Health Opportunity Costs and Expert Elicitation: A Comment on Soares et al.

Chris Sampson, PhD [corresponding author]

csampson@ohe.org

https://orcid.org/0000-0001-9470-2369

Office of Health Economics

Isobel Firth, MPH

ifirth@ohe.org

Office of Health Economics

Adrian Towse, PhD

atowse@ohe.org

Office of Health Economics

Soares et al.[1] have published a valuable study on the use of expert elicitation techniques to identify elusive quantities in public policy. Their contribution may prove to be an important step in the development of robust methods to support evidence-based decision-making.

The stated purpose of the structured elicitation is "to inform estimates of expected health opportunity costs in the UK NHS". The primary conclusion of the study is that the impact of expenditure in the NHS is likely to have been underestimated, such that the shadow price of a quality-adjusted life year (QALY) in the NHS is less than £12,936.

In this commentary, we dispute the authors' conclusion. We outline three reasons why the findings are unlikely to provide valid inputs to a revised estimate of the opportunity cost of health care expenditures.

There are no experts

An expert is someone with relevant experience that would enable them to provide realistic point estimates and confidence levels for values not available from other (empirical) sources. For example, expert elicitation exercises are used to identify clinical endpoints associated with health care interventions. A common exercise is to elicit expert opinions from clinicians, to identify a relevant measure for patient response to a therapy, based on prior observations from their experience of treating a disease .

The quantities of interest in Soares et al. related to associations between mortality, morbidity, and expenditures in health care at the system level. These quantities are not observable. This creates a difficulty for expert elicitation. It is not clear who, if anyone, could be considered an expert when judgements cannot be based on prior observation.

Participants were asked to consider parameters that related to an array of heterogeneous diseases and patient groups within a budgeting category. This introduces two important problems. First, nobody could be expected to be an expert in all of the diseases mentioned. Second, it seems unlikely that any 'expert' could come to a reasonable aggregation across diseases, based on the prevalence of each disease within the budgeting category.

The lack of expertise is illustrated in the data provided by the authors. A quarter of clinical experts explicitly stated that they were either not an expert or were only an expert in a single clinical field, and therefore lacking relevant knowledge.

The majority of policy experts were from 'governmental bodies'. These people are likely to have diverse professional experience but are unlikely to have clinical experience relevant to the questions asked. Rather, they are likely to be experts in the process of policy analysis and advocacy, which has little relevance to this exercise. These participants identified many difficulties in the elicitation process with some stating that they were not contributing their own opinions, deriving their views entirely from the clinical experts.

Clinical experts in one disease area are unlikely to have relevant knowledge at the level of aggregation required by this exercise. Given the complexity of the elicited quantities, particularly where they rely on the identification of causal relationships (e.g. between expenditure and mortality), it is difficult to see who could be an expert.

The elicited quantities are not meaningful

There are several frameworks for effective expert elicitation in health care research. Notably, the Sheffield Elicitation Framework (SHELF) is regularly used for this purpose.[2] It is not clear

whether a standardised protocol was adopted by Soares et al. Nevertheless, it is useful to consider SHELF criteria as indicators of good practice.

Quantities of interest elicited using the SHELF protocol satisfy three conditions:
1. The definition must be clear and unambiguous;
2. It should be such that the quantity of interest will have a unique value; and
3. It should be formulated to make the experts' judgements as simple as possible.

The elicitation described by Soares et al. does not appear to satisfy any of these conditions.

A key respect in which many of the quantities are neither clear nor unambiguous in their definition is in the comparison of heterogeneous groups of diseases. This was a concern raised by many participants. Experts were asked to consider the disease areas within each budgeting category where an increase in expenditure is more likely to fall. Whether or not any of the participants could reasonably interpret this point is unclear, and each may have interpreted it differently, and in a way that could introduce bias. One possibility is that the consideration of specific disease areas introduced an ambiguity effect, whereby respondents focused on favourable outcomes and ignored the diseases for which they lacked information. Qualitative responses show that some participants considered the effectiveness of specific interventions.

The researchers elicited quantities of interest that characterise the proportional relationship between two other quantities, as shown by the examples in Table 1: Examples of proportional comparisons required for three elicited quantities. While a given expenditure change might be expected to influence both quantities (X and Y), it is unlikely that the relative magnitude of effect for each is constant (i.e. that $X \propto Y$) in reality, either through time or across diseases. As such, it is difficult to conceive of unique quantities that could be identified by experts.

Quantity X	Elicited Quantity	Quantity Y
Duration of mortality effects		
Year 1 mortality rate	$\propto$	Year t mortality rate
Surrogacy		
Year t mortality burden	$\propto$	Year t QALY burden
Extrapolation		
Year 1 mortality burden in disease i	$\propto$	Year 1 mortality burden in disease j

Table 1: Examples of proportional comparisons required for three elicited quantities

The researchers characterised non-mortality consequences in a way that was not simple. That is, all consequences were related to mortality. The capacity of expenditure to improve quality of life – even in aggregate – may be independent of its capacity to extend life. Indeed, these outcomes may be substitutes. It is not clear how respondents would conceptualise these relationships.

The experts' judgements would have been simpler if based on absolute quantities. The authors justify the use of relative quantities on the grounds that they support conditional independence. However, it seems unlikely that conditional independence is a reasonable assumption, with expenditure in one disease area likely to have spillover effects into other areas and through time. Furthermore, the hypothetical change in expenditure was not

explicitly specified as being temporary. Individuals may have interpreted the change in expenditure as either temporary or permanent, and we would expect their responses to the questions to differ accordingly.

Our suggestion that the quantities of interest could not be meaningfully quantified is supported by the participants' comments. One respondent stated that they were "not sure what I have based my estimates on", while others explained the problems associated with comparing budgeting categories and disease areas. One respondent said that there was "too much to aggregate".

Participants were given very little information to support their judgements. Some of the quantities that were elicited required participants to speculate about quantities that could have been otherwise estimated, such as mortality rates for different diseases. Many of the clinical experts stated that the exercise was very difficult, while some suggested that they may not have understood the questions.

There is significant uncertainty in responses

As the authors note, there is a high level of uncertainty in the pooled values. By design, the exercise did not allow for trade-offs between mortality and morbidity or between current and future benefits; all values were bound by zero. For all quantities, credible intervals included values close to this lower bound, as well as very high positive values (which were unbounded). The values imply conflicting conclusions and are difficult to interpret.

Only 14% of clinical experts (and 32% of policy experts) were confident that their answers represented their views on mortality effects, surrogacy, and extrapolation. It is important to note that this question is not about respondents' confidence about their answers representing reality, but about their answers representing their own views. Thus, the majority of responses might be considered invalid.

Concluding remarks

Soares et al.[1] ought to be commended for their research study, not least because they chose to make a wealth of material available, which in turn has enabled us to understand the study in greater detail. We hope that our commentary can support further development of methods and practice in this area.

Taken together, the concerns that we have outlined make the authors' conclusion untenable. While being a valuable exercise in methodological development, the study cannot support any practical conclusions about the validity of the assumptions used in previous work or the optimality of any cost-effectiveness threshold used in policy.

Notably, the authors have not chosen to estimate (or even approximate) a revised central estimate for the marginal productivity of health care in the NHS. If such a task were to be undertaken, it would raise other challenges that we do not discuss here. Taken at face value, the mean pooled estimates seem to imply that NHS productivity may be several orders of magnitude greater than previously reported. Such an estimate would lack face validity.

Even if the elicitation exercise were valid, these findings would tell us little about the true opportunity cost of expenditure in the NHS. Rather, they reinforce the high level of uncertainty associated with earlier estimates.

References

1. Soares MO, Sculpher MJ, Claxton K. Health Opportunity Costs: Assessing the Implications of Uncertainty Using Elicitation Methods with Experts. *Med Decis Making* 2020; 40: 448–459.

2. Oakley JE, O'Hagan A. SHELF: The Sheffield Elicitation Framework (version 3.0), http://tonyohagan.co.uk/shelf (2016, accessed 18 July 2018).