International Journal of Technology Assessment in Health Care, 30:4 (2014), 461–468.

doi:10.1017/S0266462314000427

USE OF EXPERT KNOWLEDGE ELICITATION TO ESTIMATE PARAMETERS IN HEALTH ECONOMIC DECISION MODELS

David Hadorn, Giorgi Kvizhinadze, Lucie Collinson, Tony Blakely
Burden of Disease Epidemiology, Equity and Cost Effectiveness Programme, Department of Public Health, University of Otago, Wellington, PO Box 7343, Wellington, New Zealand

Objectives: The aim of this study was to determine the prevalence and methods of expert knowledge elicitation (EKE) for specifying input parameters in health economic decision models (HEDM).
Methods: We created two samples using the National Health System Economic Evaluations Database: (1) 100 randomly selected HEDM studies to determine prevalence of EKE and (2) sixty studies using a formal EKE process to determine methods used.
Results: Fifty-seven (57 percent) of the random sample included at least one EKE-derived parameter. Of these, six (10 percent) used a formal expert process. Thirty-four studies from our second sample of sixty studies (57 percent) described at least one aspect of the process (e.g., elicitation method) with reasonable clarity. In approximately two-thirds of studies the external experts estimated parameters de novo; the remainder confirmed or modified initial estimates provided by authors, or the method was unclear. The majority of elicitations obtained point estimates only, although a few studies asked experts to estimate ranges of parameter values.
Conclusions: The use of EKE for parameter estimation is common in HEDMs, although there is room for improvement in the methods used.

Keywords: Expert opinion, Health knowledge, Attitudes, Practice, Decision modeling, Assessment, Biomedical technology

Ever-increasing pressure on public funding for health services is likely to increase the importance of health economic decision modeling (HEDM) over the coming years. To be accepted as a valid tool for assessing the value of health services, HEDM techniques must be viewed as sufficiently credible, including application of "due diligence" in deriving estimates of key parameters (e.g., effect sizes, transition probabilities, resource use, and costs). Unfortunately, empirical evidence is often unavailable to provide sufficient guidance for estimating parameter values. In such cases, analysts often choose simply to "guesstimate" parameter values, relying on sensitivity analyses to manage the uncertainty around these values.

Alternatively, analysts may use formal elicitation techniques to obtain parameter estimates from outside experts, that is, individuals other than the study authors with credentials and/or significant experience in the relevant fields. Formal expert knowledge elicitation (EKE) has been used in a wide variety of healthcare settings, including deriving prior probabilities in Bayesian analyses of esophageal cancer (1), pulmonary embolism (2), severe head injuries (3), and hepatocellular carcinoma (4). Further examples of EKE are given in O'Hagan et al. (pp. 195–204) (5).

Taxonomy of Formal Methods

Two overlapping methodological dichotomies can be discerned within this body of work: (i) consensus versus mathematical approaches and (ii) deriving point estimates only versus deriving likely ranges or distributions of model parameters. Regarding the former dichotomy, consensus approaches entail groups of experts developing collective estimates, whereas mathematical approaches combine individual, independently derived estimates using quantitative measures, for example, the mean or median. Hybrid approaches use two or more rounds of group-level iteration and feedback (e.g., the Delphi method) followed by aggregation of the final individual estimates (6).

Regarding the second dichotomy, point estimates are the easiest and (therefore) most common output of EKE. Often this is sufficient information for modeling purposes, but more sophisticated outputs have been obtained by some researchers, including estimates of interquartile range and 95 percent uncertainty intervals (7;8). For example, Garthwaite and colleagues (8) asked experts to estimate point values and interquartile ranges for several parameters related to the management of patients with bowel cancer. Leal et al. (7) obtained estimates of proportions of patients falling within specified ranges of outcomes. A more sophisticated approach to eliciting probability

The Burden of Disease Epidemiology, Equity and Cost Effectiveness Programme (BODE^3) receives funding support from the Health Research Council of New Zealand (10/248). The authors have no conflicts of interest. We thank other BODE^3 team colleagues for comments and contributions to early versions of this work, in particular Nick Wilson and Rachel Foster-Russell.

distributions is to ask experts to place a set of twenty counters, each accounting for 5 percent probability weight, over a set of potential parameter values—the so called "bins and chips" approach (9). This is one of several approaches incorporated within the Sheffield Elicitation Framework (SHELF), a "package of documents, templates and software to carry out elicitations of probability distributions for uncertain quantities from a group of experts" (10). A similar approach was used by McKenna et al. (11) to estimate persistence of quality of life benefits in patients with angina pectoris. There are also online tools (e.g., http://statistics.open.ac.uk/Projects/Elicitation).

Little if any information exists concerning the extent to which such elicitation methods are used in contemporary HEDM studies (i.e., outside a methodological research context). We address this question in the current study. Nor is information available concerning the relative performance of different elicitation methods in terms of the accuracy of the estimates, or, for that matter, between EKE and simple author "guesstimates." This lack of information is due in large part to the absence of "gold standard" or "correct" values for EKE parameters. Indeed, if correct values were known, EKE would not be needed.

On the other hand, some efforts have been made to distinguish different levels of accuracy across experts by, for example, asking experts to estimate quantities whose values are independently known based on objective information (12). For example, Aspinall (13) described how he asked experts to estimate time to failure for thousands of small old earth dams in the United Kingdom. The estimates were differentially weighted based on how well the experts had performed in a preliminary series of so-called *seed questions*, to which the "correct" answer was known, for example, "how long did the Teton Dam in Idaho take to fail (which it did in 1976) after it started leaking?". The mean "performance-weighted" estimate of time to failure (70 days) was much higher than the estimate obtained from pooling individual estimates with equal weights (6 days). Unfortunately Aspinall (13) offered no information concerning which estimate was better. A similar weighting approach was used in a Canadian study in which EKE was used to estimate the risk of iatrogenic prion disease (e.g., Jacobs-Creutzfeld disease) (14). Again, however, no definitive conclusions were reached concerning the relative accuracy of performance-weighted versus unweighted estimates.

Expert Knowledge Elicitation Guidelines

Despite the potential desirability of undertaking EKE when study-based evidence is lacking, there is as yet no consensus concerning the "best" approach for doing so and indeed surprisingly little guidance available. For example, a draft report entitled "Model parameter estimation and uncertainty" by the International Society for Pharmacoeconomic and Outcomes Research – Society for Medical Decision Making's Joint Modeling Good Research Practices Task Force (15) stated:

> "Sometimes there is very little information on a parameter, either because there are very few studies informing the parameter estimation, or because there are no data at all and expert opinion must be relied upon. In these situations, it is imperative that the uncertainty related to such parameter estimates be fully explored. Analysts should adopt a conservative approach such that, in the absence of evidence on a given parameter, a very broad range of possible estimates is considered" (lines 233–236).

No further consideration was given in this document to processes of EKE, though the authors briefly noted that work has been done toward developing formal elicitation practices, citing O'Hagan et al. (5).

Based on their own experiences, Garthwaite et al. (8) suggested eight principles of good EKE: well-chosen experts; advance discussions with experts on the appropriate questions to ask; elicitation in such a way that requirements for statistical coherence were satisfied (but without the expert having to focus on these requirements); flexible elicitation methods to adapt to expert's preference; some elicitations validated, both by eliciting the opinions of more than one expert and comparing their answers and by comparing an expert's opinion with data; elicit intervals (in addition to point estimates) so as to quantify an expert's uncertainty; HEDM allowing appropriately for uncertainty in EKE parameters; and that the elicitation process and the resulting assessments are reported in detail.

Objectives

As noted above, we are not aware of any studies that have described current EKE practices in contemporary HEDM studies. Such information would be useful for assessing the discipline's current level of practice, and possibly for (further) developing guidelines on good EKE practice. Therefore, the first objective of our study was to estimate the prevalence of informal and formal EKE across a random sample of HEDM publications. The second more substantial objective was to describe the methods used in studies that used a formal EKE process. In this regard, we sought to determine, for each relevant study: (i) how many experts were used; (ii) qualifications of the external experts, for example, clinical specialists, generalists, nonclinical; (iii) what process was used, for example, Delphi, preparation of background paper or formal questionnaire, whether experts gave individual estimates, what types of parameters were estimated; (iv) whether experts estimated parameters de novo or modified authors' estimates; (v) how experts specified uncertainty around parameters; (vi) what form of sensitivity analysis was carried out by the authors on the EKE parameters.

METHODS

Part 1: To obtain an indication of the prevalence of formal EKE in contemporary HEDM studies, we randomly sampled studies listed in the UK National Health System

https://doi.org/10.1017/S0266462314000427 Published online by Cambridge University Press

Economic Evaluations Database (NHS-EED) between 2008 and 2010. NHS-EED is compiled using a protocol designed to detect and incorporate essentially all HEDMs published in peer-reviewed journals (http://www.crd.york.ac.uk/CRDWeb/AboutNHSEED.asp), but also includes studies that are not "full" HEDMs (e.g., costing studies). We had to randomly sampled 216 studies from the NHS-EED to obtain our target of 100 HEDM studies; we excluded the other 116 papers as they did not meet the "HEDM inclusion criteria" of explicit decision modeling and quantitative estimates of costs, clinical or preventive care processes, preferences, or health outcomes. (See Supplementary Table 1, which can be viewed online at http://dx.doi.org/10.1017/S0266462314000427, for list of the 100 studies.) We then ascertained whether EKE was used to estimate any of these parameters and if so whether a formal process was used.

We classified the topic areas of these studies as follows: (i) Therapeutic intervention – pharmaceutical; (ii) Therapeutic intervention – nonpharmaceutical; (iii) Diagnostic intervention; (iv) Screening or monitoring; (v) Prevention program; (vi) Vaccination.

For describing the types of parameters used in the various models, we developed the following taxonomy: (i) *Clinical practice*, including how tests, treatments, hospitalization, and other interventions are typically used in the relevant practice environment (e.g., appropriate use and doses of pharmaceuticals); also includes how patients "flow" from one health state or intervention to another and the likely duration of states or treatments (e.g., length of time spent in hospital or duration of impairment). (ii) *Unit price* or *cost* of interventions. (iii) *Epidemiology*, including incidence, prevalence, event rates, transition probabilities (e.g., alive-dead), change in quality of life, and effect size (e.g., absolute or relative risk reduction). (iv) *Utilities*, that is, health state valuations.

For describing the highest level of EK used for each of the above four types of parameters: *Nil*. No expert knowledge used; all input parameters were linked to cited empirical studies. *Informal*. Assumed or estimated by the authors without reference to an empirical study or data analysis, and no formal external expert process used (as defined below). *Formal*. At least one parameter was estimated by two or more external experts (i.e., not co-authors), having at least one of the following characteristics: (i) qualifications of two or more experts given, (ii) a formal process was described (e.g., "Delphi," how experts selected, whether experts approached separate or as a group); (iii) multiple rounds of rating were used; (iv) a background paper was prepared to assist experts; (v) a formal questionnaire was administered to experts.

For describing the highest level of stated uncertainty around EK-derived parameters: (i) *No specific quantitative estimates*, qualitative model parameter only (e.g., used to specify model structure); no uncertainty estimates; (ii) *Point estimates only*; no uncertainty estimates; (iii) *Deterministic* – uncertainty specified by deterministic values, for example, best, worst, high, low, or range; (iv) *Probabilistic* distributions provided (including simple uniform, triangular and trapezoidal, gamma, normal, etc.).

For describing the highest level of sensitivity analysis used for EK-derived parameter: (i) *None or unidentifiable*; (ii) *One-way deterministic* – only one parameter varied at a time, no probability distribution used (e.g., defined scenarios, highest/lowest plausible values, deterministic range); (iii) *One-way probabilistic* – only one parameter varied at a time, probability distribution stated; (iv) *Probabilistic sensitivity analyses* (PSA)/N-Way – formal probability distribution used to determine impact of changing values of EK parameter on model output and/or two or more parameters jointly varied.

Both forms of probabilistic analysis ([iii] and [iv]) require repeated sampling from the specified distribution (e.g., Monte Carlo or bootstrapping).

Part 2: Our target papers for inclusion were those HEDM studies in the NHS-EED between 2006 and 2011 (inclusive) that would have been classified as "formal" EKE in Part 1 (see criteria above). It was not efficient to just randomly sample studies from the NHS-EED until we found 60 such studies (our target sample size given time and resource constraints). Thus, we developed a keyword search strategy to apply to the NHS-EED as follows. We trialed six search strategies including and excluding different search terms such as "expert", "panel", "opinion", "elicit", "formal", and "Delphi", using different combinations of the Boolean operators "AND" and "OR" to vary the sensitivity and specificity of the search. The selected terms were searched in all fields (i.e., title, abstract, free text) for the years 2006–2011. We reviewed the first twenty-five papers from each test search to determine the proportion that were both: HEDM studies AND meeting our criteria for using a formal EKE process (see criteria above under Part 1). The search strategy ("expert" AND "panel" [not necessarily together] OR "Delphi") had the highest "hit rate," with 79 percent of the first 25 papers being HEDM studies using a formal EKE process. (By contrast, for example, only 28 percent of the first 25 papers from the search strategy "expert AND opinion" OR "expert AND panel" OR "delphi" were HEDM studies using a formal EKE process, due to "false positive" studies resulting from use of the term "opinion". Similarly, we found that use of "expert" without "panel" resulted in too many false-positive hits.) Our preferred ("expert" AND "panel" [not necessarily together] OR "Delphi") search yielded a total of ninety-four HEDM studies that actually used a formal EKE process. For these ninety-four studies, we randomly selected sixty to review (listed in Supplementary Table 2, which can be viewed online at http://dx.doi.org/10.1017/S0266462314000427). Full copies of each article were reviewed, including any appendices or supplementary material on EKE methods.

A draft pro forma for scoring the studies, based on the items listed in the bullet points at end of the Introduction, was independently trialed on ten studies by G.K. and L.C.,

resulting in a 70 percent agreement rate. Scoring criteria were then modified to facilitate inter-rater agreement and a further twenty studies scored independently by G.K. and L.C. Agreement rate on these studies rose to 96 percent. The first ten studies were rescored based on these modifications and GK and LC then reviewed another thirty papers (fifteen each), discussing any scoring queries along the way. The final pro forma is shown in Supplementary Table 3, which can be viewed online at http://dx.doi.org/10.1017/S0266462314000427. Data from the sixty studies were analyzed using Microsoft Excel.

Many of the studies in our samples failed to provide clear information on the process used for EKE of model parameters. In many cases we were unable to distinguish between a "zero" value for a parameter (e.g., number of face-to-face meetings) versus "not stated" (NS). We coded NS unless we could be sure that the true parameter value was zero.

We distinguished between uncertainty and sensitivity analyses, restricting the former term in this study to apply to estimates around *parameters* (e.g., inter-quartile range, 95 percent confidence interval). This uncertainty can be either *within-expert* (with distributions showing personal subjective probability density functions over possible parameter values, e.g., bins and chips), or *between-experts* with distributions showing and summarizing the different values obtained from individuals. Sensitivity analysis, on the other hand, refers to the process of assessing the effects of different possible parameter values on outputs of the *models*, especially incremental cost-effectiveness. We divided these sensitivity analyses into two types: (i) deterministic (or scenario) analysis, in which different *fixed* values are assigned to the parameters, and (ii) probabilistic uncertainty analysis, in which repeated random sampling is conducted from *probabilistic distributions* of possible values.

RESULTS

Part 1: Fifty-seven of our initial sample of 100 randomly selected HEDM studies (57 percent) used EKE to estimate at least one model parameter. Of these fifty-seven studies, only six (10 percent) used a formal EKE process for at least one EKE-derived parameter, although this process was usually poorly described. Put another way, most instances of EKE were actually "just" author estimations (although we think it fair to credit authors with being experts in their own right – so long as they have thoroughly reviewed existing knowledge). Thirty-nine studies (69 percent) reported information on uncertainty around the EKE parameters and thirty-six studies (64 percent) undertook some form of sensitivity analysis.

Part 2: The sixty randomly selected formal EKE process HEDM studies from our more focused search were co-authored by 306 individuals from twenty-five countries. The majority of authors were from the United States ($n = 79$; 26 percent), United Kingdom ($n = 65$; 21 percent), and Germany and Spain (each $n = 24$; 8 percent). The most common study type in our sample was cost-effectiveness analysis (CEA) of pharmaceutical interventions ($n = 37$; 62 percent), followed by CEAs of vaccines ($n = 6$; 10 percent) screening interventions ($n = 4$; 7 percent) and diagnostic interventions ($n = 4$; 7 percent). Nine interventions (15 percent) did not fall into any of these categories, such as surveillance strategies for cancer (16).

The number of external experts used for EKE was provided in thirty-seven studies (62 percent), with a mean of 8.4 experts per study (median 5, with an interquartile range of 5–9). The composition (occupation or qualifications) of the external experts was stated in forty studies (67 percent) with twenty-two (55 percent) using clinical specialists only, three (8 percent) using specialists and general practitioners (GPs), two (5 percent) using GPs only, one (3 percent) using specialists plus nonclinical practitioners, one using a combination of these three types of expert, and one using lay experts only (parents of children with a particular condition).

Thirty-four studies (57 percent) described at least one aspect of the EKE process. Twenty-four studies (40 percent) described the elicitation process as Delphi. The authors of 18 percent of studies stated they provided written information to experts before elicitation, 30 percent of studies stated the use of some form of structured survey or questionnaire, and 23 percent of studies stated they contacted experts at least twice. Many studies used a form of group consensus (more so when the process was described as Delphi), but 17 percent of studies stated they elicited parameters separately and independently from experts. Supplementary Figure 1, which can be viewed online at http://dx.doi.org/10.1017/S0266462314000427, shows selected process characteristics *when stated* for the sixty studies.

Table 1 shows the parameter classes that external expert's estimated. Assuming that when we failed to find evidence of EKE for a given parameter class that it truly was absent in the study, then forty-six (77 percent), twenty-seven (45 percent), fifteen (25 percent), and twelve (20 percent) of the sixty studies used EKE for at least one parameter within the clinical, epidemiology, cost and utility parameter classes, respectively (Table 1). (Fourteen studies also used expert experts to elicit some information about model structure.) For many of these studies, we could not determine the actual number of EKE parameters by class, restricting us to forty, twenty-five, thirteen, and eleven studies, respectively. Within these studies, we could separately identify a total of 638 formal EKE process parameters. For the 329 clinical practice parameters across 40 quantifiable studies, this was an average of 8.2 clinical parameters per study (Table 1). External experts were most commonly used for eliciting clinical parameters, followed by epidemiological and cost parameters, and least commonly for utility parameters.

Table 2 shows the method used to derive parameter estimates from the expert experts. For example, of the twenty-seven studies that used a formal EKE process for epidemiological parameters, two confirmed parameter estimates provided by the authors, one study modified the authors' estimates,

Table 1. Number of Studies Using a Formal EKE Process by Parameter Class, and the Total Number, Mean, Median, and Range of Parameters among Those Studies

	Number of studies (% of 60)						
		Where we could detect ≥ 1 parameter of this class elicited:		Summary statistics on parameters, for studies we could determine the actual number and it was ≥ 1 (i.e. 'C' studies)			
Parameter class	A. Where we could detect no parameters of this class elicited (assumed zero)	B. But the actual number not specified (NS)	C. And the actual number could be determined	Total	Mean	Median	Range
Clinical	14 (23%)	6 (10%)	40 (67%)	329	8.2	2	1–56
Epidemiology	33 (55%)	2 (3%)	25 (42%)	155	6.2	2	1–39
Cost	45 (75%)	2 (3%)	13 (22%)	117	9.0	1	1–47
Utility	47 (78%)	2 (3%)	11 (19%)	37	3.4	1	1–17

Note. The model structure is qualitative and therefore the mean, median, and range are not applicable.

Table 2. Method of Parameter or Model Structure Derivation, by Parameter Class

		Method of parameter(or model structure(MS) derivation by experts (% of studies using each method for given parameter class)			
Parameter class	Number of studies	Confirmed parameter/MS estimate(s) provided by authors[a]	Modified the parameter/MS estimate(s) provided by authors[a]	De novo[b]	Unclear
Model Structure	14	7 (50%)	1 (7%)	1 (7%)	5 (36%)
Clinical practice	46	2 (4%)	1 (2%)	33 (72%)	10 (22%)
Epidemiology	27	2 (7%)	1 (4%)	18 (67%)	6 (22%)
Cost	15	2 (13%)	0 (0%)	10 (67%)	3 (20%)
Utilities	13	4 (31%)	1(8%)	7 (54%)	1 (8%)

[a]Where the authors provided their own best estimate based on a literature search or other means.
[b]Where authors did not provide initial estimates.

and 18 (67 percent of the twenty-seven studies) provided de novo estimates. De novo estimation was also the most common method across the three other nonmodel structure parameter classes (Table 2). The method of elicitation for model structure was more likely to be confirmatory (50 percent) or unclear (36 percent).

Uncertainty and Sensitivity Analysis

The type of uncertainty provided by the experts around each parameter was reported for ninety-nine of the 640 (15 percent) separate parameters we could identify (Table 3; as this table quantifies separate parameters (as opposed to studies with at least one parameter example in all other Tables), totals are also provided). Of these, eighty-one parameter estimates (82 percent) were elicited in the form of point estimates only from the experts (i.e., no uncertainty elicited from the experts themselves, although the authors may have subsequently determined uncertainty distributions based on the multiple values provided by the external experts). Likely ranges were obtained for fifteen parameter estimates (15 percent) and distributions were elicited directly from the experts for three parameter estimates (3 percent).

Table 4 shows the types of deterministic sensitivity analyses classified by study—not parameter. For example, of the forty-six studies that used EKE for clinical parameters, fourteen (30 percent) reported a deterministic or scenario sensitivity analysis that involved at least one of the EKE clinical parameters. A further twenty-six of these studies (57 percent) conducted deterministic sensitivity analyses, but we were unable to reliably

https://doi.org/10.1017/S0266462314000427 Published online by Cambridge University Press

Table 3. Type of Uncertainty Specified for 99 Separate Parameters with Sufficient Information for Characterization

	Level of uncertainty obtained from experts (% of studies using a formal EKE process for a given parameter class)		
Parameter class	Point estimate only[a]	Range (e.g. minimum to maximum)	Distribution[B] (e.g. histogram of shape)
Clinical practice	39 (91%)	4 (9%)	0
Epidemiology	20 (77%)	6 (23%)	0
Cost	12 (80%)	3 (20%)	0
Utilities	10 (67%)	2 (13%)	3 (20%)
Total	81 (82%)	15 (15%)	3 (3%)

[a]If not stated, point estimate was assumed.
[b]Expert's distribution (histogram) to indicate within-expert probability density function.

Table 4. Studies Classified by Parameter Class by Reported Deterministic Sensitivity Analysis and Probabilistic Sensitivity Analysis Involving EK-Derive Variables

	Deterministic sensitivity analysis			Probabilistic sensitivity analysis[a]			
Parameter class	Reported for EKE parameter	Used, but unclear if included EKE parameters	Not used	Univariate PSA reported for EKE parameters	Multivariate PSA reported for EKE parameters	Used, but unclear if EKE parameters included	Not used
Clinical practice ($n = 46$ studies)	14 (30%)	26 (57%)	6 (13%)	1 (2%)	8 (17%)	15 (33%)	22 (48%)
Epidemiology ($n = 27$ studies)	11 (41%)	11 (41%)	5 (18%)	0	8 (30%)	10 (37%)	9 (33%)
Cost ($n = 15$ studies)	4 (27%)	7 (47%)	4 (27%)	1 (7%)	5 (33%)	2 (13%)	7 (47%)
Utilities ($n = 13$ studies)	5 (38%)	3 (23%)	5 (38%)	1 (8%)	6 (46%)	4 (31%)	2 (15%)

Note. The classification of studies by deterministic and probabilistic sensitivity analysis was separate. For example, a study using deterministic methods might also use probabilistic methods. For a study with two or more EK parameters, we only required evidence of sensitivity analysis for one parameter for coding.
[a]If both univariate and multivariate reported, scored only as univariate

determine if EKE parameters were included. These percentage distributions were reasonably similar across the four parameter classes.

The use of probabilistic sensitivity analysis is also shown in Table 4, and was classified separately from deterministic sensitivity analyses, meaning the same study could be classified as both deterministic and probabilistic sensitivity analysis. Very few studies undertook univariate probabilistic sensitivity analyses specifically on an EKE parameter. Multivariate PSA involving an EKE parameter was more common, ranging from 17 percent to 46 percent across the parameter classes; however, this approach does not necessarily identify the contribution of uncertainty in EKE parameter(s) to overall model output uncertainty, as all variables with probabilistic uncertainty (both EKE and empiric) are simultaneously sampled. Approximately a third of studies including at least one EKE parameter used PSA, but it was unclear whether uncertainty in the EKE parameter(s) was part of the PSA (although one would assume it was).

DISCUSSION

We found that EKE was used for at least one parameter in over half ($n = 57$) of 100 randomly selected HEDM studies, but only six (10 percent) of these studies use a formal EKE process. Uncertainty analysis around EKE parameters was conducted in approximately two-thirds of the studies using either formal or informal EK. Our more in-depth assessment of 60 HEDM studies using a formal EKE process found that the most commonly elicited parameters concerned clinical practice or epidemiology, and that EK estimates were most often derived "de novo" as opposed to confirming authors' suggested values. Most EKE was of point estimates only; approximately a quarter of the sixty studies elicited from experts some range or distribution about a central estimate. We could identify that approximately a third of studies using a formal EKE process subjected their EKE parameters to each of deterministic and probabilistic sensitivity analyses. It was difficult to identify the impact of uncertainty in EKE parameters on final model outputs such as the ICER.

https://doi.org/10.1017/S0266462314000427 Published online by Cambridge University Press

We found only one study that used a more sophisticated type of EKE as discussed in the Introduction; this involved a variant of the "bins and chips" method (11). We found no studies that differentially weighted experts' estimates based on responses to "seed questions" whose correct answers are known. And only one parameter in one study had its uncertainty distribution directly elicited from the experts (Table 3). Thus, there is to some extent a "disconnect" between mainstream HEDM and methodological EKE research.

A potential limitation of Part 2 of our study was our filtering search strategy to identify formal EKE process studies. As noted in the Methods section, this was necessary for efficiency reasons to reduce the number of false-positive returns. However, in so doing we may have skewed our sample somewhat (e.g., toward those using a Delphi process given it was included in the search strategy) compared with all formal EKE process studies. Nevertheless, we believe the selected studies and findings meet the objectives of this study.

Our findings, coupled with the lack of attention to best elicitation practice (see Introduction), could be taken to suggest that formal EKE is an optional component of HEDM analyses, even when no empirical evidence is available concerning the values of important model parameters. We would, however, be hesitant to draw this conclusion in view of the increasing social significance of HEDM in prioritizing healthcare services. Also, progress is being made in making elicitation techniques more sophisticated (10).

Although we cannot draw definitive conclusions from our findings as to what constitutes best EKE practice, the following recommendations seem congruent with contemporary practice and with the range of EKE methods described in the Introduction. First, and most importantly, whatever process is used should be described thoroughly (using Web-only appendices or supplemental material if necessary). As noted above, we found it difficult to determine what was actually done in the majority of studies.

Second, and consistent with the guidance from International Society for Pharmacoeconomic and Outcomes Research – Society for Medical Decision Making's Joint Modeling Good Research Practices Task Force cited in the Introduction, we recommend that analysts estimate the likely range of values of EKE parameters with a "generous" level of uncertainty specified. If any parameters, when varied through their likely ranges, result in substantial differences in key outputs of the economic decision model (e.g., >10–20 percent difference in net costs, change in health outcomes, or incremental cost-effectiveness ratios), a formal EKE process would seem justified. For parameters found to have less substantial impacts on outputs, formal EKE is likely not needed.

Third, if EKE from external experts is warranted, one approach that may be cost-effective (and which is compatible with contemporary practice) would be to (i) identify five to nine people with the required expertise (usually including one or more specialists), (ii) obtain estimates from each expert individually (by means of e-mail in most cases), and (iii) integrate these estimates mathematically, reporting median estimates and, where feasible, probability distributions for subsequent analyses, for example, interquartile or 95 percent ranges. These latter descriptive statistics will provide an indication of between-expert uncertainty. Where feasible, within-expert uncertainty could also be derived by asking experts to estimate their own personal subjective likely ranges and distributions (e.g., using a bin and chips approach). A significant degree of statistical sophistication will be required for this step, although software such as SHELF may be of assistance here (10). That all said, we know of no research on the marginal added value (in terms of accuracy or validity) or cost effectiveness of such processes; this is a research question in its own right.

Based on current observed practice and the lack of evidence of differential validity of varying approaches, the following process features should be considered optional and of uncertain value: (i) conducting face-to-face meetings (which could in any case be counterproductive due to personality factors and small group dynamics), (ii) conducting multiple rounds of ratings, (iii) providing initial estimates of parameter values to the experts, and (iv) obtaining estimates of likely range or distribution of parameters directly from the experts (i.e., rather use the variation across experts).

From a research perspective, we suggest that whenever feasible HEDM studies planning to use EKE incorporate comparisons of estimates obtained with and without one or more of these latter features. The validity of such EKE processes would ideally be determined by comparing EKE estimates with true values where known (e.g., in the "seed questions" described above) or where studies are underway that may provide an evidence-based answer to the questions at some future point. Use of such "gold standard" parameters would generate information concerning the most difficult and important aspect of EKE: knowing whether the accuracy of parameter estimates varies according to the EKE process used.

SUPPLEMENTARY MATERIAL

Supplementary Table 1:
http://dx.doi.org/10.1017/S0266462314000427
Supplementary Table 2:
http://dx.doi.org/10.1017/S0266462314000427
Supplementary Table 3:
http://dx.doi.org/10.1017/S0266462314000427
Supplementary Figure 1:
http://dx.doi.org/10.1017/S0266462314000427

CONTACT INFORMATION

David Hadorn, MD, PhD
Giorgi Kvizhinadze, PhD
Lucie Collinson, MBChB MPH

Tony Blakely, PhD (tony.blakely@otago.ac.nz),University of Otago, Wellington, 23a Mein Street, Newtown, Wellington 6242, New Zealand

CONFLICTS OF INTEREST

The authors have no conflicts of interest.

REFERENCES[1]

1. van der Gaag LC, Renooij S, Witteman CL, Aleman BM, Taal BG. Probabilities for a probabilistic network: A case study in oesophageal cancer. *Artif Intell Med*. 2002;25:123-48.
2. Christiansen F, Nilsson T, Mare K, Carlsson A. Adding a visual linear scale probability to the PIOPED probability of pulmonary embolism. *Acta Radiol*. 1997;38:458-63.
3. Harmanec D, Leong TY, Sundaresh S, et al. Decision analytic approach to severe head injury management. *Proc AMIA Symp*. 1999:271-275.
4. Tan SB, Chung YF, Tai BC, Cheung YB, Machin D. Elicitation of prior distributions for a phase III randomized controlled trial of adjuvant therapy with surgery for hepatocellular carcinoma. *Control Clin Trials*. 2003;24:110-121.
5. O'Hagan A, Buck CE, Daneshkhah A, et al. *Uncertain judgements: Eliciting experts' probabilities*. Chichester, England: John Wiley & Sons Ltd; 2006.
6. Clemen RT, Winkler RL. Combining probability distributions from experts in risk analysis. *Risk Anal*. 1999:19:187-203.
7. Leal J, Wordsworth S, Legood R, Blair E. Eliciting expert opinion for economic models: An applied example. *Value Health*. 2007;10:195-203.
8. Garthwaite PH, Chilcott JB, Jenkinson DJ, Tappenden P. Use of expert knowledge in evaluating costs and benefits of alternative service provisions: A case study. *Int J Technol Assess Health Care*. 2008;24:350-357.
9. Johnson SR, Tomlinson GA, Hawker GA, et al. A valid and reliable belief elicitation method for Bayesian priors. *J Clin Epidemiol*. 2010;63:370-383.
10. O'Hagan A. *SHELF: The Sheffield Elicitation Framework version 2.0*. Sheffield, UK: University of Sheffield; 2010.
11. McKenna C, McDaid C, Suekarran S, et al. Enhanced external counterpulsation for the treatment of stable angina and heart failure: A systematic review and economic analysis. *Health Technol Assess*. 2009;13:iii-iv, ix-xi, 1-90.
12. Cooke RM. *Experts in uncertainty: Opinion and subjective probability in science*. Oxford: Oxford University Press; 1991.
13. Aspinall W. A route to more tractable expert advice. *Nature*. 2010;463:294-295.
14. Tyshenko MG, Darshan S. *Summary report of the expert elicitation workshop results for iatrogenic prion disease risks in Canada*. Expert Elicitation Workshop Summary Report. Ottawa, Canada: University of Ottawa; 2009.
15. Briggs A, Fenwick E, Karnon J, et al. *DRAFT model parameter estimation and uncertainty: Report of the ISPOR-SMDM Modeling Good Research Practices Task Force - 6 Model Parameter Estimation and Uncertainty*. Glasgow, Scotland, UK: Health Economics & Health Technology Assessment, Institute of Health & Wellbeing, University of Glasgow; 2010.
16. Andersson KL, Salomon JA, Goldie SJ, Chung RT. Cost effectiveness of alternative surveillance strategies for hepatocellular carcinoma in patients with cirrhosis. *Clin Gastroenterol Hepatol*. 2008;6:1418-1424.

[1] *Note.* The references listed here correspond to the citations in the main body of the report, and concern EKE background and methods. See Supplementary Tables 1 and 2 for lists of the actual HEDM studies selected and examined in Parts 1 and 2 of this study, respectively.