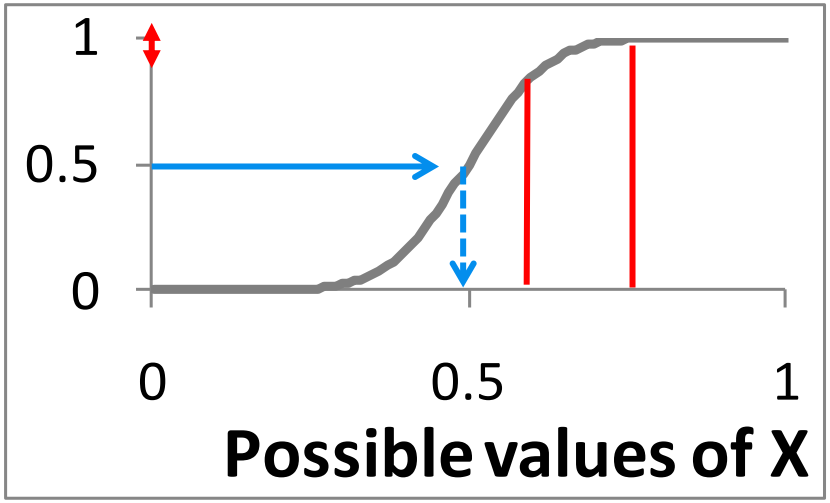
- Multiple experts better than one, but how many?

- Applications included a varied sample of experts: from a range of relevant specialties [10,12,20], clinical settings [9,10,20], and geographical areas/countries [10,23]
    - capture heterogeneity in beliefs (reflecting underlying heterogeneity in patient populations) and avoid dependency between experts [10].
- Sampling was purposeful:
    - collaborators in the research, recommendation [10,14,18], professional associations [24], specialist conferences [15,23].
- Sample sizes: 2 [11] - 23 [9]

Experiences of elicitation in CE:
Elicitation method

- Only a small number of summaries can be elicited
- Approaches based on probabilities:
  - variable interval methods

    The facilitator specifies a probability, p, and asks the expert for a value x such that - P[X ≤ x]=p



  - fixed interval methods
    - The facilitator specifies a value of the random variable (or an interval of values) and asks the expert for how much of the probability, p, should be allocated - P[X ≤ x]=p

- Performance of the alternative methods unclear

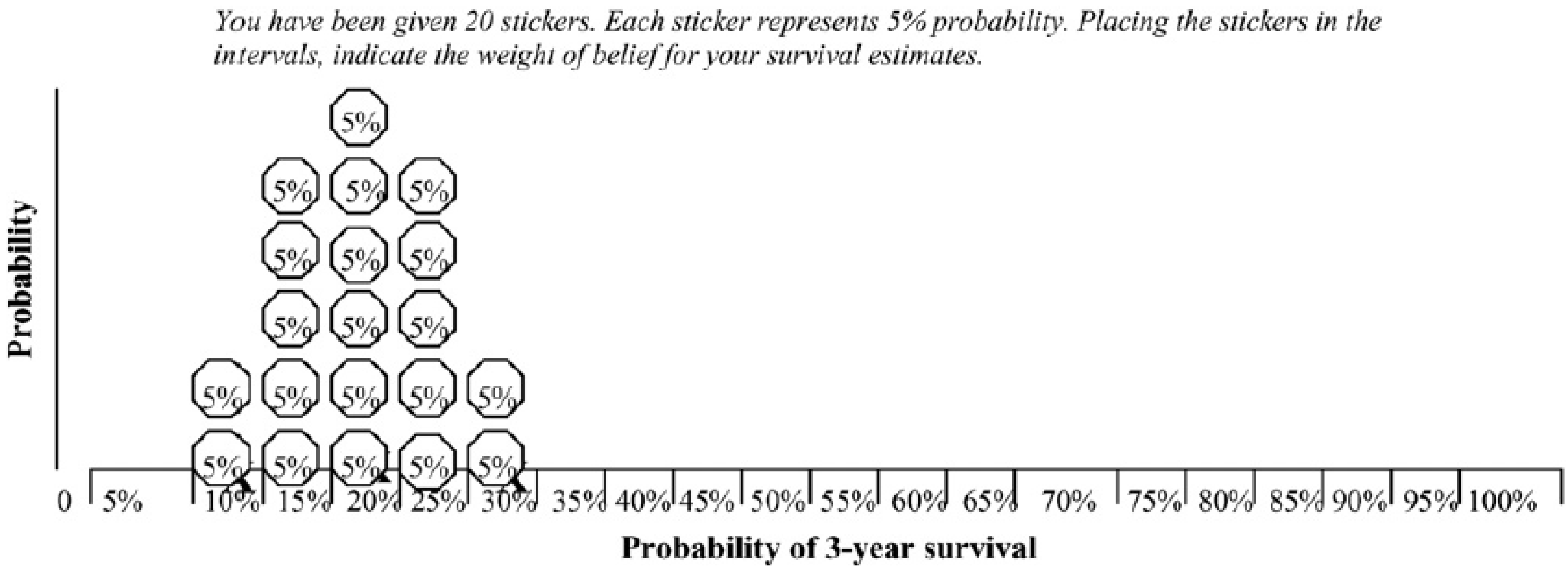
Experiences of elicitation in CE: Elicitation method

- Example of a variable interval method: Bisection (quartiles) [O'Hagan 2006]

1. Can you determine a value (median) such that X is equally likely to be less than or greater than this point? $P[X \le x_1] = 0.5$
2. Suppose you were told that X is below your assessed median. Can you now determine a new value (lower quartile) such that it is equally likely that X is less than or greater than this value? $P[X \le x_2] = 0.25$
3. Suppose you were told that X is above your assessed median. Can you now determine a new value (upper quartile) such that it is equally likely that X is less than or greater than this value? $P[X \le x_3] = 0.75$

Experiences of elicitation in CE:
Elicitation method

- Example of a fixed interval method: chips and bins



Experiences of elicitation in CE: Elicitation method

- What did applications use?
  - 8 used VIMs[13,14,17,21–23,28]: Quartiles or credible intervals
  - 9 used FIM [9–12,18,19,24–27]: chips and bins and 4 complementary intervals
  - 3 did not report method
- No clear guidance on how to choose between methods
  - Four application pilot different methods (find contradictory results)
  - Others justify choice on:
    - citations of generic methods research,
    - previous use in CE modelling, and
    - claims of lower burden or intuitiveness for experts (chips and bins)

Experiences of elicitation in CE:
Consensus vs mathematical aggregation

- Different approaches
  - Consensus – where synthesis is implicit
  - Delphi type processes – implicit synthesis with controlled interaction, based on rounds of individual revision after group summaries are presented
  - Mathematical – where synthesis is explicit (methods used can influence results)

Experiences of elicitation in CE:
Consensus vs mathematical aggregation

[Montebellier 2018, Elicitation: Science and art of structuring judgement, Springer]

- Benefits of group interaction
  - Error checking, Motivation, pooling of information and perspectives
- Biases associated with group interaction
  - Groups may increase or attenuate individual biases
  - Group biases:
    - false consensus,
    - groupthink,
    - polarization,
    - escalation of commitment,
    - overconfidence (particularly where task is complex)
  - Facilitation or Delphi may alleviate group biases

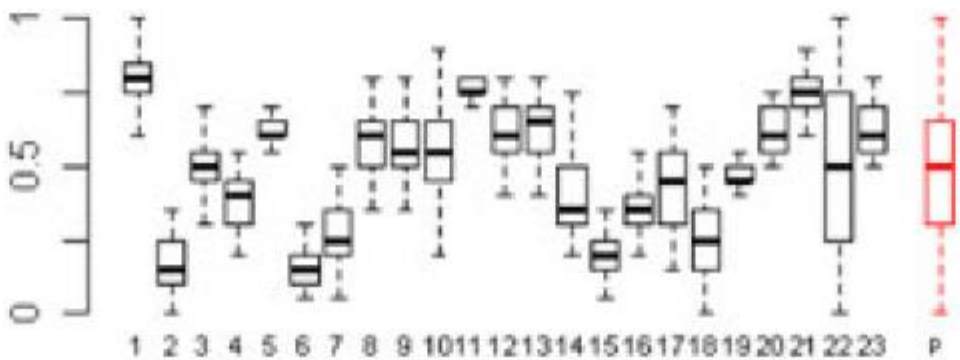
Experiences of elicitation in CE:
Consensus vs mathematical aggregation

- 14 studies [9-14,18,19,22-24,27,28,35] elicited individually,
- 3 aimed to achieve consensus among experts [15,17,21],
- 3 did not explicitly report the method [16,20,25].

- Choice of mathematical:
  - desirability to reflect variation within and between experts [12],
  - consensus leads to overconfident results (i.e., narrow distributions) [10]
  - difficult to convene experts and provide experienced facilitation.
  - In one pilot study [9] consensus produced incoherent probability statements (median time to healing > time taken for 70% of patients to heal).

Experiences of elicitation in CE:
Consensus vs mathematical aggregation

- Between expert variation



Soares et al

TABLE 6 *Mean and standard deviation (SD) of the elicited values for each expert separately, and linearly pooled results across experts*

	Mean probability of sustaining year 1 QoL benefits in subsequent years[a] (SD)		
	Year 2	**Year 3**	**Year 4+**
Expert 1	0.670 (0.091)	0.600 (0.082)	0.526 (0.088)
Expert 2	0.807 (0.047)	0.886 (0.052)	0.886 (0.052)
Expert 3	0.785 (0.039)	0.700 (0.057)	0.675 (0.075)
Expert 4	0.605 (0.104)	0.605 (0.104)	0.605 (0.104)
Expert 5	0.908 (0.036)	0.905 (0.035)	0.898 (0.043)
Pooled result	0.757 (0.126)	0.742 (0.150)	0.719 (0.168)

a Conditional on sustaining benefits in the previous year and receiving top-up procedures as considered appropriate.

McKenna et al

Experiences of elicitation in CE: Conclusions

- Across applied examples:
    - Elicitation conducted mostly when evidence is absent
    - Generally poor reporting of methods
    - Lack of homogeneity in methodology

Experiences of elicitation in CE: Conclusions

- Critical aspects for HTA
  - Judgements required on a large number of parameters
  - Substantive experts are not expected to have very strong quantitative skills
  - Between-expert variation is warranted and expected due to heterogeneity
  - Decision makers seek for assurance on validity
  - Little integration with behavioural psychology research

Experiences of elicitation in CE: Conclusions

- SEE is essential for accountable decision making

- Methodological guidance is needed now!
  - Homogeneity in methods across appraisals
  - Reflect specificities of HTA
  - Flexible to accommodate constraints of the decision making context
    - For example, in early assessment it is desirable for experts to have (or be trained in) adaptive skills
    - Decision making context: e.g. budget, timelines

- Further methodological research is important
  - Integration with behavioural research -- debiasing techniques
  - Define best practice in elicitation
  - Particularly for the less normative experts

Thanks!

Marta Soares, Centre for Health Economics, University of York

LSHTM, 29 June 2018