CHIL
CENTRE FOR HEALTH
ECONOMICS IN LONDON

Framework and practical tool for eliciting expert priors in clinical trials with MNAR outcomes

Alexina Mason

joint work with James Carpenter, Richard Grieve and Manuel Gomes

Centre for Statistical Methodology Seminar
26 October 2018

Outline

Background

- Missing data in RCTs can lead to
  - underestimation of uncertainty
  - biased effectiveness estimates
- The analysis of RCTs typically assumes outcomes are ‘missing at random’ (MAR)
  - probability of observing outcome does not depend on patient’s outcome, after conditioning on the observed data
- However, concern MAR may be implausible in many settings
  - in particular for patient-reported outcomes, e.g. quality of life
- So, methodological guidelines recommend sensitivity analysis recognising data could be ‘missing not at random’ (MNAR)
- But, recently Bell *et al.* (2014)
  - reviewed 77 RCTS published in prestigious medical journals
  - found none conducted MNAR sensitivity analysis

Aims of this research

- Why is MNAR sensitivity analysis not performed?
- Requires specifying plausible assumptions - a challenge!
- Elicited expert opinion can help frame relevant MNAR scenarios
- OUR AIM: to encourage triallists to conduct MNAR sensitivity analyses by providing
  - a framework for eliciting expert opinion about MNAR outcomes in clinical trials
  - a practical, easy-to-use, elicitation tool
- We have developed these using two typical trials
  - IMPROVE
  - POPPI

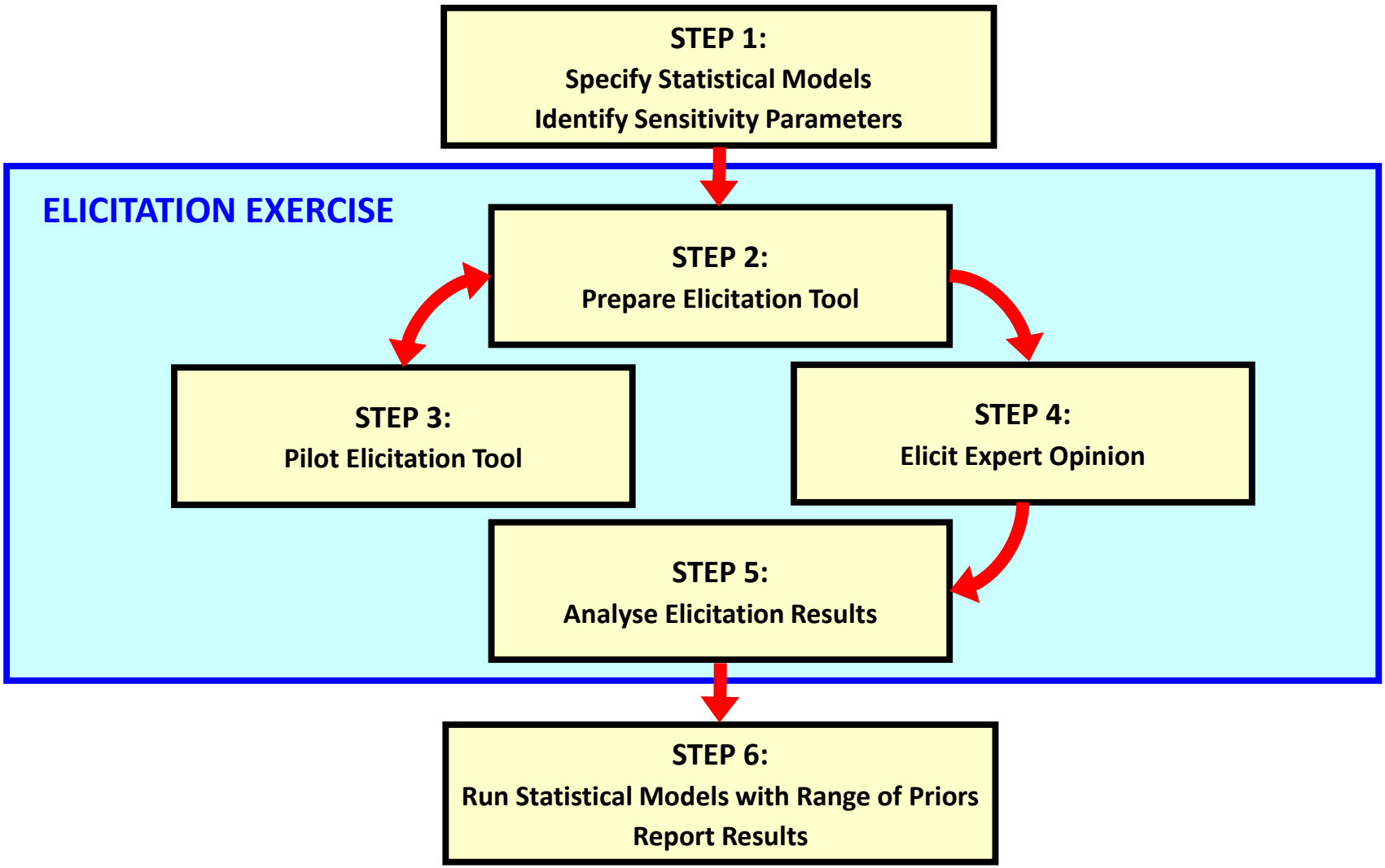
Illustration 1: the IMPROVE trial

- Aim of the IMPROVE trial (Grieve *et al.*, 2015)
  - to assess the effectiveness of an emergency endovascular strategy (eEVAR) compared with open repair (OPEN) to treat patients with a ruptured abdominal aortic aneurysm
- Pragmatic, parallel design, multi-centre RCT (30 sites, N=613)
- Outcomes included quality of life (QoL) at 3 months
  - measured using the EQ-5D-3L health questionnaire
- Primary analysis found a clinically significant difference favouring eEVAR
- 21% of surviving patients did not complete EQ-5D-3L questionnaire at 3 months
  - primary analysis used multiple imputation assuming MAR
  - unclear whether gains depend on the assumptions about the missing data

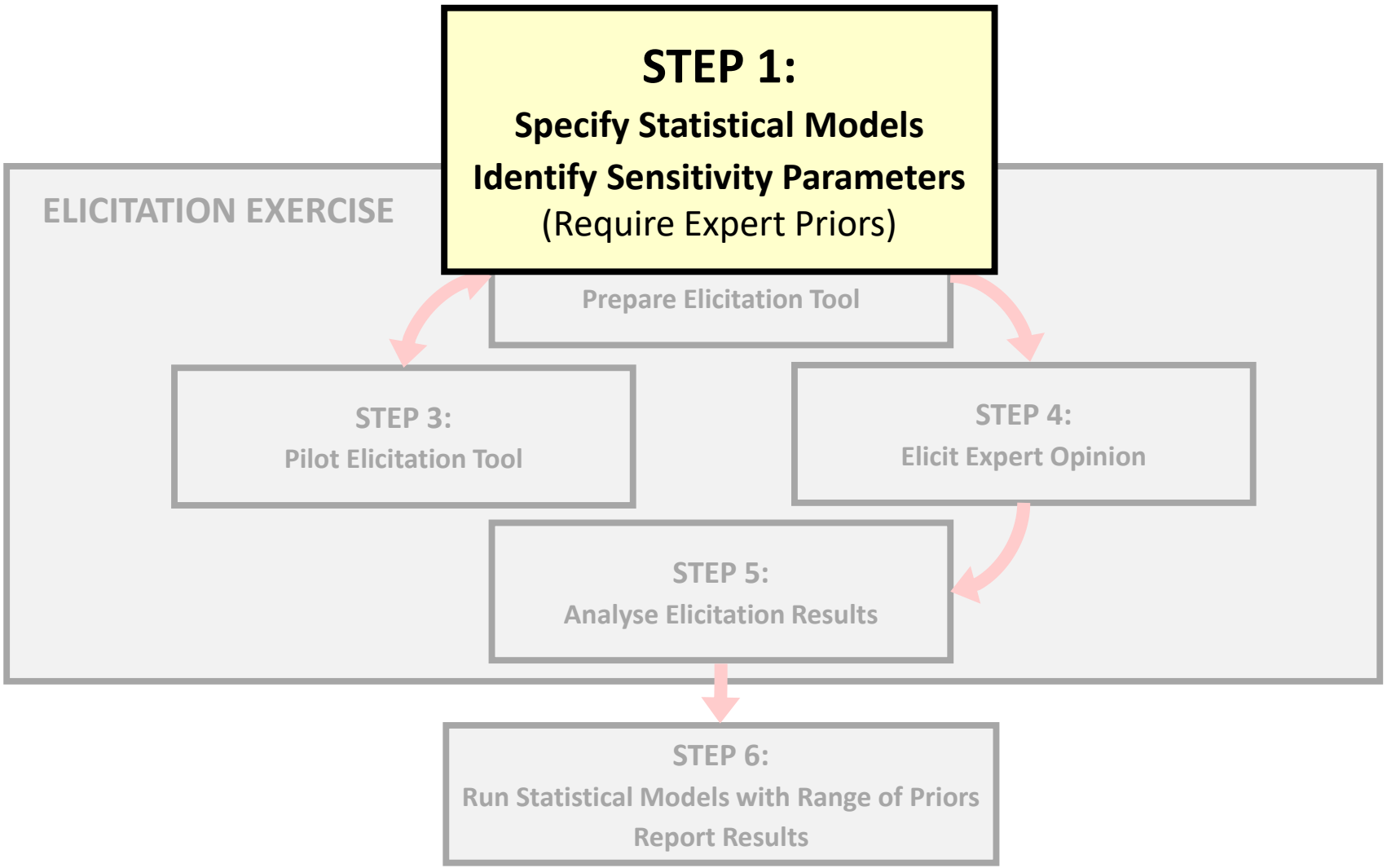
Illustration 2: the POPPI trial

- Aim of the POPPI trial
  - to investigate whether a complex psychological intervention, commenced early in ICU, would reduce patient-reported PTSD symptom severity
- Parallel design, multi-centre, cluster RCT (24 sites, N=1,458)
- Two important outcomes collected by self report questionnaire at 6 months
  - Primary outcome: patient-reported PTSD symptom severity, measured using the PTSD Symptom Scale (PSS-SR)
  - Health-related quality of life, measured using the EQ-5D-5L health questionnaire (HrQoL)
- Loss to follow-up anticipated to be 20% among survivors
- Statistical analysis plan specified
  - primary analysis using multiple imputation assuming MAR
  - MNAR sensitivity analysis informed by expert opinion

Framework overview



Framework: Step 1 - Specify Statistical Models



Joint models

- Notation:
  - Let $\boldsymbol{z} = (z_{ij})$ denote a rectangular data set of interest
    $i = 1, ..., n$ individuals and $j = 1, ..., k$ variables
  - Let $\boldsymbol{m} = (m_{ij})$ be a binary indicator variable such that

$$m_{ij} = \begin{cases} 0: & z_{ij} \quad \text{observed} \\ 1: & z_{ij} \quad \text{missing} \end{cases}$$

  - Let β and θ denote vectors of unknown parameters
- Then the joint model (likelihood) of the full data is $f(\boldsymbol{z}, \boldsymbol{m}|\beta, \theta)$
- MNAR requires modelling
  1. data of interest
  2. missing data mechanism

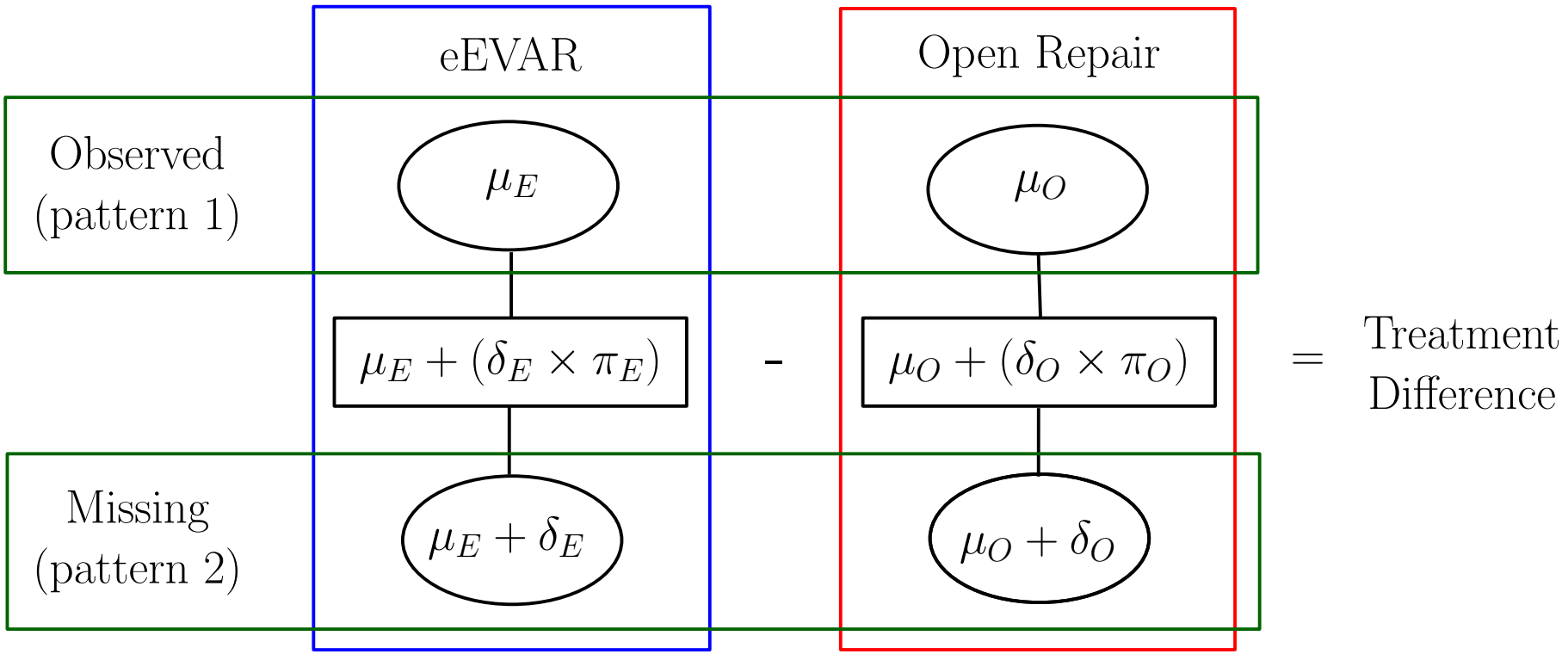
Pattern mixture models

- Two factorisations of the joint model are commonly used
  1. selection models
     $f(\boldsymbol{z}, \boldsymbol{m}|\beta, \theta) = f(\boldsymbol{m}|\boldsymbol{z}, \beta, \theta) f(\boldsymbol{z}|\beta, \theta)$
  2. pattern mixture models
     $f(\boldsymbol{z}, \boldsymbol{m}|\beta, \theta) = f(\boldsymbol{z}|\boldsymbol{m}, \beta, \theta) f(\boldsymbol{m}|\beta, \theta)$

  see Molenberghs *et al.* (2015) for further details
- We have chosen to use pattern mixture models
  - allow a different model for $\boldsymbol{z}$ for each pattern of missing values
  - assumptions about the missing data are more explicit
  - corresponds more directly to what is actually observed (i.e. the distribution of the data within subgroups having different missing data patterns)

Estimating a treatment difference using a PMM

Example from the IMPROVE trial



μ = mean QoL for patients who returned their QoL questionnaires
δ = difference in mean QoL between patients with observed and missing QoL
π = proportion of patients with missing QoL
E = eEVAR treatment group; *O* = open repair treatment group

Pattern mixture model equations (IMPROVE)

Pattern 1: patients who completed their QoL questionnaire:

$$QoL_i \sim N(\mu_i, \sigma^2)$$

$$\mu_i = \eta_{t(i)} + \beta_a age_i + \beta_s sex_i + \beta_{h,hardman(i)}$$

$\eta_1, \eta_2, \beta_a, \beta_s, \beta_{h,1}, \ldots, \beta_{h,5}, \sigma^2 \sim$ prior distribution

$t(i)$ trial arm indicator of individual i

age and *sex* are fully observed

Hardman index (*hardman*) is partially observed

$\Rightarrow$ covariate imputation model required

Pattern 2: patients who did not complete their QoL questionnaire:

$$\mu_i = \eta_{t(i)} + \delta_{t(i)} + \beta_a age_i + \beta_s sex_i + \beta_{h,hardman(i)}$$

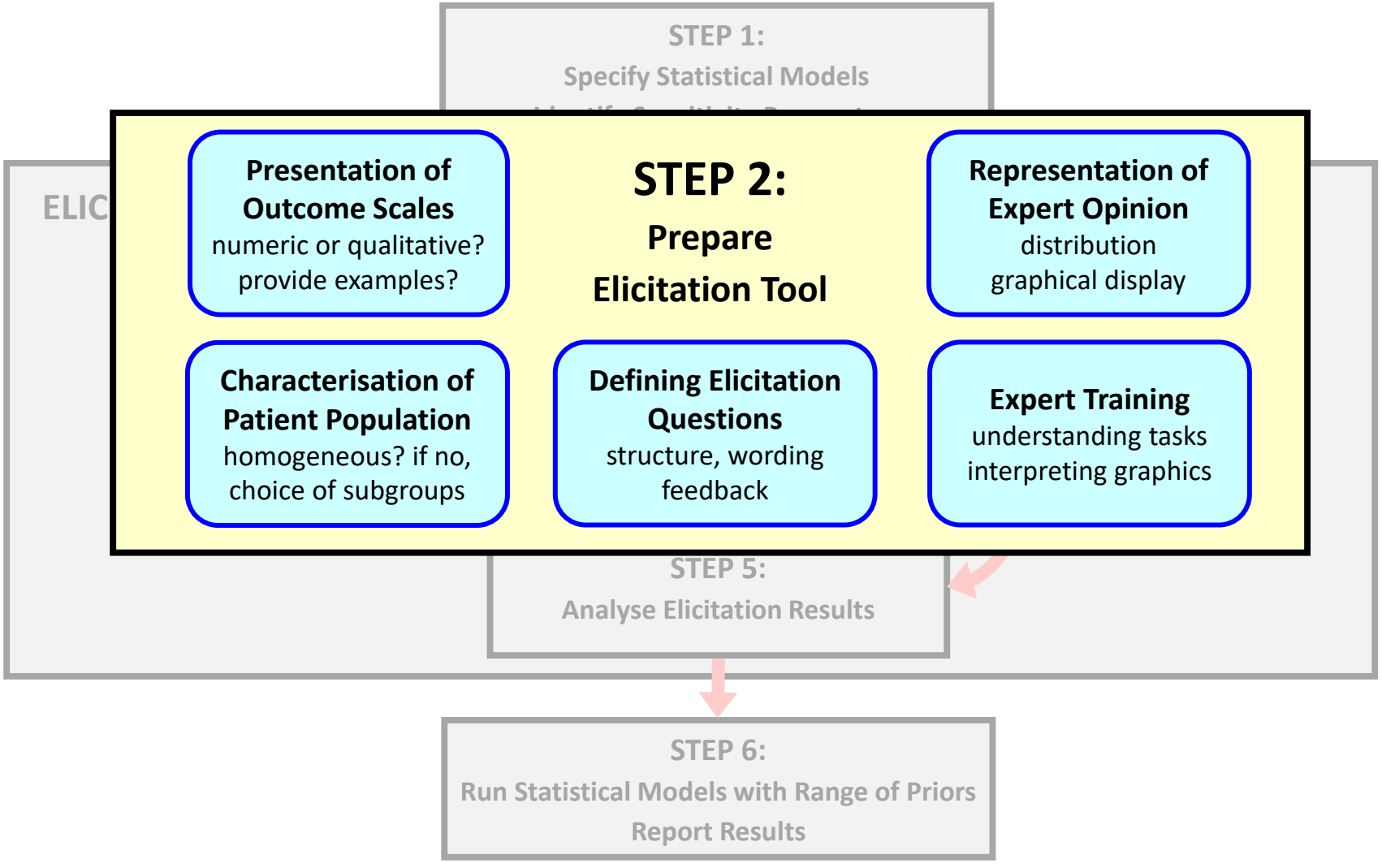
What information is required from experts?

- δ are unidentifiable sensitivity parameters
  - relate to difference in mean score between patients who did and did not return their questionnaire
- Comparisons made for patients who are similar according to observed characteristics
- To convert expert opinion about this difference into an informative prior, we must elicit
  1. the value that the expert believes is most likely
  2. the expert's uncertainty about this value

What information should be elicited from experts?

- In the context of a typical RCT, require
  - estimates by randomised arm
  - estimate of the correlation between treatment arms
- Allows specification of a joint prior
- For IMPROVE, we elicited the difference in mean QoL score for patients who did NOT return their questionnaire compared to those whose questionnaire was returned for:
  1. patients in OPEN REPAIR arm (marginal distribution)
  2. patients in eEVAR arm (marginal distribution)
  3. patients in eEVAR arm given the difference in the OPEN REPAIR arm (conditional distribution)

Framework: Step 2 - Prepare Elicitation Tool



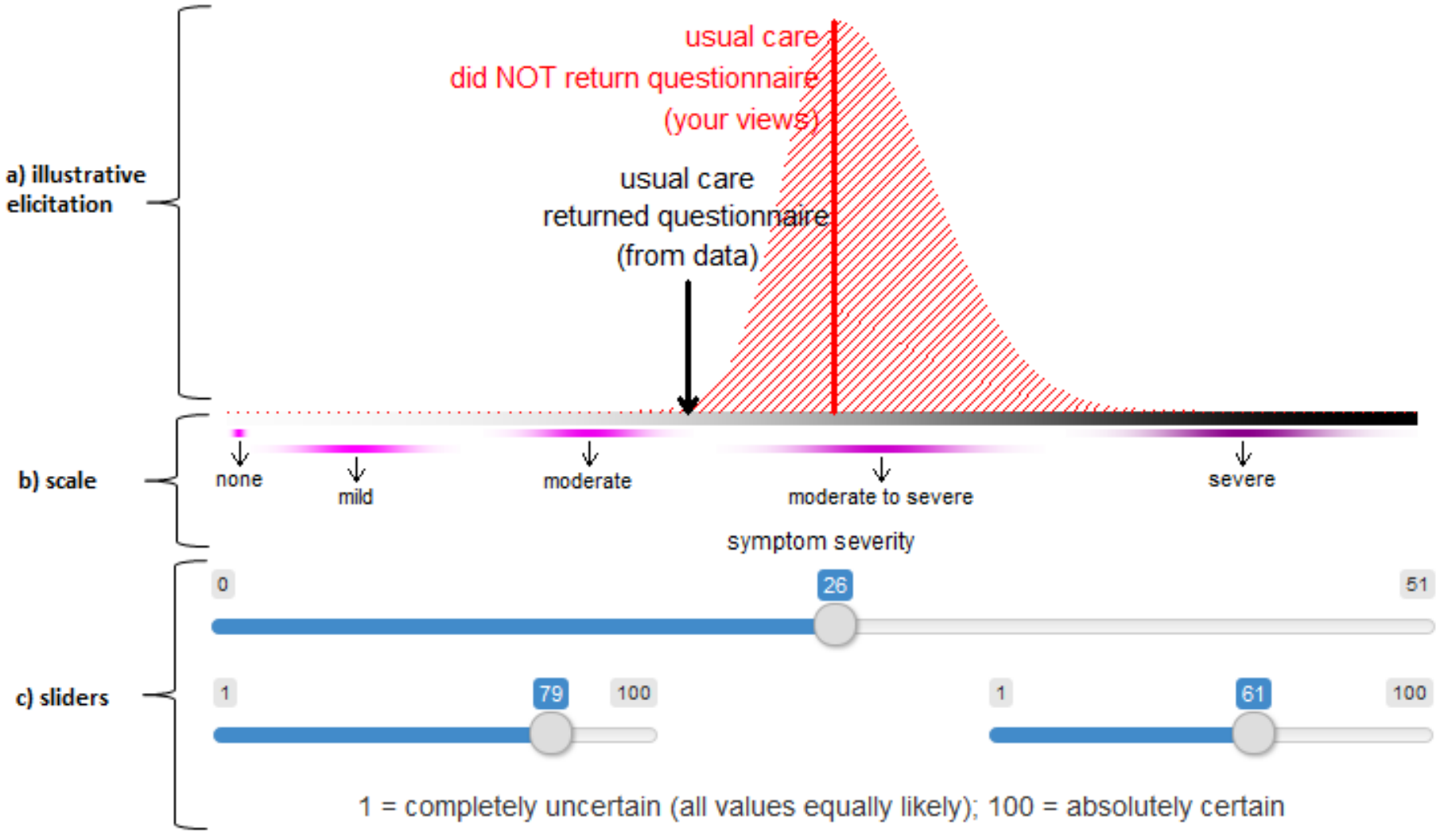
Step 2: Presentation of Outcome Scales

- Can develop easy-to-use web based elicitation tool using Shiny
  - a web application framework within the statistical software, R
- Adopting a graphical approach
  - helps experts represents their views intuitively
  - naturally produces prior distributions for statistical models
- Outcome scales
  - must be interpretable to non-statistical expert
  - can be customised using descriptive labels and/or examples
- Examples shortly

Step 2: Representation of Expert Opinion

- Tension between
  - sufficient flexibility to represent expert beliefs accurately and
  - over-complicating the elicitation exercise and subsequent analysis
- In broad terms, choice is
  - parametric distributions, e.g. normal distribution, often calculated from elicited quantiles
  - non-parametric distributions, typically represented as a histogram and elicited using ‘chips and bins’
- What shapes could distributions representing expert uncertainty realistically take?
  - individual scores plausibly bimodal - ‘too sick’ and ‘too well’ may not respond
  - but expert’s uncertainty about average scores will be uni-modal
- Other considerations - bounded scales

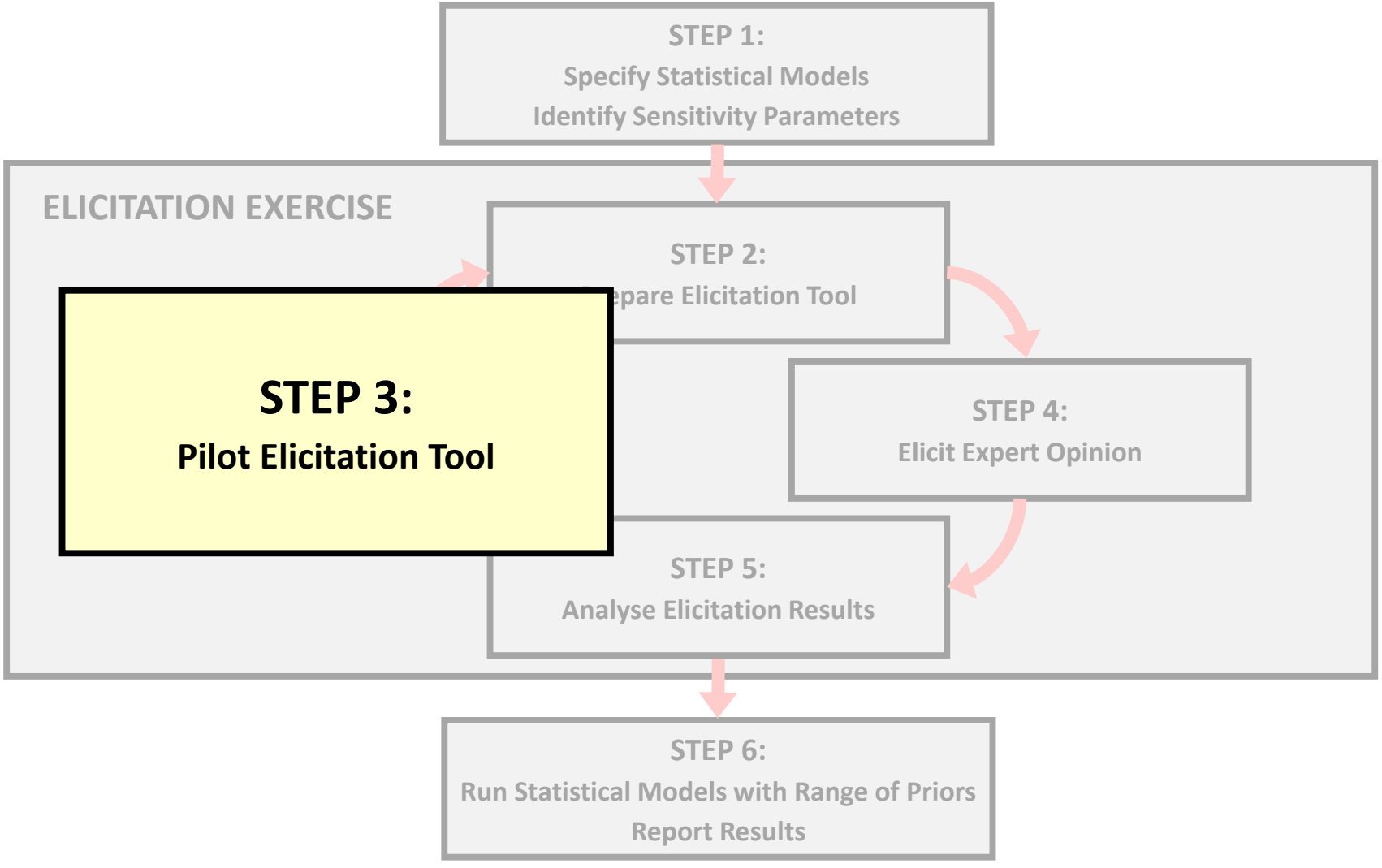
Step 2: Outcome scale for PSS-SR scores



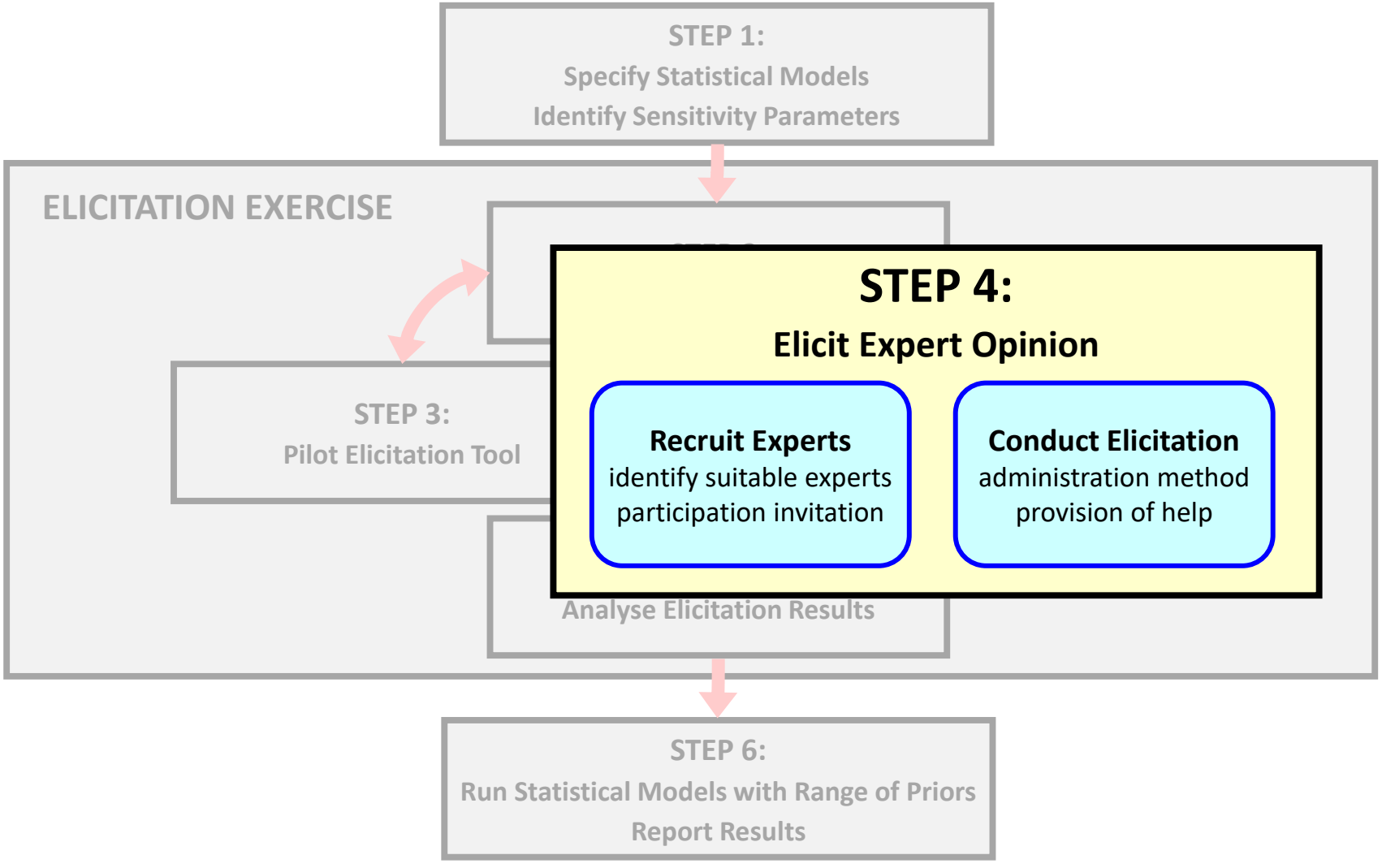
Step 2: Other important elements

- Characterisation of patient population
  - is the patient population homogeneous?
  - should we elicit by sub-group?
- Defining elicitation questions
  - consider structure, wording, use of feedback
  - crucial wording ensures expert provides views about average, not individual scores
  - include free text questions - useful for assessment
- Expert training - do the experts
  - understand the elicitation tasks?
  - correctly interpret the graphics?

Framework: Step 3 - Pilot Elicitation Tool



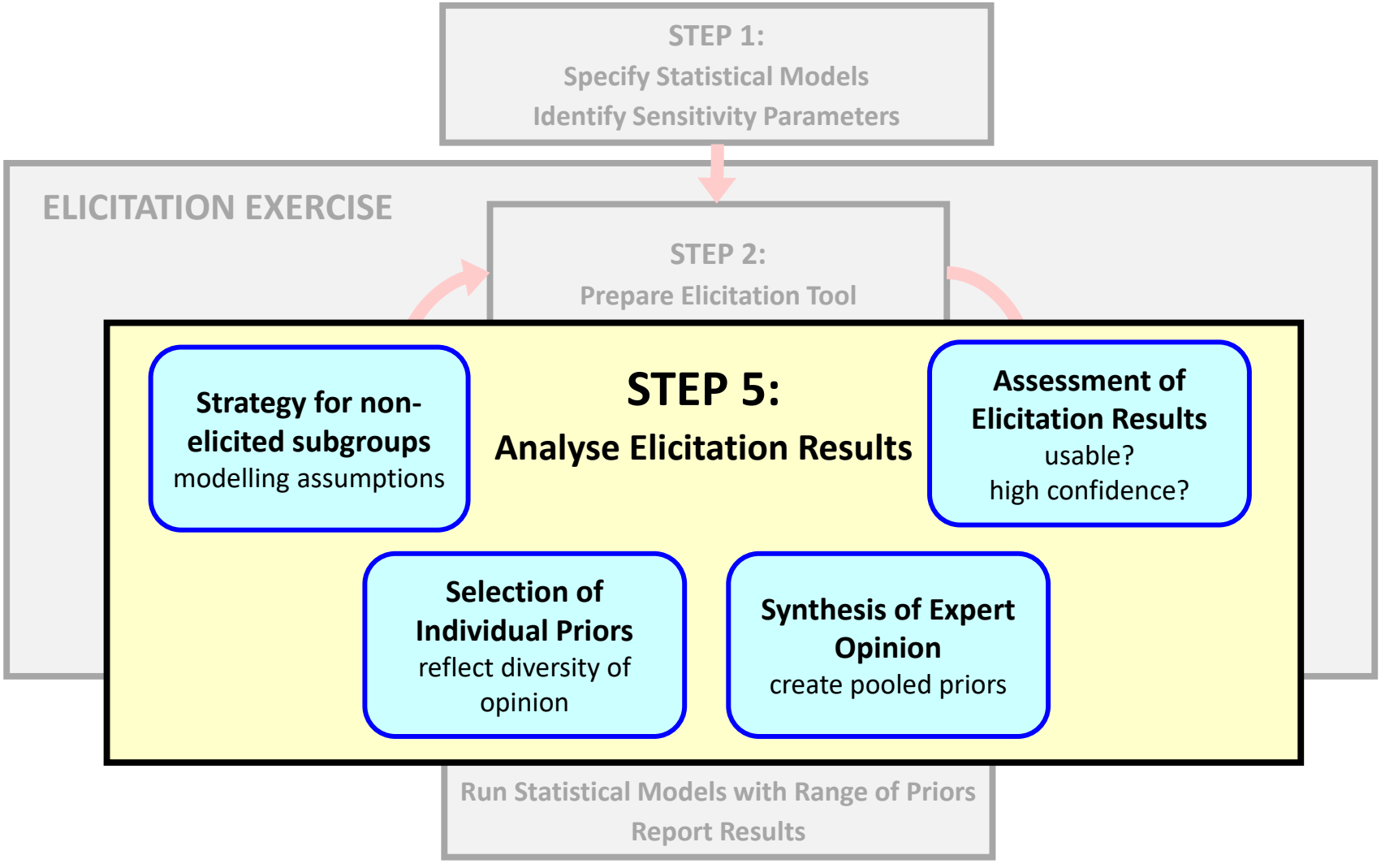
Framework: Step 4 - Elicit Expert Opinion



Step 4: Recruit Experts and Conduct Elicitation

- Recruit Experts
  - require good coverage of experts with interest in the research questions
  - identify experts with relevant knowledge about trial outcomes, beyond that recorded in the data
  - chief investigator emails participation invitation to potential respondents
- Conduct Elicitation
  - in-person (conference breaks) or remote (send web link)
  - consent taken electronically
  - reminders emailed

Framework: Step 5 - Analyse Elicitation Results



Step 5: Strategy for non-elicited subgroups

- Following issues considered in Step 2
  - Should we elicit by subgroup?
  - If so, how should we define the subgroups?
  - If impractical to elicit for all subgroups, which should we choose?
- How should we treat non-elicited subgroups?
  - assign a prior or mixture of priors from elicited subgroups
  - assume a priori observed differences in distributions of sub-groups carry through to differences in the priors
- How should we deal with subgroup correlation?

Step 5: Assessment of Elicitation Results - Cooke

- Crucial question: do the experts understand the questions?
- Colson and Cooke (2018):

  validation of expert judgements means both that the judgements reflect the beliefs of the expert and that those beliefs reflect reality

- Cooke's 'classical' method compares elicited judgements to observed data using seed (calibration) questions
- Finding suitable seed questions
  - recognised as difficult
  - adds substantially to demands on experts' time

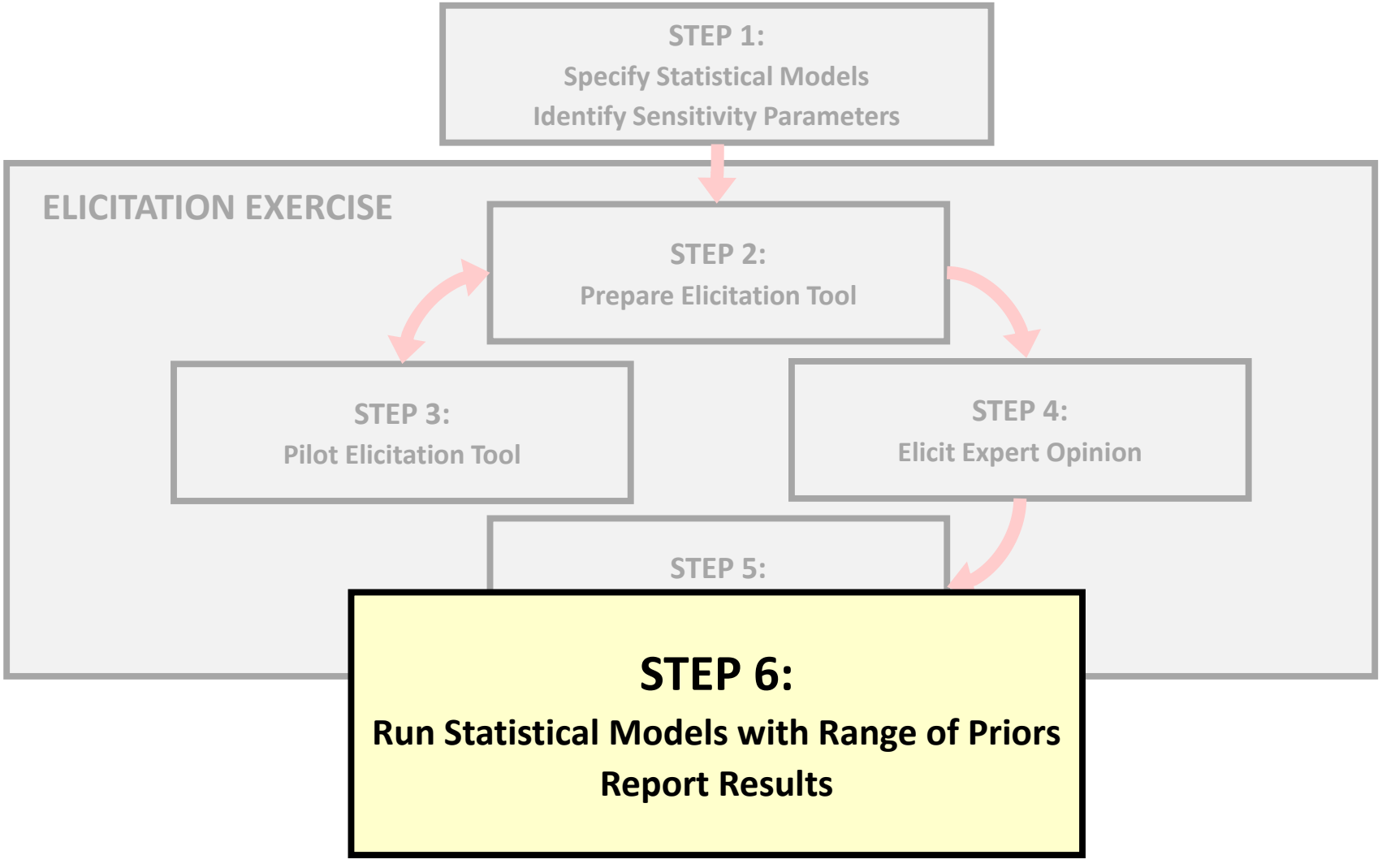
Step 5: Assessment of Elicitation Results - alternative

- Use criteria based on a combination of the experts' quantitative and qualitative responses to identify:
  - experts who have provided 'usable' responses
  - subgroup in whom we have 'high confidence'
- 'usable' experts
  - exclude experts who clearly misunderstood the task
  - multiple unmoved sliders or evidence from qualitative responses
- 'high confidence' experts
  - evidence of high level of engagement with elicitation exercise from qualitative answers
  - consistency between quantitative and qualitative answers
  - consistency in approach to all subgroups

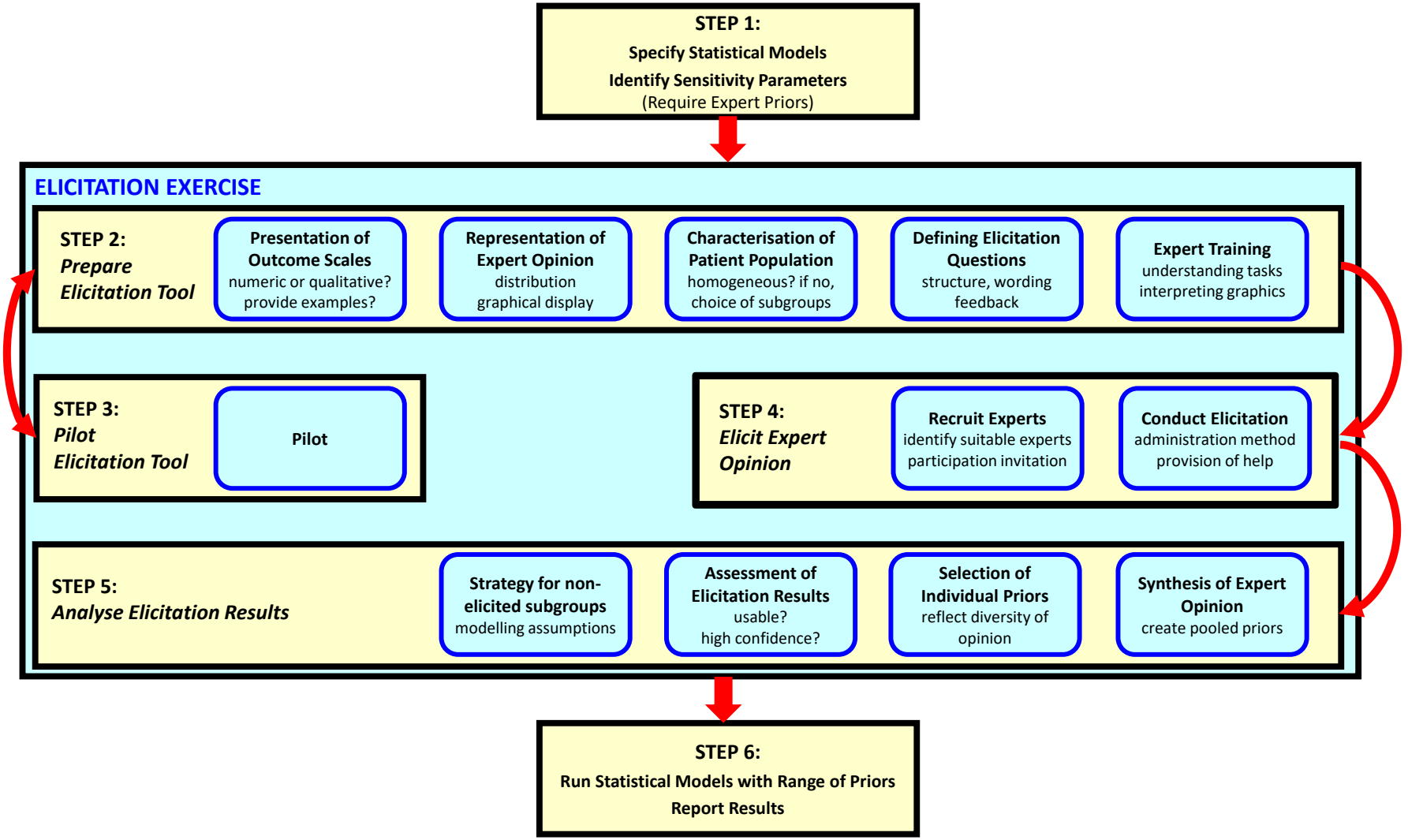
Step 5: Selection of Individual Priors and Synthesis of Expert Opinion

- Individual priors:
  - choose to reflect diversity of opinion
  - e.g. ‘most enthusiastic’ and ‘most sceptical’
- Pooled priors:
  - average of individual distributions for multiple experts
  - linear pooling, with equal weights
  - specified as a mixture of distributions
  - e.g. all ‘high confidence’ experts, all doctors

Framework: Step 6 - Run Statistical Models



Framework in Detail

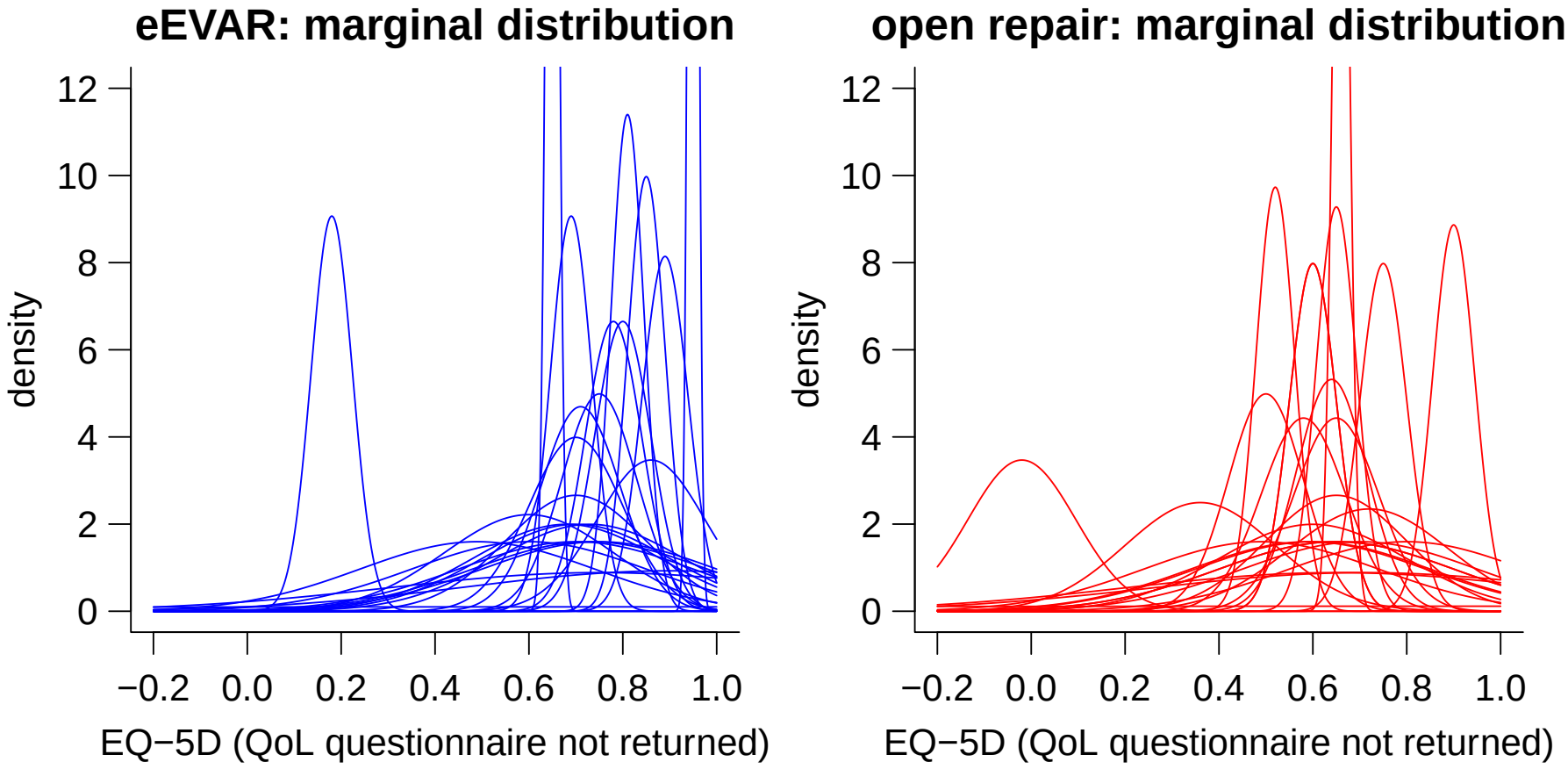


Elicitation Tool DEMO

The experts

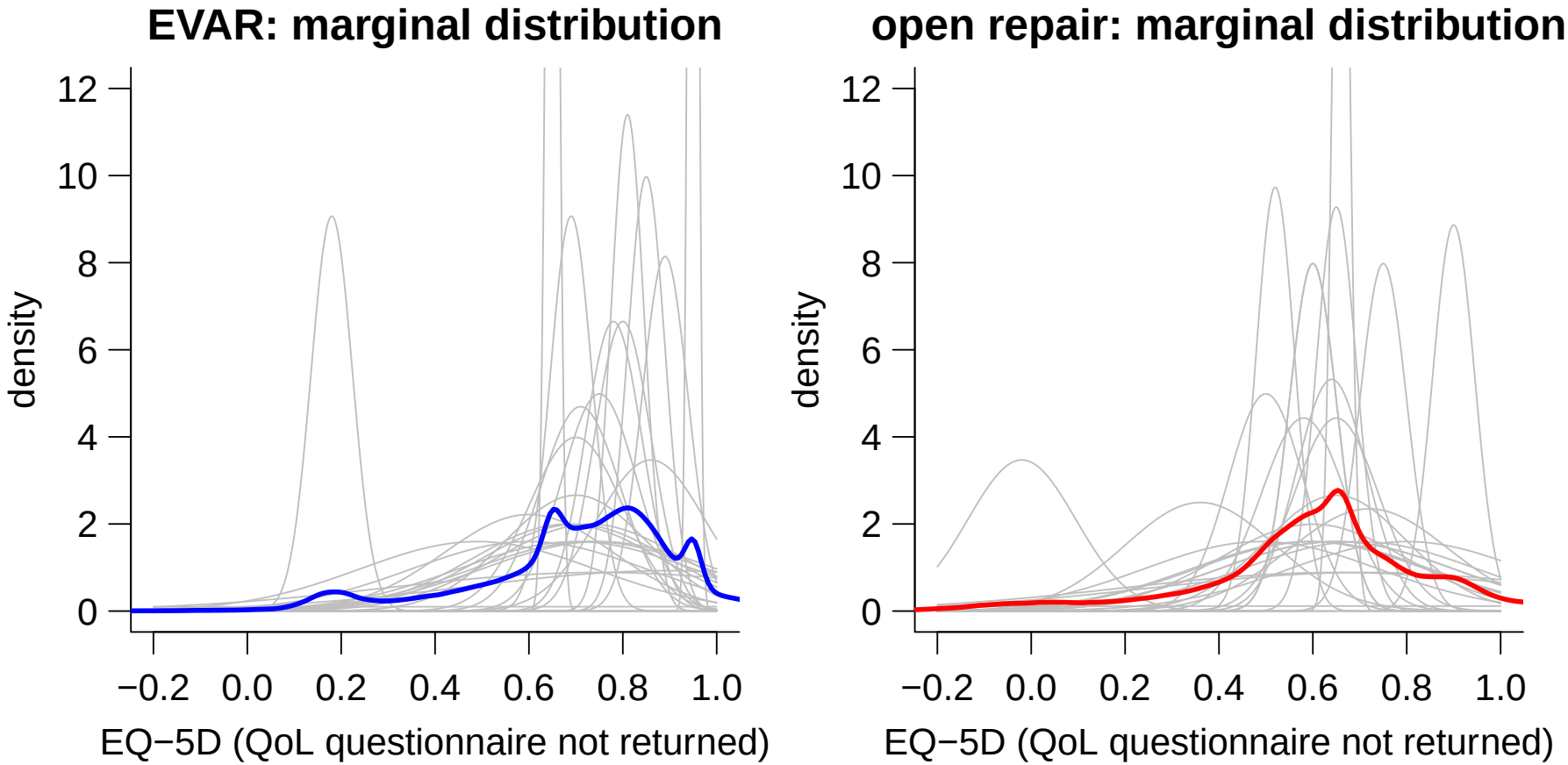
- IMPROVE:
  - 46 potential experts, with ongoing trial involvement, identified
    - principal investigators (mainly consultant vascular surgeons)
    - trial co-ordinators (vascular nurse specialists or research nurses)
  - 26 responses submitted, 15 at conference (The Vascular Society Annual Scientific Meeting in November 2015)

IMPROVE: individual expert priors



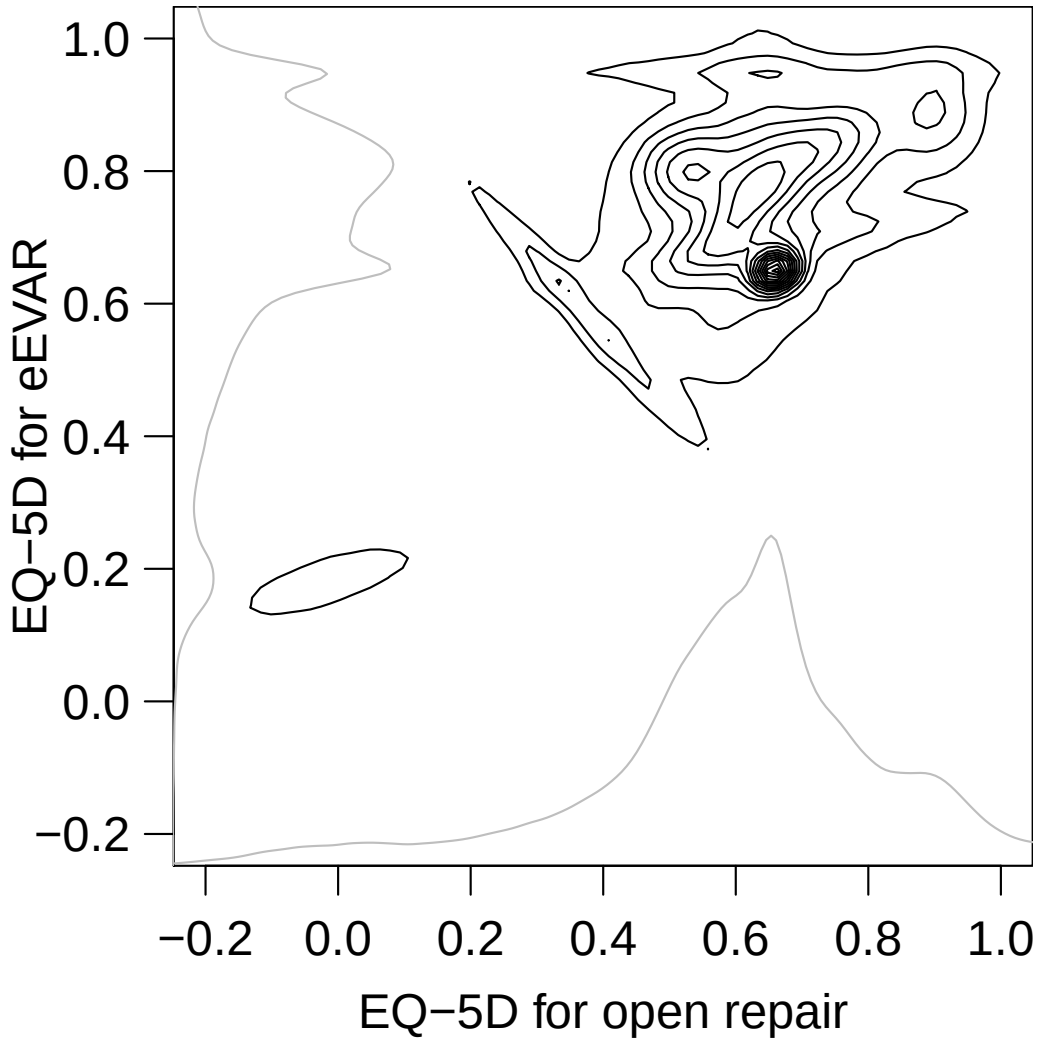
- Elicited QoL scores for patients with missing data lower on average than observed QoL scores
- But considerable uncertainty in these elicited values

IMPROVE: pooled marginal distributions



- Views of all experts combined using mathematical aggregation
  - linear pooling with equal weights
  - mixture of normal distributions

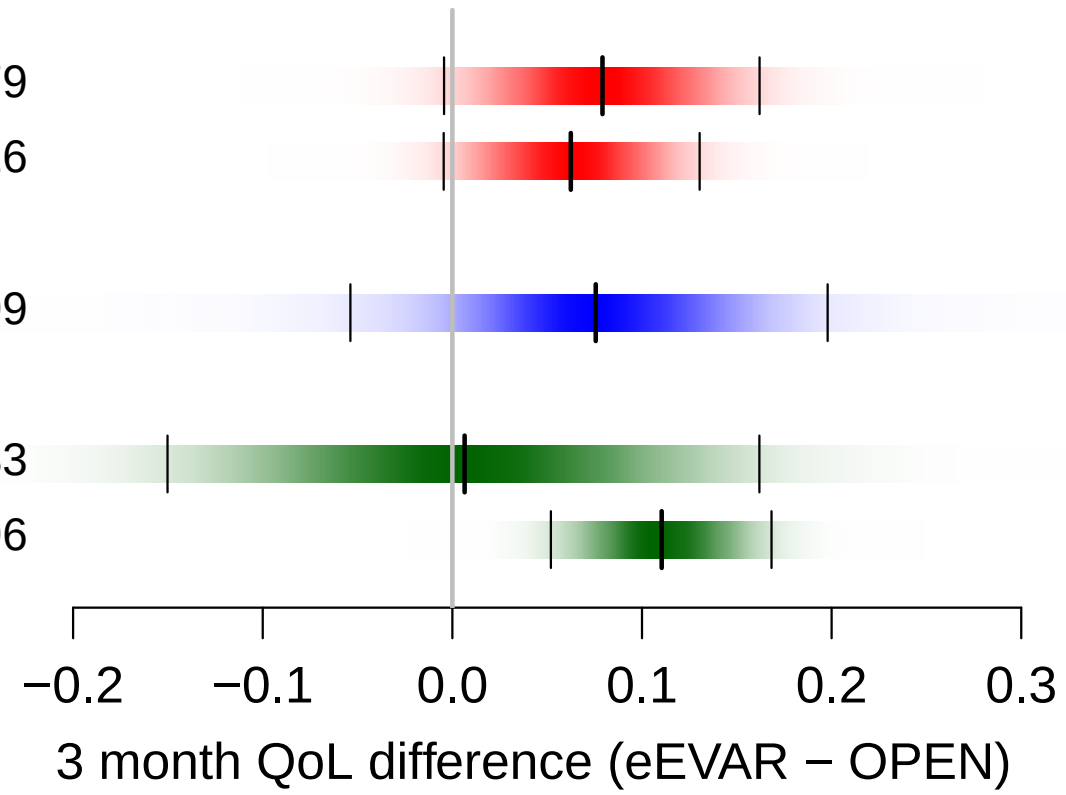
IMPROVE: joint pooled prior



- Allows values in the two arms to be related

IMPROVE: difference in mean QoL at 3 months

	Probability of finding a clinically relevant difference in mean QoL between arms*
Complete Case Analysis	0.879
Primary analysis (MAR)	0.826
Sensitivity analysis (MNAR)	
all experts	0.799
Extreme sensitivity analysis (MNAR)	
sceptical expert	0.383
enthusiastic expert	0.996



*0.03 is considered the minimum clinically important difference (Walters and Brazier, 2005)

Comment

- Successfully demonstrated a practical approach for
  - eliciting expert opinion from a range of experts
  - conducting MNAR sensitivity analysis
- Allows more realistic assumptions to be made about the missing data
- It allows us to quantify how experts would interpret the study
  - taking into account their understanding of the differences between the observed and missing values
- Models naturally extend to more complex settings

Challenges and extensions

- Address concerns raised by Heitjan (2017)
  - evaluate framework using structured qualitative interviews
  - improve training and feedback
  - extend approach to assessment of elicitation results
- Extension to multiple time-points
  - number of sensitivity parameters can quickly become large with increasing time-points
  - simplifying assumptions may be required - focus on key parameters
  - discuss with subject-matter experts which parameters are most relevant
- Sub-groups - best strategy unclear
- Generalise to other types of trials and endpoints

Acknowledgements

- Pinar Ulug (IMPROVE Trial Manager) and Janet Powell (IMPROVE Chief Investigator)
- David Harrison, Paul Mouncey, Alvin Richards-Belle and Kathy Rowan (ICNARC)
- All the IMPROVE and POPPI 'Experts' who made this research possible
- LSHTM staff who assisted with pre-piloting

References

- Bell, M. L., Fiero, M., Horton, N. J., and Hsu, C.-H. (2014).
  Handling missing data in RCTs; a review of the top medical journals.
  BMC Medical Research Methodology, **14**, 118.
- Colson, A. R. and Cooke, R. M. (2018).
  Expert Elicitation: Using the Classical Model to Validate Experts' Judgments.
  Review of Environmental Economics and Policy, **12**, (1), 113–32.
- Grieve, R., Gomes, M., Sweeting, M. J., Ulug, P., Hinchliffe, R. J., Thompson, M. M., Thompson, S. G., Ashleigh, R., Greenhalgh, R. M., and Powell, J. T. (2015).
  Endovascular strategy or open repair for ruptured abdominal aortic aneurysm: one-year outcomes from the IMPROVE randomized trial.
  European Heart Journal, **36**, (31), 2061–9.
- Heitjan, D. F. (2017).
  Commentary on Mason et al.
  Clinical Trials, **14**, (4), 368–9.
- Mason, A. J., Gomes, M., Grieve, R., Ulug, P., Powell, J. T., and Carpenter, J. (2017).
  Development of a practical approach to expert elicitation for randomised controlled trials with missing health outcomes: Application to the IMPROVE Trial.
  Clinical Trials, **14**, (4), 357–67.
  DOI: 10.1177/1740774517711442.
- Molenberghs, G., Fitzmaurice, G., Kenward, M. G., Tsiastis, A., and Verbeke, G. (ed.) (2015).
  Handbook of Missing Data Methodology. Chapman & Hall.
- Walters, S. J. and Brazier, J. E. (2005).
  Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D.
  Qual Life Res, **14**, (6), 1523–32.