International Journal of Technology Assessment in Health Care

cambridge.org/thc

Method

Cite this article: Jankovic D, Payne K, Kanaan M, Bojke L (2022). Eliciting uncertainty for complex parameters in model-based economic evaluations: quantifying a temporal change in the treatment effect. *International Journal of Technology Assessment in Health Care* **38**, e21, 1–9. https://doi.org/10.1017/S0266462322000022

Received: 30 July 2021
Revised: 22 November 2021
Accepted: 9 January 2022

Key words:
Health outcomes research; Modeling; Decision uncertainty; Fall prevention

Author for correspondence:
Dina Jankovic,
E-mail: dina.jankovic@york.ac.uk

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

CAMBRIDGE UNIVERSITY PRESS

Eliciting uncertainty for complex parameters in model-based economic evaluations: quantifying a temporal change in the treatment effect

Dina Jankovic[1], Katherine Payne[2], Mona Kanaan[3] and Laura Bojke[1]

[1]Centre for Health Economics, University of York, York YO10 5DD, UK; [2]Manchester Centre for Health Economics, Division of Population Health, Health Services Research & Primary Care, The University of Manchester, Manchester, UK and [3]Department of Health Sciences, University of York, York, UK

Abstract

Background. In model-based economic evaluations, the effectiveness parameter is often informed by studies with a limited duration of follow-up, requiring extrapolation of the treatment effect over a longer time horizon. Extrapolation from short-term data alone may not adequately capture uncertainty in that extrapolation. This study aimed to use structured expert elicitation to quantify uncertainty associated with extrapolation of the treatment effect observed in a clinical trial.

Methods. A structured expert elicitation exercise was conducted for an applied study of a podiatry intervention designed to reduce the rate of falls and fractures in the elderly. A bespoke web application was used to elicit experts' beliefs about two outcomes (rate of falls and odds of fracture) as probability distributions (priors), for two treatment options (intervention and treatment as usual) at multiple time points. These priors were used to derive the temporal change in the treatment effect of the intervention, to extrapolate outcomes observed in a trial. The results were compared with extrapolation without experts' priors.

Results. The study recruited thirty-eight experts (geriatricians, general practitioners, physiotherapists, nurses, and academics) from England and Wales. The majority of experts (32/38) believed that the treatment effect would depreciate over time and expressed greater uncertainty than that extrapolated from a trial-based outcome alone. The between-expert variation in predicted outcomes was relatively small.

Conclusions. This study suggests that uncertainty in extrapolation can be informed using structured expert elicitation methods. Using structured elicitation to attach values to complex parameters requires key assumptions and simplifications to be considered.

Introduction

Decision-analytic model-based economic evaluation (MBEE) is intended to generate relevant costs and treatment effects for competing alternatives, over an appropriate time horizon (1–3). The characterization of uncertainty (parameter, structural, methodological, and decision) is a fundamental component in the design and conduct of MBEE (4). Methods for incorporating uncertainty into decision models have received substantial attention over the past two decades (5;6). Quantifying what this uncertainty looks like is more problematic with continued debate about appropriate methods and what to do in the absence of available evidence.

Temporal uncertainty arises when the time horizon for the specified MBEE exceeds the observed end point for available evidence. This type of uncertainty poses specific challenges (7). The finite, and often limited, duration of follow-up in randomized controlled trials (RCTs), used to populate the effectiveness parameter for treatments in MBEE, means it is necessary to extrapolate and predict the lifetime impact on outcomes (1). Ideally, external data from an appropriately designed observational study should be used. Such data are rarely available. Instead, outcomes can be extrapolated directly from the trial or instead, through scenarios representing assumptions or statistical models of the relative effect of treatment after the evidence end point (7;8). Neither of these approaches appropriately quantify uncertainty around the generated predictions for use in MBEE. Extrapolation ignores uncertainty in the predictive accuracy of the statistical model fitted to short-term data. Scenario analysis underestimates uncertainty if none of the scenarios accurately represent the expected trajectory of the treatment effect (8). One option is to delay approval decisions until temporal uncertainty has been resolved when longer-term data are made available. However, delays have consequences for patients who will not realize the benefit of the treatment or incur costs while waiting for research to report (7).

In the absence of empirical evidence, an alternative approach is to use structured expert elicitation (SEE) of experts' beliefs to characterize uncertainty in the progression of the

treatment effect to make decisions in the absence of required data or while waiting for data collection to be completed (7;8). Experts' beliefs can be captured as probability distributions (experts' priors) to represent their uncertainty in the expected value of a parameter. There is a growing interest in using expert elicitation to characterize uncertainty in MBEE, but there are only a few applied examples that attempt to elicit temporal uncertainty (9–13). The published studies used a range of methods to extrapolate different types of parameters, predominantly when "short-term" outcomes are available to experts. This study elicited experts' beliefs to quantify temporal uncertainty regarding the relative effect of a multifaceted podiatry intervention designed to reduce the rate of falls and fractures in the elderly, compared with treatment as usual (TAU). It is the first such study to our knowledge that elicited the relative effect (rate and risk ratios) before the short-term effect had been reported.

Methods

This SEE was designed in accordance with a published reference protocol (14) and reported in line with published recommendations (15).

Intervention

A multifaceted podiatry intervention involving education, exercise, foot orthoses, and footwear was designed to reduce the rate of falls and fractures in the elderly, with an associated impact on costs and health-related quality of life (HRQoL). The clinical and cost-effectiveness of the podiatry intervention in the first year of treatment has been evaluated in a published trial (16;17). In the trial, the podiatry intervention led to a modest reduction in the rate of falls [rate ratio = .88; 95 percent confidence interval (CI), .73–1.05] and an increase in the risk of fractures after a fall (risk ratio = 1.32; 95 percent CI, .65–2.76), although neither effect was statistically significant. These trial results were not available when designing the SEE exercise reported here.

Decision Problem

The podiatry intervention was designed for indefinite use, with a potentially long-term effect on mortality through a reduced risk of falls. The appropriate time horizon for this decision problem and associated analysis was, therefore, a life time (1). The decision problem for SEE was to generate estimates of the lifetime impact of the podiatry intervention compared with TAU, in the absence of empirical evidence. Crucially, this SEE assumed that the long-term effect of the podiatry intervention was correlated with the observed effect in the trial.

Elicitation Protocol

A pre-specified protocol to direct the elicitation exercise was produced (see Supplementary File 1) and piloted (see Supplementary File 2). The goal was to develop a SEE protocol that could be completed by an individual working in isolation within 1 h. Protocol development was guided by two physiotherapists with expertise in fall prevention.

Defining the Relevant Experts

Experts were defined as clinicians and/or researchers who: had applied knowledge of foot and ankle physiology; understood the risk factors for falling; understood the role of fall prevention interventions; and had direct experience of delivering behavioral interventions to patients. These identified experts included: clinicians representing a range of settings (fall prevention, fall treatment, and regular contact with at-risk population); different professions (geriatricians, nurses, physiotherapists, and researchers); and different levels of patient contact and professional experience (details in Supplementary File 1).

Identifying the Relevant Elicitation Parameters

The target parameter to be quantified in this SEE was defined as the relative change in the treatment effect for the podiatry intervention relative to TAU, over a lifetime horizon. A change in the treatment effect is not directly observable in RCTs or clinical practice (hence an unobservable parameter), and so it is difficult to estimate accurately (18). The parameter of interest was, in keeping with published recommendations (14), broken down into quantities that can be observed and measured: the defined outcomes for patients that do (do not) receive the podiatry intervention.

Two outcomes were elicited to capture the treatment effect: (i) rate of falls and (ii) the risk of having a fracture after a fall (risk of fractures thereafter). Elicited outcomes with (and without) the podiatry intervention were used to derive two treatment effects: (i) the rate ratio of falls and (ii) the relative risk of fractures. The temporal change in the treatment effect was then derived from treatment effects at two time points: (i) the treatment effect at 1 year after starting treatment (comparable to that observed in the RCT) and (ii) the treatment effect at a second (specified) time point.

Eliciting the Relative Change in the Treatment Effect

Six steps (see Figure 1) were followed guided by the SEE exercise protocol (Supplementary File 1) to quantify the relative change in the treatment effect.

Step 1: Eliciting 1-Year Outcomes for TAU (Baseline)

The rate of falls was derived from the frequency distribution of falls in a population by weighting each possible number of falls by its probability. It was assumed that each individual patient can experience more than one fall in a given time period, potentially exceeding ten falls (19;20) per year. Eliciting the probability of more than ten falls would be time-intensive. To reduce the burden on experts, possible outcomes were grouped, confirmed to be reasonable by a physiotherapist specializing in fall prevention, into three categories capturing the conditional probability of falling: at least once ($P(x > 0)$, where x is the number of falls); more than five times ($P(x > 5|x > 0)$); more than ten times ($P(x > 10|x > 5)$. Conditional probabilities were elicited to prevent statistical incoherence (e.g., $P\ (x > 5) > P\ (x > 0)$), and assumed to be independent. Eliciting correlation between the conditional probabilities was deemed to be prohibitively cognitively burdensome, requiring in-depth training of experts in the concept of correlation (21).

Experts' priors were elicited for each of the three categories of outcome as relative frequencies to derive the probability distribution of the rate of falls (details provided in Supplementary File 1).

The risk of fracture after a fall was elicited as odds to be consistent with how these values are reported in the literature, and to allow comparison of experts' priors when assessing uncertainty around different types of quantities.

https://doi.org/10.1017/S0266462322000022 Published online by Cambridge University Press

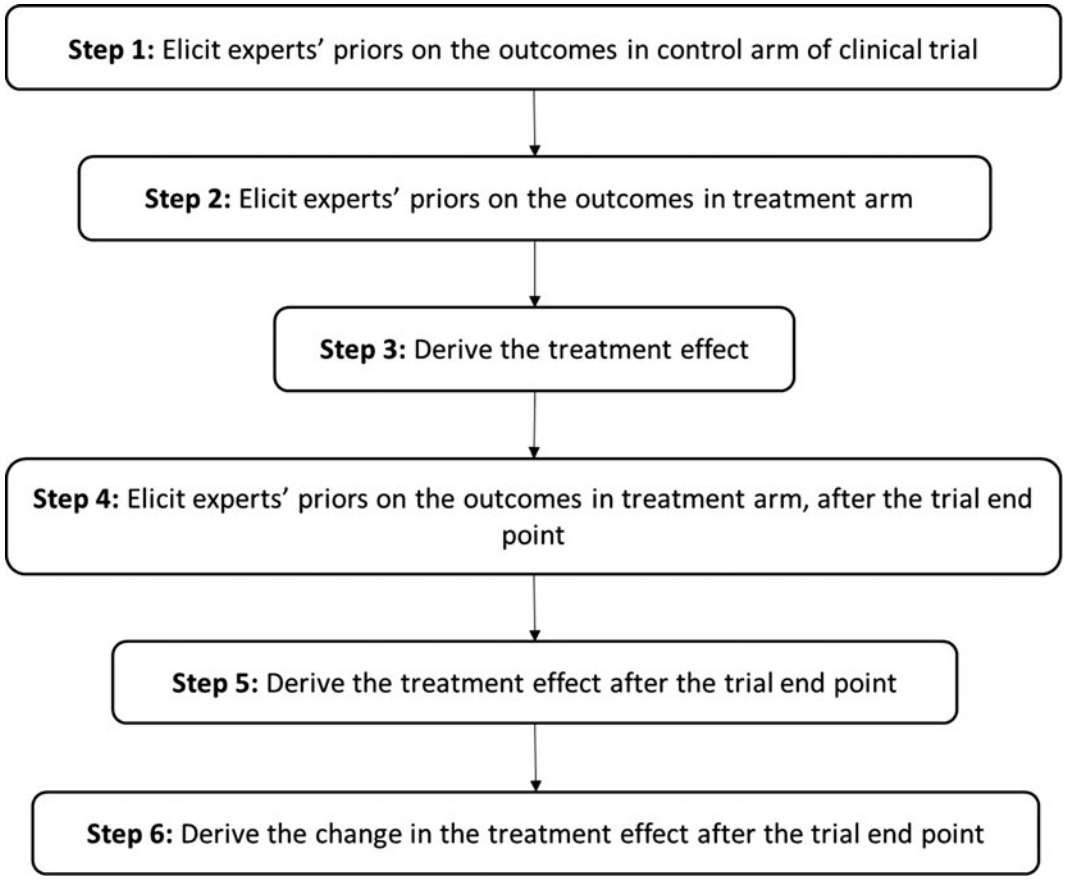


Figure 1. Methods for eliciting the treatment effect and changes in the treatment effect.

Step 2: Eliciting 1-Year Outcomes for the Podiatry Intervention

Step 1 was repeated to elicit the rate of falls and odds of fractures for the podiatry intervention, relative to the baseline (TAU). Experts were asked to express their beliefs about outcomes 1 year after starting the podiatry intervention, assuming that the proportions of falls and odds of fractures without the intervention were equal to their mode (a most likely scenario). The treatment effect was assumed to be independent of the baseline outcomes.

Step 3: Deriving the Treatment Effect 1 Year after Starting Treatment

The treatment effect 1-year after starting the podiatry intervention relative to TAU was measured in two ways: rate ratios for falls (RtR) and relative risk of having a fracture after a fall (RR).

Step 4: Capturing the Potential Impact of Time

Experts were asked whether they believed the effect of the podiatry intervention would change over time using multiple-choice questions (MCQs). Dependent on the answers to the MCQs, a sub-sample of experts were asked to elicit outcomes at a subsequent time-point, determined by the experts.

Experts were asked to elicit outcomes in patients who continued to receive the podiatry intervention after the trial end point, conditional on the outcomes in TAU remaining the same, to capture the change in the treatment effect. Age-related changes in falls and fractures were adjusted for at the analysis stage (not reported in this paper).

Step 5: Deriving the Treatment Effect at Second Time Point

The treatment effect for the podiatry intervention at a follow-up time point after the trial completion was derived using the process described in Step 3.

Step 6: Deriving the Temporal Change in the Treatment Effect

The change in the treatment effect (ΔTE) was assumed to be linear and relative to the treatment effect observed in the trial of the podiatry intervention, derived using the following equation:

$$\Delta\mathrm{TE} = \frac{\mathrm{TE}_{t2} - \mathrm{TE}_{t1}}{\mathrm{TE}_{t1}} \quad (1)$$

where TE_{t1} indicates the treatment effect 1 year after starting the podiatry intervention, and TE_{t2} indicates the treatment effect at the second time point.

The second time point at which the treatment effect was elicited, $t2$, varied between experts by design. To make experts' priors on the change in the treatment effect comparable, ΔTE was used to derive the annual change in the treatment effect, ΔATE using the following equation:

$$\Delta\mathrm{ATE} = \frac{\Delta\mathrm{TE}}{t2} \quad (2)$$

The treatment effect can take any value between zero and infinity, and so ΔTE and ΔATE could take any value between −1 and infinity, where $\Delta\mathrm{ATE} < 0$ indicates the treatment effect

https://doi.org/10.1017/S0266462322000022 Published online by Cambridge University Press

would decrease (potentiate when $TE_{t1}<1$, depreciate when $TE_{t1} > 1$), $\Delta ATE = 0$ indicates no change in the treatment effect, and $\Delta ATE > 0$ indicates that the treatment effect would increase over time (depreciate when $TE_{t1} < 1$, potentiate when $TE_{t1} > 1$).

Selecting the Experts

A minimum sample of thirty experts was set as the target sample size based on feasibility while ensuring a sufficient sample size to reflect the views of a wide range of experts. The protocol in Supplementary File 1 describes the expert recruitment process.

Collecting the Expert Beliefs

Data was collected in September and October 2016. The elicitation was conducted with each expert individually to avoid biases such as "peer-pressure" potentially introduced using a group-approach and to capture the variation in beliefs between different experts. Experts were given the choice of completing the exercise with the help of the investigator, either in person or on the phone. Experts did not receive any financial rewards for completing the exercise but were provided with lunch if completing the exercise in person.

Data for the expert elicitation exercise were collected using a bespoke web application (the elicitation tool) produced in Shiny package for R (22), which is an extension of the MATCH code developed by Morris et al. (23). The elicitation tool trained experts on concepts of uncertainty and how to express their beliefs in the required format, before taking them through the elicitation questions (see protocol in Supplementary File 1 for details).

Experts' priors were elicited using the "Chips and Bins" method (24), suggested to be more intuitive and at least as effective as other commonly used methods in experts not trained in probabilities and statistics (14). The Chips and Bins method provides experts with a range of possible parameter values on the *x*-axis of a grid and asks them to distribute the "chips" across the intervals ("bins") to indicate their uncertainty and generate a "histogram" (see protocol in Supplementary File 1 for examples of questions).

Prior Aggregation

The objective of data aggregation was to generate one probability distributions for each treatment effect (RtR and RR). The priors elicited from individual experts were aggregated mathematically, using unweighted linear pooling (24;25).

Data Analysis

Data analysis focused on four elements: comparison of experts' priors to trial outcomes; analysis of experts' priors on the change in the treatment effect; extrapolating the treatment effect; and sensitivity analysis.

Comparison of Experts' Priors to Trial Outcomes

The baseline outcomes and the treatment effect were compared with the results of the trial of the podiatry intervention using four summary statistics describing the proportion of experts:

- whose median values differed (lower/higher) to the observed trial value (representing under/over-estimate of treatment effect);
- who included the observed value in their 50 percent credible interval (CrI) or plausible range (representing overconfidence);
- whose values did not overlap with the 95 percent CI in the trial; and
- who included the entire 95 percent CI in their plausible range.

Analysis of Experts' Priors on the Change in the Treatment Effect

The derived annual changes in rate ratios and relative risk were described by summarizing qualitatively whether experts believed the treatment effect would depreciate (or potentiate) over time, and the magnitude of the change. The direction of change in the treatment effect implied by experts' medians was compared with their stated responses to assess the internal consistency of the priors.

Extrapolating the Treatment Effect

The treatment effect at different time points t was derived using the following equation.

$$TE_t = TE_{t-1}(1 + \Delta ATE) \quad (3)$$

Negative treatment effect predictions were manually adjusted to zero. When experts believed that the treatment effect would depreciate over time, the treatment effect was truncated at 1, assuming it never changed direction (e.g., from beneficial to harmful).

ΔATE was adjusted to standardize its interpretation. As described in Step 6 above, a positive ΔATE value could indicate both potentiation and depreciation of the treatment effect, depending on whether the podiatry intervention was beneficial or harmful. For extrapolation, it was assumed that experts who believed the treatment effect would depreciate over time believed so even if the direction of the treatment effect indicated in their prior was inaccurate (i.e., the median rate ratio was greater than 1, or the median risk ratio was less than 1). Therefore, when an expert's prior on the treatment effect was inaccurate, their prior on ΔATE was inverted before being applied to the observed treatment effect in Equation 3.

The predicted rate of falls and the risk of fractures derived from priors are presented graphically. The differences were analyzed qualitatively, describing how experts' beliefs, and assumptions made in the analysis, affected the predicted rate of falls.

Sensitivity Analyses

One-way sensitivity analyses (see Supplementary File 3) were used to understand the impact of changing two key assumptions used to derive the temporal change in the treatment effect:

- conditional independence between different outcomes (1–5, 6–10, and >10 falls), and between the treatment effect of the podiatry intervention and the baseline outcomes;
- the rate of falls derived from experts' priors on the probabilities of falling, accurately represented their beliefs.

Results

A total of 38 experts completed the SEE exercise (see Supplementary File 4 for sample details). Of these, three individuals could not provide values for the risk of fractures due to technical difficulties with the software.

Experts' Beliefs about the Treatment Effect 1 Year after Starting the Intervention

Table 1 compares experts' priors on the baseline rate of falls, odds of fracture, and the treatment effect (RtR for falls and RR for

https://doi.org/10.1017/S0266462322000022 Published online by Cambridge University Press

Table 1. Comparison of experts' priors and trial outcomes for four elicited parameters.

	Rate of falls ($n = 38$)	Rate ratio ($n = 38$)	Odds of fracture ($n = 35$)[a]	Risk ratio ($n = 35$)[a]
Mean observed in the trial (95% CI)	1.57 (1.37–1.78)	.88 (.73–1.05)	55.9 (42–71)	1.32 (.65–2.76)
Experts' medians: mean (range)	1.05 (.09–2.92)	.84 (.31–2.91)	22.8 (1–149)	1.13 (.31–3.42)
Median of pooled prior	.82	.73	12.8	.99
Experts' plausible range	0–7.05	0–8.67	0–199	.08–19.67
Plausible range of pooled prior	0–6.20	.01–7.62	0–199	.09–15.65
Number (proportion) of experts whose *median* was *lower than the observed value*	29 (.76)	25 (.66)	31 (.89)	29 (.83)
Number (proportion) of experts whose *median* was *higher than the observed value*	9 (.24)	13 (.34)	4 (.11)	6 (.17)
Number (proportion) of experts who *included the observed value in their 50% CrI*	3 (.08)	13 (.34)	1 (.03)	7 (.2)
Number (proportion) of experts who *included the observed value in their plausible range*	20 (.53)	27 (.71)	3 (.09)	27 (.77)
Number (proportion) of experts whose *LL was above* and *UL was below the lower 95% CI*, respectively	1 (.03), 11 (.29)	1 (.03), 3 (.08)	2 (.06), 23 (.66)	0 (0), 0 (0)
Number (proportion) of experts who *included the entire 95% CI in their plausible range*	15 (.39)	23 (.61)	1 (.03)	11 (.31)

CrI, credible interval; LL, lower limit; UL, upper limit; CI, confidence interval.
[a]Three experts did not complete this part of the exercise due to technical difficulties.

fractures) with the observed trial results (further diagrams provided in Supplementary File 5).

On average, experts' medians on the rate of falls and odds of fractures (1.05 and 22.8, respectively) were lower than those observed in the trial (1.57 and 55.9, respectively). The experts predicted the nature of the treatment effect correctly, as they believed the podiatry intervention would decrease the rate of falls (mean RtR = .84, 95 percent CI, .31–2.91) and increase the risk of fractures (mean RR = 1.13, 95 percent CI, .31–3.42).

Analyzing the elicited CrIs, experts were more likely to include the observed treatment effect in their CrIs (71 and 77 percent of priors on RtR and RR, respectively) than the observed baseline outcomes (53 and 9 percent of priors, respectively). Furthermore, experts were more likely to have some overlap between their prior and the observed CI when assessing the treatment effect (89 percent of priors on RtR, 100 percent of priors on RR) than the baseline outcomes (68 percent of priors on the rate of falls, 28 percent prior on RR).

Overall, the uncertainty across all experts combined was greater than that observed in the trial, as the elicited medians and plausible ranges both covered a wider range than the trial-based CIs, for all four parameters. However, few experts included the full CI in their prior (61 percent of experts for RtR and 3–39 percent of experts for all other parameters).

Experts' Beliefs about the Change in the Treatment Effect

In general, experts were certain that the treatment effect would change (36/38 experts) and diminish after the trial (32/38 experts), and the mean time after which the treatment effect would diminish completely was 3.0 years. Two experts believed the effect of the podiatry intervention would potentiate, and plateau after 3.2 years (mean), while two were uncertain whether it would potentiate or diminish. One expert was uncertain whether the treatment effect would change, and one was certain it would stay the same.

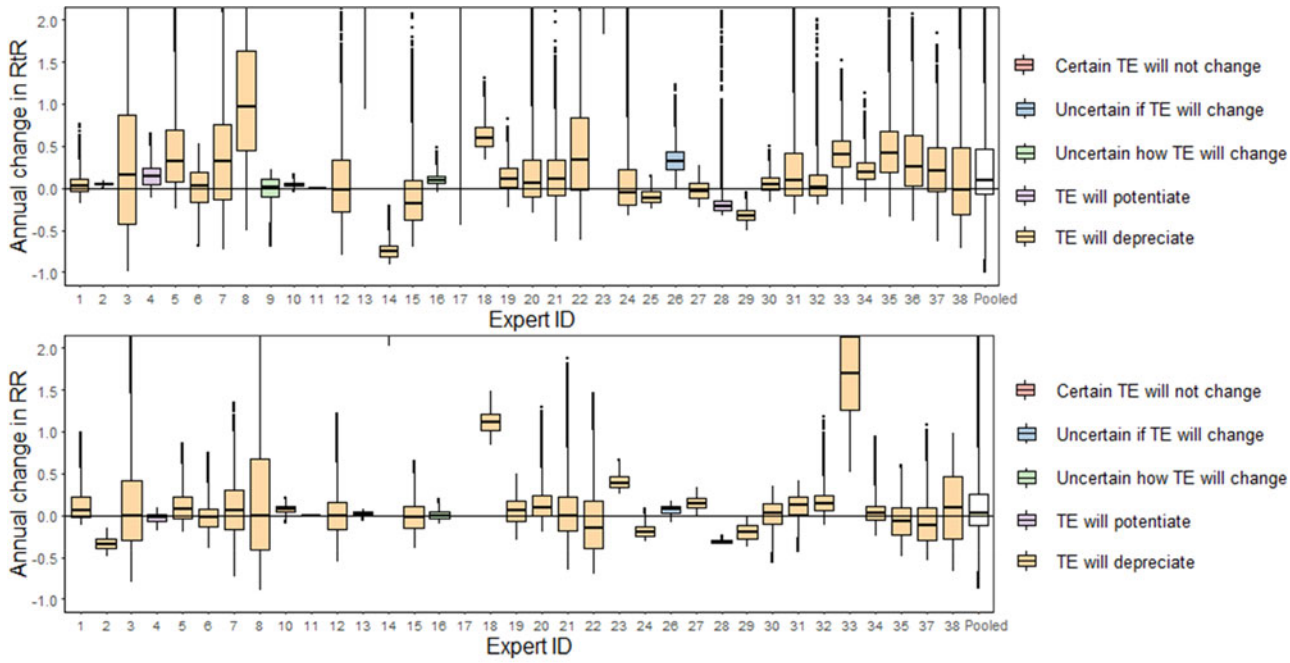
Figure 2 shows experts' priors on the annual change in the treatment effect. Priors elicited from thirteen experts were inconsistent with their stated treatment effects. Twelve experts believed that the treatment effect would depreciate, yet their priors (medians) suggested the opposite, for at least one of the two outcomes. One expert (Expert 26 in Figure 2) was uncertain whether the treatment effect would change over time, yet their priors indicate that they were confident that the treatment effect would diminish.

Extrapolation: Combining Trial Outcomes and Experts' Priors

Figure 3 shows the rate of falls and the risk of fractures observed in the trial, extrapolated over time using experts' priors. By year 5, the median predicted the rate of falls was between 1.3 and 1.5 for 87 percent (33/38) of experts, and for 97 percent (33/35) the risk ratio was between .04 and .09.

Three priors led to predicted outcomes outside the possible range of the parameter—one for the rate of falls (predicted RtR < 0) and two for fractures (predicted RR > 1). The majority of experts believed the treatment effect would depreciate. The mean number of years for the predicted rate of falls to revert back to pretreatment levels was 3.2 years, compared with the mean 3.0 years in their stated responses. For those experts who believed the treatment effect would depreciate both in their stated responses and in their priors, the mean difference between time taken for the treatment effect to diminish in verbal responses and priors was −.08 years (range: −4 to 3 years). For the risk of fracture, fourteen experts expressed priors that reverted back to pretreatment risk. Of these fourteen experts, the mean number of years it took for the treatment effect to diminish was 4.9 years, and the mean difference between time taken for the

https://doi.org/10.1017/S0266462322000022 Published online by Cambridge University Press



Box = median and interquartile range, whiskers = range. Horizontal line = no change in treatment effect. Box colour indicates experts' stated beliefs. RR = risk ratio; RtR = rate ratio; TE= treatment effect.

Figure 2. Experts' priors on the annual change in the rate ratio and relative risk.

treatment effect to diminish in verbal responses and priors was 2.6 years (range: −1 to 8 years).

Figure 4 shows the predicted rate of falls and the risk of fractures over time, derived using experts' aggregate priors compared with two alternative scenarios—indefinite treatment effect and treatment effect diminishing 2 years after starting treatment. The median rate of falls and risk of fractures derived from experts' priors were between the two scenarios, as experts generally believed that the treatment effect would depreciate gradually. The point estimates in all three scenarios are comparable (likely because the treatment effect in the trial was small); however, uncertainty was much greater when priors were combined with trial results than in the two alternative scenarios when uncertainty was extrapolated from short-term data.

Sensitivity Analysis

The two one-way sensitivity analyses suggested no correlation between experts' beliefs about the conditional probabilities, and between the baseline outcomes and the treatment effect. Furthermore, results were not sensitive to the method used to derive the rate of falls from their priors (see Supplementary File 6 for details).

Discussion

This study used SEE to inform uncertainty in the treatment effect over time for a multifaceted podiatry intervention designed to reduce the rate of falls and fractures in the elderly. The outcomes (rate of falls and risk of fractures) over time were derived by combining results from a published trial and the elicited priors on the temporal change in the treatment effect.

Key Findings

The experts' median predicted rate of falls and risk of fractures (Figure 4) were within the range of those derived from two contrasting assumptions about the trajectory of the treatment effect observed in the trial (indefinite treatment effect and immediate reverting to pretreatment outcomes). The priors, however, implied greater uncertainty than either of the two alternative assumptions.

The between-expert variation in predicted outcomes (in Figure 4) was relatively small, probably because the treatment effect in the trial was small and the majority of experts believed the effect would depreciate, leaving a relatively narrow range of plausible values for the treatment effect.

The internal consistency of the elicited priors varied. Priors elicited from fourteen (out of thirty-eight) experts were not consistent with their stated responses. Furthermore, the consistency between the time taken for the outcomes to revert back to pretreatment levels implied by the priors, and by experts' stated responses differed for the two outcomes. The elicited priors on the rate of falls were fairly consistent but priors elicited for the risk of fracture were not consistent. This probably occurred because experts underestimated the effect of the podiatry intervention on fractures (compared with the trial), and so the elicited change in the treatment effect was more impactful when applied to the effect observed in the trial.

The study elicited multiple outcomes (rate of falls and odds of fracture, for different comparators) allowing a comparison between experts' assessments of the different types of parameters. On visual inspection, the results shown in Table 1 implied that experts' priors on the treatment effect generally had greater overlap with trial results than priors on baseline outcomes. Furthermore, priors on the rate of falls were more likely to overlap with trial results than those on odds of fractures. It is not clear

https://doi.org/10.1017/S0266462322000022 Published online by Cambridge University Press

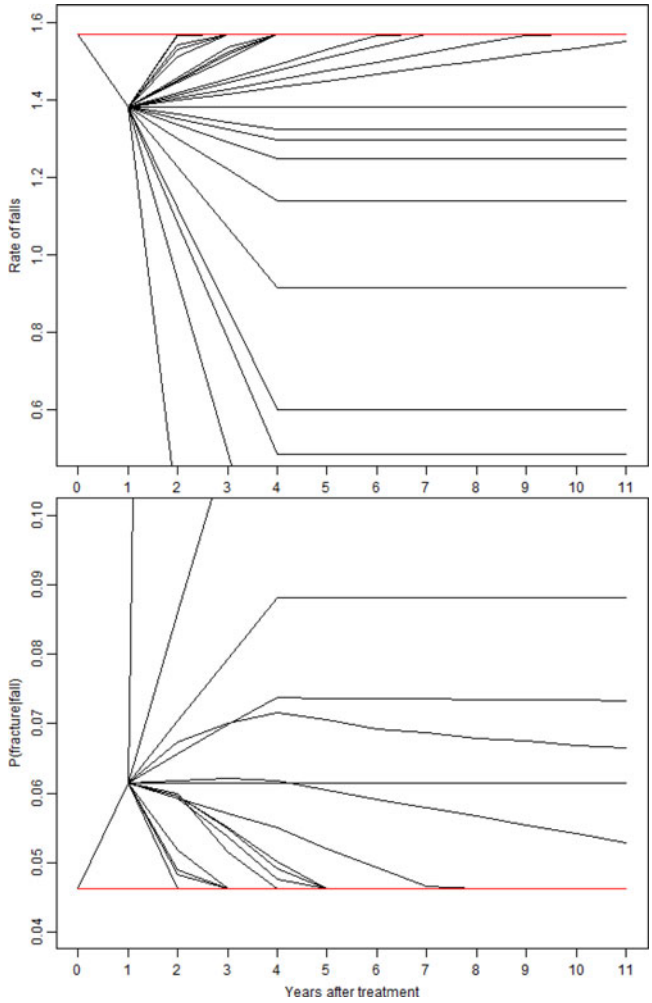


Figure 3. Median outcomes over time, derived from trial results and experts' priors. For experts who believed the treatment effect would potentiate, the predicted rate of falls and the risk of fractures were assumed to plateau after 4 years. The red line indicates no effect.

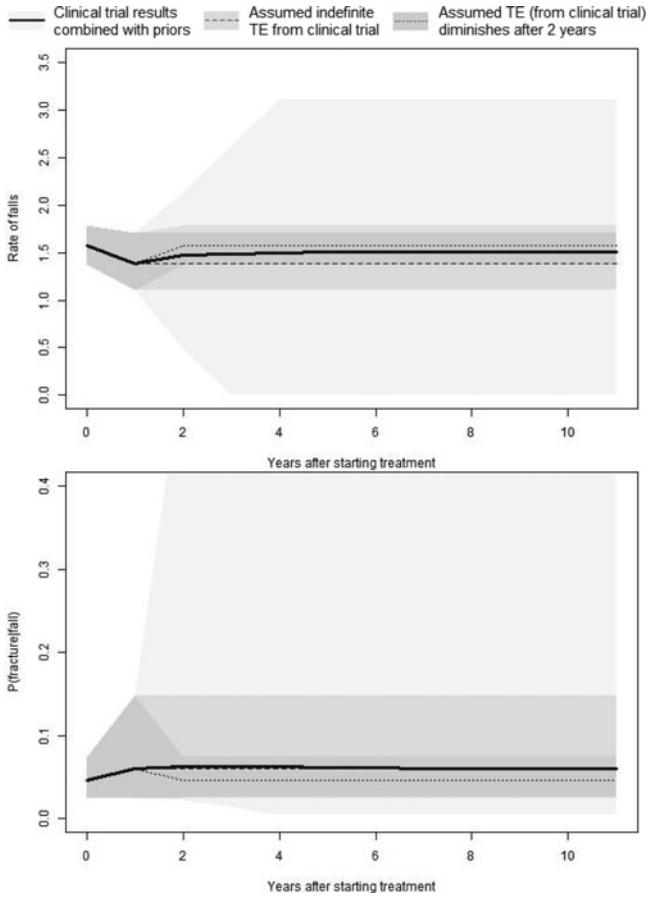


Figure 4. Uncertainty in outcomes over time, derived from trial results and experts' aggregate priors. Line = median. Shading = 95% CI. TE = treatment effect.

from the findings whether this result was because experts found rates easier to assess than odds, or whether the observed rate was simply more representative of the population they observed in practice.

Comparison with the Existing Literature

We are not aware of any published studies where rates and odds are elicited for use in MBEE in health care (26). We are aware of five other studies that elicited temporal uncertainty in this context, by eliciting clinical outcomes (multinomial (11) and continuous (12)), the relative treatment effect (9;10), and survival (13) at multiple time points. The existing studies differed in how they dealt with dependency between outcomes at different time points. One study (12) explicitly elicited the dependency by asking experts to revise their priors on "long-term" outcomes, after varying "short-term" outcomes. In the remaining four studies (9–11;13), experts were presented with the parameter values at one time point so their priors were relative to the observed probability distributions, but no correlation was assumed when sampling values at each time point for use in the probabilistic decision model. Their approach was not applicable to this study without assuming independence between outcomes at different time points, as trial results were not available at the time of elicitation.

Limitations

Some simplifications were necessary to deliver the results in a timely manner and to reduce burden on experts, such as the assumptions of conditional independence between outcomes and the treatment effect, and the treatment effect and its temporal change. Sensitivity analyses detailed in Supplementary File 6 tested for between-expert correlation between the outcomes assumed to be independent. The results suggest that the assumptions were plausible; however, it is not possible to test for within-expert correlation *post hoc*, for example, whether experts would have adjusted their estimates of the treatment effect for different values of the baseline rate of falls. It is not clear whether the assumptions made represent the most appropriate methods to ensure that the resulting priors represent experts' beliefs.

In addition, assumptions were made when applying the change in the treatment effect to the effect observed in the trial. The change in the treatment effect was assumed to be linear over time. When experts' beliefs about the direction of the treatment effect were incorrect, we assumed that experts' beliefs about the nature of change in the treatment effect (i.e., whether it would potentiate or depreciate) did not change. This assumption was applied to many priors because the observed treatment effect (both for falls and fractures) was not statistically significant, and

https://doi.org/10.1017/S0266462322000022 Published online by Cambridge University Press

so random samples from their 95 percent CI indicated the intervention could be both beneficial and harmful.

This study did not explore reasons for internal inconsistencies in experts' values. Experts were able to update their priors at any point in the exercise but they were not provided with verbal feedback about what their priors implied, in terms of the treatment effect and its change. The majority of the priors that lacked internal consistency implied that the treatment effect would potentiate rather than diminish over time, potentially underestimating the rate at which the rate of falls and fractures would return to pretreatment levels. There is no guidance in elicitation literature on whether such priors should be included in pooled estimates. The exclusion of priors that lacked internal consistency and other basis for differential weighting is a topic for further research.

Policy Implications

This study shows that SEE can be used to inform temporal uncertainty for use in MBEE. Without experts' input, uncertainty in the treatment effect is assumed to be stable over time (8)—an assumption that is inappropriate in many instances. The elicitation of experts' uncertainty provides an alternative to extrapolation from short-term data alone (in this case, the trial). Experts may observe patients over a longer term compared with experimental evidence, or apply knowledge about comparable interventions to assess plausible long-term outcomes.

The study also demonstrates an approach to eliciting temporal uncertainty before short-term outcomes have been observed while allowing for correlation between conditional outcomes at different time points. The methods are applicable for elicitation conducted at early stages of intervention development, for example, earlier in the development pathway for medicines.

Conclusion

This study suggests that using an SEE exercise can be used to characterize uncertainty in extrapolation of the treatment effect. However, accounting for correlation between outcomes at different time points is complex, and likely to require simplifications. Further evidence, focusing on applied examples, is needed to inform the optimum method for eliciting temporal uncertainty using elicited priors.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0266462322000022.

Acknowledgments. We are grateful to all experts who took time out of their busy days to take part in this study, and those who encouraged their colleagues to do so. This includes, in no particular order, Katie Robinson at AGILE, Chartered Society of Physiotherapy, Catherine Anguish, Marie Clarke, Jayne Bunday and Susan Briggs from the Derbyshire Community Health Service, Ceri Griffiths and the friendly team from the Tenby Cottage Hospital in Wales, Dr Alastair Dickson, Julia Gray from Manchester University NHS Foundation Trust, Dr Helen Hawley-Hague from the University of Manchester, Jo Jennings from the South Warwickshire NHS Foundation Trust, Tessa Jervis from Sheffield Teaching Hospitals NHS Foundation Trust, Dr Jonathan Treml from the University Hospitals Birmingham, Dr Jane Youde from the Derby Teaching Hospitals NHS Foundation Trust, and all experts who preferred to stay anonymous. We would also like to thank Geraint Collingridge, Dr James Reid, Dr Huma Naqvi and Dr David Broughton at the British Geriatrics Society who engaged their membership and provided opportunities for in-person engagement, and all BGS members who participated.

Funding. Financial support for this study was provided in part by a grant from NIHR CLAHRC YH and the NIHR Applied Research Collaboration Yorkshire and Humber. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

Conflict of Interest. There are no conflicts of interest.

References

1. **Sculpher MJ, Claxton K, Drummond M, McCabe C**. Whither trial-based economic evaluation for health care decision making? *Health Econ*. 2006;**15**:677–87.
2. **Claxton K**. Exploring uncertainty in cost-effectiveness analysis. *Pharmacoeconomics*. 2008;**26**:781–98.
3. **Petrou S, Gray A**. Economic evaluation using decision analytical modelling: Design, conduct, analysis, and reporting. *BMJ*. 2011;**342**:d1766.
4. **Philips Z, Bojke L, Sculpher M, Claxton K, Golder S**. Good practice guidelines for decision-analytic modelling in health technology assessment. *Pharmacoeconomics*. 2006;**24**:355–71.
5. **Claxton K, Sculpher M, McCabe C, Briggs A, Akehurst R, Buxton M, et al.** Probabilistic sensitivity analysis for NICE technology assessment: Not an optional extra. *Health Econ*. 2005;**14**:339–47.
6. **Griffin SC, Claxton KP, Palmer SJ, Sculpher MJ**. Dangerous omissions: The consequences of ignoring decision uncertainty. *Health Econ*. 2011;**20**:212–24.
7. **Mahon R**. Temporal uncertainty in cost-effectiveness decision models: Methods to address the uncertainties that arise when the appropriate analysis time horizon exceeds the evidence time horizon in cost-effectiveness decision models as applied to healthcare. University of York; 2014. Available from: http://etheses.whiterose.ac.uk/8268/1/Ronan Mahon Thesis %28Feb2015%29.pdf [accessed 2020 Dec 14).
8. **Bojke L, Claxton K, Sculpher M, Palmer S**. Characterizing structural uncertainty in decision analytic models: A review and application of methods. *Value Health*. 2009;**12**:739–49.
9. **Stevenson MD, Oakley JE, Lloyd Jones M, Brennan A, Compston JE, McCloskey EV, et al.** The cost-effectiveness of an RCT to establish whether 5 or 10 years of bisphosphonate treatment is the better duration for women with a prior fracture. *Med Decis Making*. 2009;**29**:678–89.
10. **Stevenson MD, Oakley JE, Chick SE, Chalkidou K**. The cost-effectiveness of surgical instrument management policies to reduce the risk of vCJD transmission to humans. *J Oper Res Soc*. 2009;**60**:506–18.
11. **Wilson ECF, Usher-Smith JA, Emery J, Corrie PG, Walter FM**. Expert elicitation of multinomial probabilities for decision-analytic modelling: An application to rates of disease progression in undiagnosed and untreated Melanoma. *Value Health*. 2018;**21**:669–76.
12. **Bojke L, Claxton K, Bravo-Vergel Y, Sculpher M, Palmer S, Abrams K**. Eliciting distributions to populate decision analytic models. *Value Health*. 2010;**13**:557–64.
13. **Cope S, Ayers D, Zhang J, Batt K, Jansen JP**. Integrating expert opinion with clinical trial data to extrapolate long-term survival: A case study of CAR-T therapy for children and young adults with relapsed or refractory acute lymphoblastic leukemia. *BMC Med Res Methodol*. 2019; **19**:182.
14. **Bojke L, Soares MFO, Claxton KP, Colson A, Fox A, Jackson C, et al.** *Developing a reference protocol for expert elicitation in healthcare decision making: A mixed-methods study*. Southampton, UK: NIHR Journals Library; 2021. (Health Technology Assessment, No. 25.37). doi:10.3310/hta25370.
15. **Iglesias C, Thompson A, Rogowski W, Payne K**. Reporting guidelines for the use of expert judgement in model-based economic evaluations. *Pharmacoeconomics*. 2016;**34**:1161–72.
16. **Corbacho B, Cockayne S, Fairhurst C, Hewitt CE, Hicks K, Kenan A-M, et al.** Cost-effectiveness of a multifaceted podiatry intervention for the prevention of falls in older people: The REducing falls with orthoses and a multifaceted podiatry intervention trial findings. *Gerontology*. 2018; **64**:503–12.

https://doi.org/10.1017/S0266462322000022 Published online by Cambridge University Press

17. **Cockayne S, Adamson J, Clarke A, Corbacho B, Fairhurst C, Green L, et al.** Cohort randomised controlled trial of a multifaceted podiatry intervention for the prevention of falls in older people (The REFORM Trial). *PLoS ONE*. 2017;**12**:e0168712.
18. **Kadane J, Wolfson LJ**. Experiences in elicitation [Read before The Royal Statistical Society at a meeting on "Elicitation" on Wednesday, April 16th, 1997, the President, Professor A. F. M. Smith in the Chair]. *J R Stat Soc Ser D Stat*. 1998;**47**:3–19.
19. **Spink MJ, Menz HB, Lord SR**. Efficacy of a multifaceted podiatry intervention to improve balance and prevent falls in older people: Study protocol for a randomised trial. *BMC Geriatr*. 2008;**8**:30.
20. **Spink MJ, Menz HB, Fotoohabadi MR, Wee E, Landorf KB, Hill KD, et al.** Effectiveness of a multifaceted podiatry intervention to prevent falls in community dwelling older people with disabling foot pain: Randomised controlled trial. *BMJ*. 2011;**342**:d3411.
21. **Clemen RT, Fischer GW, Winkler RL**. Assessing dependence: Some experimental results. *Manage Sci*. 2000;**46**:1100–15.
22. **Chang W, Cheng J, Allaire J, Xie Y, McPherson J**. shiny: Web application framework for R. R package version 1.5.0. 2020. Available from: https://cran.r-project.org/package=shiny [accessed 2020 Dec 14].
23. **Morris DE, Oakley JE, Crowe JA**. A web-based tool for eliciting probability distributions from experts. *Environ Model Softw*. 2014;**52**:1–4.
24. **O'Hagan A, Buck CE, Daneshkhah A, Eiser JR, Garthwaite PH, Jenkinson DJ, et al.** *Uncertain judgements: Eliciting experts' probabilities*. Chichester: John Wiley & Sons; 2006.
25. **Genest C, Zidek JV**. Combining probability distributions: A critique and an annotated bibliography. *Statist. Sci*. 1986;**1**:114–135.
26. **Soares MO, Sharples L, Morton A, Claxton K, Bojke L**. Experiences of structured elicitation for model-based cost-effectiveness analyses. *Value Health*. 2018;**21**:715–23.

https://doi.org/10.1017/S0266462322000022 Published online by Cambridge University Press