Centre for Health Economics, University of York

Health opportunity costs in the NHS: assessing the implications of uncertainty using elicitation methods with experts

Work funded by the Department of Health's Policy Research Unit in Economic Evaluation of Health and Care Interventions

Protocol

1. Background

1.1 Recent research in the UK developed a framework of analyses for estimating the expected health opportunity costs for the NHS. The study quantified how changes in the NHS expenditure in a given year are allocated between disease areas, and how changes in expenditure by disease area in a particular year affect disease specific mortality rate in that year (all else being equal). [Claxton et al. 2015, Martin et al. 2008]. However, the benefits of changes in NHS expenditure cannot be assumed only to relate to an effect on mortality rate in the same year of expenditure. There should also be consideration of the possibility of (A) changes in expenditure having an effect beyond the year of expenditure; (B) how the effects of changes in expenditure on mortality relate to effects on a broader measure of health that incorporates both duration and quality of life impacts (quality adjusted life years); (C) how changes in expenditure affect health in disease areas for which previous work could not measure a mortality effect; and (D) the additional life-years associated with any reduction in the rate of mortality. There are only very limited data available, if any, on each of these items.

1.2 In the absence of formal empirical evidence, however, it is important that the estimate of the health opportunity costs in the NHS, suitable to support policy decisions, reflects the judgements on these key quantities of individuals close to policy.

1.3 To get quantitative expressions of belief on these different uncertainties we propose an elicitation exercise. Elicitation is a systematic process for formalising and quantifying, typically in probabilistic terms, expert judgements about uncertain quantities [O'Hagan 2006]. Such structured approach to the design and conduct of the elicitation aims to minimise the use of cognitive heuristics when formulating judgments in the face of uncertainty, and avoid the biases that this form of evidence is known for.

1.4 The quantities elicited aim to fully reflect uncertainty in knowledge of each expert. This allows a comprehensive assessment of the implication of uncertainty both within and between experts. The primary output of the elicitation exercise is thus a set of subjective probability distributions.

1.5 This document is a protocol for the elicitation exercise that will cover aspects related to the design (Sections 2 to 6), conduct (Sections 7 to 8) and analysis and feedback (Sections 10 to 12).

2. Summary of the approach to elicitation

2.1 We propose two groups of experts. Firstly, policy experts: individuals drawn from organisations that develop or implement policy, or that have a major interest in policy in this area; but that are not expected to have specific substantive expertise in key clinical areas. Secondly, clinical experts acting as substantive experts in key disease areas.

2.2 The approach to the elicitation we pursue here is to ask policy experts to elicit their judgement on the quantities of interest once they have considered the information from clinical experts. This approach means that policy experts' elicited judgements will combine the judgements from the different substantive (clinical) experts and their own beliefs – implicitly weighting these. In order words, the policy experts' judgements will reconcile their own *prior* beliefs with the 'data' from clinical experts and in this way distributions elicited can be considered *posterior* distributions (within a Bayesian framework).

2.3 The practical implication of this approach is that judgements from clinical experts need to be elicited prior to those of policy experts, at a first workshop. At the final workshop, the judgements from clinical experts will be summarised and presented to policy experts to help them formulate their judgements. The same elicitation tool will be used in both workshops.

3. What do we want to elicit?

3.1 The work by Claxton et al. [2015] identified the key structural assumptions to estimates of the NHS expected opportunity costs:

A: Duration of mortality effects of expenditure in a particular year

B: Whether the impact of expenditure on the mortality burden of disease is a reasonable surrogate for its impact on morbidity (in disease areas where mortality effects could be identified)

C: Whether the proportionate effect of expenditure on QALY burden of disease is likely to be similar in those disease areas for which mortality effects could not be identified (extrapolation).

D: The survival effects of the estimated mortality effects

We will refer to these as the parameters of interest throughout. These are not necessarily elicited directly as there are often alternative quantities that can more feasibly and usefully inform such parameters – the quantities to elicit are considered in more detail in Section 5.

3.2 Whilst the parameter values are very likely to differ between disease areas, the research team deemed burdensome for the experts to elicit for each of the 23 programme budget categories, PBCs, identifying disease areas. Seven disease areas (detailed in point 4) were

selected for being important drivers of the central estimate of health opportunity costs in the NHS, and sensitive to the surrogacy and extrapolation assumptions, across more recent time waves of analyses (unpublished EEPRU work). Both policy and clinical experts will be asked to elicit all parameters separately for each of these seven main PBCs, and a pooled estimate for the remaining PBCs. Clinical experts will be asked to record their area of expertise, which will be considered in summarising results for policy experts.

4. Which experts?

4.1 For the reasons detailed above (point 2) two subsets of experts are to be recruited: policy and clinical experts. Column 1 in Table 1 details the organisations from which we will invite policy experts. We suggest that a total of 15-20 policy experts would be appropriate for this exercise (at least 2 from each organisation).

Table 1: Suggested policy organisations and areas of expertise for clinical experts

Organisations from which policy experts are drawn	**Clinical areas from which clinical experts are drawn**
Department of Health	Circulatory
NHS England (NHSE)	Respiratory
Public Health England (PHE)	Gastrointestinal
National Institute for Health and Care Excellence (NICE)	Neurological
Joint Committee on Vaccination and Immunisation (JCVI)	Mental health
NHS Clinical Commissioners (NHSCC)	Endocrinology
Association of the British Pharmaceutical Industry (ABPI)	Musculoskeletal
Patients' organisations	Primary care

4.2 Clinical experts from the 7 selected clinical areas will be identified (item 3.2 and Table 1). Primary care has been added to the list for the broad expertise these individuals have of all clinical areas. Recruitment will aim for a total sample size of 15-20 clinical experts, 2 experts from each clinical area and primary care experts overrepresented in relation to others. Recruitment efforts will target clinicians who have participated on decision making committees (e.g. NICE guidelines and appraisal committees), or that are or have been significantly involved in research activities.

4.3 Responses from experts will be anonymous (see point 13) but we will record the organisation they belong to (policy experts) and their clinical area of expertise (clinical experts). This will give us opportunity to explore between-expert heterogeneity (see 9.2).

5. What quantities to elicit?

5.1 The quantities elicited will attempt to gather experts beliefs on each of the four assumptions detailed in 3.1. The choice of the quantities will consider the following three objectives [Soares 2011]: i) fitness for purpose; ii) quantities elicited should be directly observable as it

may be difficult for experts to formulate beliefs on quantities that are not directly observable [Kadane 1998]; iii) homogeneity in the quantities elicited – eliciting the same summaries throughout will reduce the burden of training.

Section A. Duration of mortality effects of expenditure in a particular year.

5.2 Section A aims to elicit the duration and magnitude of mortality effects of expenditure in a particular year. This relates to expenditure in a particular year and its effect on years thereafter; it does not include additional effects of expenditure in other years.

5.3 This quantity is only elicited for disease areas for which previous research was able to identify a measurable effect of changes in expenditure on mortality rates.

5.4 We will elicit using a two part question. The first part (question A1) asks experts to directly report their beliefs over the expected duration of mortality effects beyond the first year ($\bar{T}$, years). This indicates duration but does not provide information on the magnitude of effect in each year. This quantity ranges between 0 and +infinity.

5.5 The second part (question A2) asks the experts to report the mortality effects in subsequent years as a proportion of the effect on the 1st year. A limited number of time points is going to be elicited: 2nd, 3rd and 4th years. The elicited quantities can range between 0 and +infinity, where a value of 0 indicate no effect in a particular year, a value of 1 the same effect to the first year and a value bigger than 1 a bigger effect than in the first year.

Section B. Surrogacy.

5.6 This section relates to establishing experts' beliefs on how the effects of increased expenditure on health burden compare with its effects on mortality burden (surrogacy relationship). Such broad measure of health burden considers the impact of increased expenditure on both the rate of mortality (and any life-years lost as a consequence) and on the level of health-related quality of life of individuals – i.e. per patient quality-adjusted life-years.

5.7 This quantity is only elicited for disease areas for which previous research was able to identify a measurable effect of changes in expenditure on mortality rates.

5.8 This will be elicited as a proportionate effect of a change in expenditure in total health burden in relation to the proportionate effect on mortality burden, and will be elicited separately for the year of expenditure (1st year) and subsequent years (2nd, 3rd and 4th years).

5.9 This quantity ranges from 0 to +infinity, where a value of zero indicates no reduction in health burden, a value of 1 indicates that the proportionate reduction in health burden is the same as that in mortality burden; values between zero and 1 indicate that the proportionate effect of increased expenditure on health burden is less than that on mortality burden; and values higher than 1 indicate that the proportionate effect of increased expenditure on health burden is more than that on mortality burden.

Section C. Extrapolation.

5.10 This section asks experts to compare (proportionate) reductions in health burden (quality-adjusted life-years) from an increase in NHS expenditure in specific disease areas that did not have measurable mortality effects with (proportionate) reductions in health burden from increased expenditure across all disease areas that had measurable mortality effects (i.e. an average effect across all disease areas in this group).

5.11 This quantity is only elicited for disease areas for which previous research was unable to identify a measurable effect of changes in expenditure on mortality rates.

5.12 This will be elicited separately for the year of expenditure (1st year) and subsequent years (2nd, 3rd and 4th years).

5.13 This quantity ranges from 0 to +infinity, where a value of zero indicates that a specific disease areas without measurable mortality effects from increased expenditure (e.g. mental health disorders) would not see health burden reduced from increased expenditure; a value of 1 indicates that the effects on health burden of increased expenditure in the specific disease area under consideration (e.g. mental health disorders) would be equal to that observed across disease areas with measurable mortality effects from increased expenditure; values between zero and 1 indicate that the proportionate effect of increased expenditure on health burden is expected to be less in the specific disease area being considered (e.g. mental health disorders) than in those with measurable mortality effects. For example, a value of 0.5 would indicate an expected reduction in health burden in 'mental health disorders', for example, of half that in disease areas with measurable mortality effects; values higher than 1 indicate that the proportionate effect of increased expenditure on health burden is expected to be more in the specific disease area being considered (e.g. mental health disorders) than in those with measurable mortality effects. For example, a value of 1.5 would indicate an expected reduction in health burden in 'mental health disorders' that is 1.5 times that in disease areas with measurable mortality effects.

Section D. Mortality effect on survival.

5.14 This relates to evaluating the expected life years gained from averting a death.

5.15 This quantity is to be elicited for disease areas for which previous research was able to identify a measurable effect of changes in expenditure on mortality rates.

5.16 The analysis in Claxton et al [2015] used 3 years of mortality data which were first averaged before being used in the econometric analysis. This means the estimate of mortality effects includes deaths averted for at least one year (for patients in the 3rd year of the mortality data), deaths averted for at least 2 years (for patients in the 2nd year of the data) and deaths averted for at least 3 years (for those in the 1st year of the data).[Claxton et al 2015, Section 4.1 and 4.2.5 as well as 1 in Box 5.1 and references to text and footnotes]. We will elicit separately for each.

5.17 We will elicit this parameter using a two part question. The first part (question D1) asks experts: Of those patients that have seen their deaths averted by at least 3 years, 2 years and

1 year, what proportion are likely to return to (or exceed) the life expectancy of the general population of the same age and gender?

5.18 The second part asks the expert to consider only those patients that have not returned to, or exceeded, normal life expectancy (age and gender matched) and report on their life expectancy (as a proportion of the life expectancy in the general population, again, matched by age and gender). This will also be elicited separately for those patients who have survived for at least 1 year, 2 years and 3 years. This quantity can vary between 0 and 1, where 0 represents immediate death and 1 the life expectancy of the matched normal population.

6. How to elicit?

Method of elicitation

6.1 The aim of this exercise is to reflect uncertainty by eliciting probability distributions. Distributions are continuous and thus to fully specify them an infinite number of summaries would need to be asked of each expert. Methods of elicitation thus focus on asking for a (usually quite small) number of summaries, and then fitting a distribution (methods described in point 10).

6.2 We will elicit three summaries. The central tendency summary chosen was the mode (value the expert believes to be most likely, their best guess). We believe experts can more easily report the mode than the mean or median.

6.3 We will also elicit within-expert uncertainty using the lower and upper 80% credible interval. This corresponds to a variable interval method (where the expert identifies points that correspond to specified percentiles of his or her subjective distribution). Evidence shows that while eliciting confidence intervals is easy and intuitive, there is a clear tendency for central credible intervals to be too short (a bias is called overconfidence), i.e. people believe they are more accurate than is justified [Garthwaite 2005]. We recognise this limitation, but considered the time constraints of the experts we aim to engage.

6.4 We looked into the evidence on overconfidence when eliciting confidence intervals and identified that single limit estimates (i.e. asking for the lower bound first, and then the upper bound separately) are suggested to produce wider estimated than asking directly for the range [Teigen 2005]. The proposed wording for describing the summaries being elicited for each quantity is thus:

(1): My best guess for the value of this quantity is ….

(2): I am very certain (90% certain) that the true value for this quantity is <u>higher</u> than...

(3): I am very certain (90% certain) that the true value for this quantity is <u>lower</u> than...

Judgements of clinical experts to be presented to policy experts

6.5 We will allow policy experts to visualise (at their choice): 1) the judgements of each expert with clinical expertise in the area, 2) the pooled judgements of the experts in 1), or 3) the pooled judgements of all clinical experts (see point 9 for method of pooling)

7. Elicitation tool

7.1 We will use a paper questionnaire.

7.2 The wording of the questions will be piloted for clarity and adequacy.

8. Objectives of the training of experts

8.1 The training session for experts will aim to: describe the objectives of the elicitation exercise, clarify concepts such as those of uncertainty, variability and heterogeneity, familiarise the experts with the quantities we wish to elicit, describe and explain the impact of bias and heuristics, and train experts on the method of elicitation used.

9. Approach to pooling judgements from multiple experts

9.1 When several experts supply probability distributions, their responses are often pooled so as to derive a single distribution. We will pool analytically the individual answers (mathematical elicitation), instead of using a consensus technique. This is because consensus methods are known to have a number of limitations (e.g. implicit aggregation, effect of dominant individual on group dynamics, known problems where consensus returns overly-precise judgements).

9.2 The overall results of the exercise will be examined for the following groupings defined here (a priori): 1) all experts pooled, 2) only clinical experts pooled, 3) only policy experts pooled. If the data is sufficient, we will further explore heterogeneity (i.e. between-expert uncertainty) by considering the area of specialty of clinical experts and the type of organisation policy experts belong to by considering Governmental Bodies (DH, NHSE, PHE), Non Departmental Public Bodies and Independent Departmental Expert Committees (NICE, JCVI), Local Commissioners (NHSCC), industry bodies (ABPI), and patient representatives. We will also explore heterogeneity according to self-reported judgements on the validity of their responses (point 11.1), by presenting results on each section that only include those experts that indicated their answers as valid (i.e. those that responded 'yes' to the question in point 11.1).

9.3 We will linearly pool across experts using equal weights. Linear pooling preserves the individual judgements in the collective (pooled) judgement. For example, if the experts' distributions for a single quantity are identical the pooled is equal to the individuals' distributions. Also, if there is the support from at least one expert that the quantity of interest

takes a particular values, the pooled distribution will also show some support for those values. [O'Hagan 2006, Dietrich 2014]

10. Fitting of distributions

10.1 Despite, for each quantity, only a limited number of summaries being elicited from each expert (mode and upper and lower bounds of credible interval), in analyses we will fit a distribution to represent uncertainty.

10.2 The process of fitting introduces uncertainty where possible fitted distributions may reasonably reflect the judgements from experts. We will aim to reflect this uncertainty by using the 3 summaries elicited from each expert on each quantity and fitting two alternative (two-parameter) distributions: one using the lower bound (LB) of the credible interval and the mode, and another using the upper bound (UP) and the mode. This will be dealt with as a source of structural uncertainty.

10.3 We aim to pre-specify the distributions being fitted and restrict the choice of distributions to those for which the fitting would not require simulation (i.e. those with closed-from solutions for the mean and cumulative density function). For quantities that can range between 0 and 1, Beta distributions will be used. There is a closed form solution for the mode. For quantities ranging between 0 and +infinity, the Log-Normal distribution will be used. There is also a closed form solution for the mode.

11. Assessing validity of the exercise

11.1 At the end of each section of the exercise (one section for each of the four parameters of interest), we will ask experts whether they are confident the answers they gave reflect their views and uncertainties. Response options will be 'yes', 'not sure' and 'no'. If they respond, 'no' or 'not sure', we will ask them to provide us with some more detail as to why in an open question.

11.2 At the end of the exercise, we will ask experts to provide any comments about aspects of the day using an open question.

12. Feedback of results

12.1 After all experts elicit the quantities of interest, we aim to feedback results of the implications of the judgements elicited for the central estimate of cost-effectiveness threshold and for estimates of the health opportunity costs of given investments.

12.2 We plan to do so two weeks after the final workshop with the policy experts using a short webinar consisting of approximately 15 min where results are presented and allows for 15

minutes of discussion. The webinar is to be recorded (audio + slides) and the recording circulated to those who were not able to attend.

13. Data protection and anonymity

13.1 Experts will be asked to give their opinions individually (not in groups). The information provided, including personal details, will be kept anonymous and confidential, stored securely and only accessed by those carrying out the study.

References

Alpert M, and Raiffa H. A Progress Report on the Training of Probability Assessors. 1969 Unpublished Report.

Claxton K, Martin S, Soares M, Rice N, Spackman E, Hinde S, Devlin N, Smith PC, Sculpher M. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. Health Technology Assessment; 19(14). 2015.

Cooke RM. Experts in Uncertainty: Opinion and Subjective Probability in Science. Oxford University Press, USA, 1991

Dietrich F, List C. Probabilistic Opinion Pooling. 2014 http://personal.lse.ac.uk/list/PDF-files/OpinionPoolingReview.pdf

Garthwaite PH, Kadane JB, O'Hagan A. Statistical Methods for Eliciting Probability Distributions, Journal of the American Statistical Association, 100:470, 680-701, 2005

Kadane, J. and Wolfson, L. J., Experiences in elicitation. Journal of the Royal Statistical Society: Series D (The Statistician), 47: 3–19. 1998

Martin S, Rice N, Smith P. Does health care spending improve health outcomes? Journal of Health Economics; 27:826–42. 2008

O'Hagan A. Uncertain Judgements : Eliciting Experts' Probabilities. John Wiley & Sons. 2006.

Peterson CR and Miller A. Mode, median and mean as optimal strategies. Journal of Experimental Psychology, 68; 363-367 1964

Soares MO, Bojke L, Dumville J, Iglesias C, Callum N, Claxton K. Methods to elicit experts' beliefs over uncertain quantities: application to a cost effectiveness transition model of negative pressure wound therapy for severe pressure ulceration. Statistics in Medicine;30(19):2363-2380. 2011.

Speirs-Bridge A, Fidler F, McBride M, Flander L, Cumming G, Burgman M.Reducing overconfidence in the interval judgments of experts. Risk Anal;30(3):512-23. 2010

Teigen KH, Jorgensen M.When 90% confidence intervals are only 50% certain: On credibility of credible intervals. Applied Cognitive Psychology, 2005; 19:455–475.