# Supplementary file 1. Structured expert elicitation protocol

## Background

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 intervention in the first year of treatment was evaluated in the REFORM trial (1,2). In the trial, the intervention led to a small reduction in the rate of falls (rate ratio = 0.88; 95% CI: 0.73 to1.05) and an increase in the risk of fractures after a fall (risk ratio = 1.32; 95% CI: 0.65 to 2.76), although neither was statistically significant.

The intervention was designed to be received indefinitely, with a potential effect on mortality through a reduced risk in falls. The appropriate time horizon for analysis was therefore a lifetime (3), and further analysis is required to extrapolate the REFORM trial results.

In the absence of empirical evidence, a SEE will be conducted to inform the treatment effect of the multifaceted podiatry intervention over time. Specifically, the treatment effect at different time points is assumed to be correlated, and so the aim is to elicit the treatment effect after the trial end point, relative to the treatment effect observed in the trial. The trial results will not be available when the SEE is conducted.

## Identifying experts

The exercise requires an understanding of falls behaviour in the elderly and how much foot and ankle health contributes to the risk of falls. The intervention is delivered by podiatrists; however, they are unlikely to observe the effects of the intervention. Additional professions were thus considered.

The aim is to recruit clinicians who satisfy the following criteria:

* have a good knowledge of foot and ankle physiology,
* understand the risk factors for falling,
* have a knowledge of fall prevention interventions and evidence behind them,
* have experience in delivering behavioural interventions (and knowledge of how patients respond to them).

The professions were identified by reading literature on fall prevention and observing professions of authors who published in the field, and by researching which professionals worked in fall prevention teams in UK hospitals. Experts were thus defined as geriatricians, physiotherapists, nurses and occupational therapists (OTs) who specialise in preventing falls or treating patients who have suffered fall related injuries.

It was noted that experts in these four professions are most likely to see patients with history of falling and fall related injuries. In order to capture beliefs of experts who see a broader population of patients at risk of falling, general practitioners (GPs) were also targeted.

Finally, health researchers whose research focuses on fall prevention were included to explore the effect of having less clinical experience but thorough knowledge of relevant literature on experts’ priors.

## Identifying the relevant elicitation parameters

The target parameter that requires elicitation is defined as the relative change in treatment effect over a life-time horizon. Published guidance for SEE in healthcare decision making (4) advises that the elicited quantities should be simple and observable (avoiding complex parameters such as ratios), that elicitation should capture dependence between variables, and that complex quantities may need to be decomposed into simpler quantities. A change in treatment effect is not directly observable in RCTs or clinical practice (hence an unobservable parameter) and so it is difficult to estimate accurately (5). The parameter of interest is, therefore, broken down into quantities that can be observed and measured: the defined outcomes for patients that do (do not) receive the intervention.

Two outcomes to capture treatment effect will be elicited - rate of falls and the risk of having a fracture after a fall (risk of fractures thereafter). The treatment effect will be derived by eliciting outcomes with and without treatment. Two treatment effects will be used to derive the temporal change in the treatment effect: the treatment effect at one-year after starting treatment (comparable to that observed in the RCT), and the treatment effect at a second (specified) time-point.

### Eliciting the relative change in treatment effect

The following six steps will be followed to elicit the relative change in treatment effect.

#### Step 1: Eliciting 1-year outcomes for the no treatment (baseline)

Eliciting the expected rate of falls is anticipated to be cognitively challenging, instead the rate of falls can be derived from the frequency distribution of falls in a population by weighting each possible number of falls by its probability. The frequency distribution of falls is skewed, with long tails depicting low but non-zero probability for high number of falls. Peterson and Miller (1964) found that, when eliciting means and medians of skewed distributions, experts’ mean estimates tend to be biased towards the median (i.e. not adjusted for outliers). Therefore, experts will be asked to elicit the multinomial distributions for the number of falls, and use these distributions to derive the rate of falls. This approach (described below) was piloted, alongside direct elicitation of the rate of falls (see Supplementary file 2).

Each individual patient can experience more than one fall in a given time-period (7) (potentially exceeding ten falls (7,8) per year). Eliciting the probability of more than ten falls would be time intensive. To reduce the burden on experts, possible outcomes will be grouped, confirmed to be reasonable by a physiotherapist specialising in fall prevention, into three categories capturing the conditional probability of falling: at least once (P( >0); more than five times P( >5| >0); more than ten times P( >10| >5), where is the number of falls. Conditional probabilities will be elicited to prevent statistical incoherence (e.g. P( >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 (9).

Experts’ priors will be elicited for each of the three categories of outcome as relative frequencies (denominator set at 1,000 to allow expression of probabilities less than 1%), as these have been suggested to be the most intuitive to represent interactions (two time points in each arm of the trial) between different outcomes (probabilities and conditional probabilities) (10).

Experts’ priors on conditional probabilities will be used to derive the rate of falls in the following three steps:

* derive joint probabilities P( >5) and P( >10) using Equation A1.1;
* predict the probability of having more than 1, 2, 3, 4, 6, 7, 8, 9, 11 or more falls using the regression model in Equation A1.2; and
* derive the probability of having exactly 1, 2, 3, 4, 6, 7, 8, 9, 11 or more falls using Equation A1.3.

|  |  |
| --- | --- |
|  | Equation A1.1 |
|  |  |
|  | Equation A1.2 |

Where logit (x) = log(x/(1-x)); the logit transformation was used to ensure the predicted probabilities do not exceed the limits of the parameter (0-1).

|  |  |
| --- | --- |
|  | Equation A1.3 |

The rate of falls was calculated by multiplying each number of falls by its corresponding probability. Ten thousand samples were drawn directly from each prior, to reflect experts’ uncertainty.

The risk of fracture after a fall will be 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. The odds will then be converted into probabilities using Equation A1.4.

|  |  |
| --- | --- |
|  | Equation A1.4 |

#### Step 2: Eliciting 1-year outcomes for the intervention

Step 1 was repeated to elicit the rate of falls and odds of fractures for the treatment, relative to the baseline (no treatment). Experts will be asked to express their beliefs about outcomes one year after starting treatment, assuming that the proportions of falls and odds of fractures without treatment are equal to their mode (most likely scenario). Treatment effect was assumed to be independent of the baseline outcomes.

#### Step 3: Deriving the treatment effect one-year after starting treatment

The treatment effect one-year after starting treatment will be measured in two ways: rate ratios for falls (RtR, see Equation A1.5) and relative risk of having a fracture after a fall (RR, see Equation A1.6).

|  |  |
| --- | --- |
|  | Equation A1.5 |

where is a random sample from the rate of falls with treatment, and is the mode rate of falls without treatment (derived from experts’ mode P(x>0), P(x>5|P>0) and P(x>10|P>5)).

|  |  |
| --- | --- |
|  | Equation A1.6 |

where and are random samples from the risk of fracture with and without treatment.

#### Step 4: Capturing the potential impact of time

Experts will be asked whether they believed the treatment effect would change over time using multiple-choice questions (MCQs), see section 4 for details. Dependent on the answers to these questions, a sub-sample of experts will be asked to elicit outcomes at a subsequent time-point, determined by the experts.

Experts will be asked to elicit outcomes in patients who continued to receive treatment after the trial end-point, conditional on the outcomes in the control arm remaining the same, to capture the change in the treatment effect. Age-related changes in falls and fractures will be adjusted for at the analysis stage.

#### Step 5: Deriving the treatment effect at second time point

The treatment effect at a follow-up time-point after the trial completion will be derived using the process described in Step 3.

#### Step 6: Deriving the temporal change in treatment effect

The change in the treatment () effect will be assumed to be linear and relative to the treatment effect observed in REFORM trial, derived using Equation A1.7.

|  |  |
| --- | --- |
|  | Equation A1.7 |

Where indicates treatment effect one year after starting treatment, and indicates treatment effect at the second time point.

The second time point at which the treatment effect will be elicited, *t2,* will vary between experts by design. To make experts’ priors on the change in the treatment effect comparable, ΔTE will be used to derive the annual change in treatment effect, using Equation A1.8.

|  |  |
| --- | --- |
|  | Equation A1.8 |

can take any value between zero and infinity, and so ΔTE and ΔATE could take any value between -1 and infinity, where ΔATE<0 indicates the treatment effect would decrease (potentiate when <1, depreciate when >1), ΔTE=0 indicates no change in the treatment effect, and ΔTE>0 indicates that the treatment effect would increase over time (depreciate when <1, potentiate when >1).

## Framing of the questions

Elicitation questions relating to outcomes of REFORM trial are shown in Box A1.1 and Box A1.2.

Box A1.1. Elicitation questions: Baseline outcomes.

#### *Question 1*

‘Consider 1000 patients over the age of 70, randomly selected in the UK, who participate in the trial but DO NOT RECEIVE the intervention.’

‘Out of 1000 people who participate in the trial and DO NOT receive the intervention, how many do you think will have a fall in one year, during the trial?’

#### *Question 2*

*‘Now let’s assume that out of 1000 patients exactly [L] fall at least once.’* *(Note that this is the number you stated to be the most likely in the grid above.)*

*‘How many out of these [L] individuals do you think will fall MORE THAN FIVE TIMES in one year?’*

#### *Question 3*

*‘Now let’s assume that out of 1000 patients exactly [L] fall at least once, and [M] fall more than five times.’*

*‘How many out of these [M] individuals do you think will fall MORE THAN TEN TIMES in one year?’*

#### *Question 4*

*‘This section aims to find out about the severity of falls.‘*

*‘For patients who do not receive the intervention, one in how many falls, on average, do you think will result in a fracture?’*

Box A1.2. Elicitation questions: Outcomes one year after starting treatment

#### *Question 1*

*‘Consider 1000 patients over the age of 70, randomly selected in the UK, who participate in the trial for a year and are offered to continue with the intervention for [T] years after the trial.’*

*‘After this period, how many of them do you think will have a fall in one year?’*

#### *Question 2*

*‘Now let’s assume that out of 1000 patients exactly [L] fall at least once.’*

*‘How many out of these [L] individuals do you think will fall MORE THAN FIVE TIMES in one year?’*

#### *Question 3*

*‘Now let’s assume that out of 1000 patients exactly [L] fall at least once, and [M] fall more than five times.’*

*‘How many out of these [M] individuals do you think will fall MORE THAN TEN TIMES in one year?’*

#### *Question 4*

*‘This section aims to find out about the severity of falls.‘*

*‘For patients who do receive the intervention, one in how many falls, on average, do you think will result in a fracture?’*

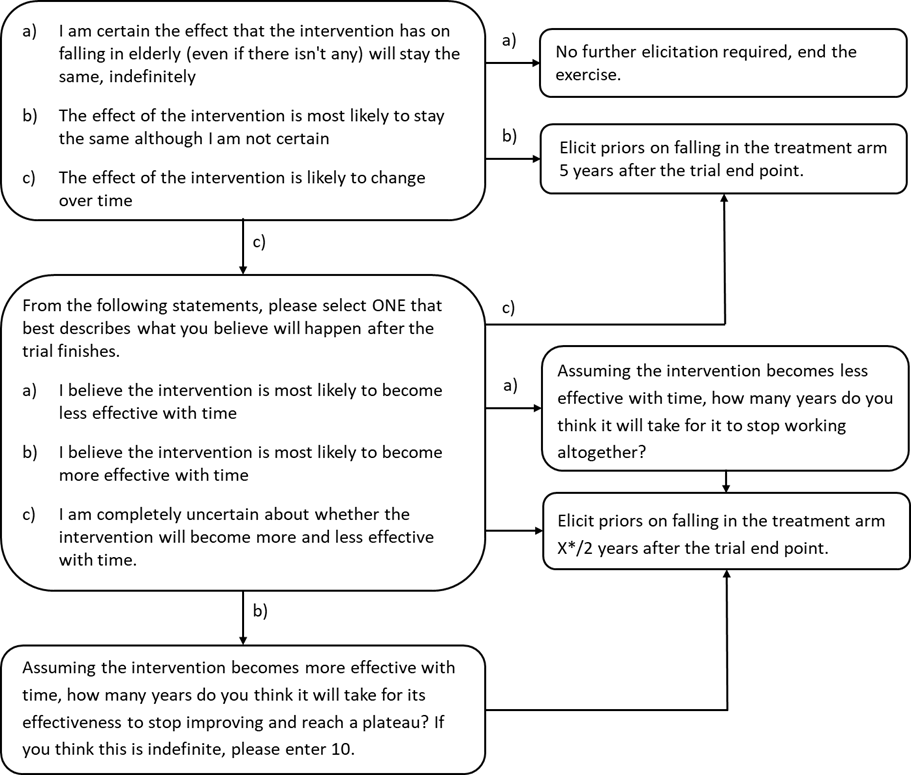
To elicit experts’ beliefs about outcomes after the REFORM trial, experts will be presented background information in Box A1.3, and a set of multiple choice questions (Figure A1.1). Experts who believe the treatment effect could change over time will be presented with a summary of their previous answers as shown in Box A1.4, before eliciting their beliefs.

**Box A1.3. Outcomes after the REFORM trial: Background information for experts.****Figure A1.1. MCQs and algorithm for the second time point for which probabilities were elicited.**

‘This section is about what happens after the trial has finished, given that patients remain enrolled in the intervention.’

‘Let's assume that all patients who were in the trial were offered to continue with the intervention in the same way: they continue to wear the appropriate footwear, foot orthoses (if required) and do self-directed exercise. Let's also assume that they no longer have to report their falling behaviour and they only see their podiatrist if required. We would like to know whether you believe that the intervention would have the same effect (if any) on the probability of falling as it did in the trial.’

‘Please take into account any factors that can influence the effect of the intervention on the risk of falling. This could be patient compliance after the trial ends, possibility that effectiveness of treatment will wear off, the possibility that the intervention will become more effective with time, or anything else that you think may be relevant.’



\* X = time point indicated by the expert.

Box A1.4. Assumptions about the rate of falls observed in the trial, provided to experts.

‘Questions in this section refer to what you think will happen [T] years after the trial finishes, provided all patients who DID RECEIVE the intervention in the trial (the treatment arm) are offered to continue with the intervention.’

‘We will assume that the risk of falling remains the same for those people who do not receive the intervention. We would like to know what you think will happen to those individuals who DO RECEIVE the intervention (the treatment arm) in the trial and are offered to continue with it after the trial has finished.’

‘In your responses, assume the following about trial outcomes:’

‘Out of 1000 patients who DO NOT receive the intervention [X] fall more than once, [Y] fall more than five times and [Z] fall more than ten times.’

‘Out of 1000 patients who DO receive the intervention [A] fall more than once [B] fall more than five times and [C] fall more than ten times.’

‘You can look back at these numbers at any point while answering the next question by clicking back on the 'Treatment after the trial' tab on the side panel.’

‘You can also change your responses to previous questions by clicking back on relevant tabs, but please note that these are not 'correct' answers. These numbers were obtained from your previous responses and are different for every expert who completes this exercise. If you chose to change your answers, please make sure you save the new answers.’

[T] is the second time point in years.

## Elicitation method

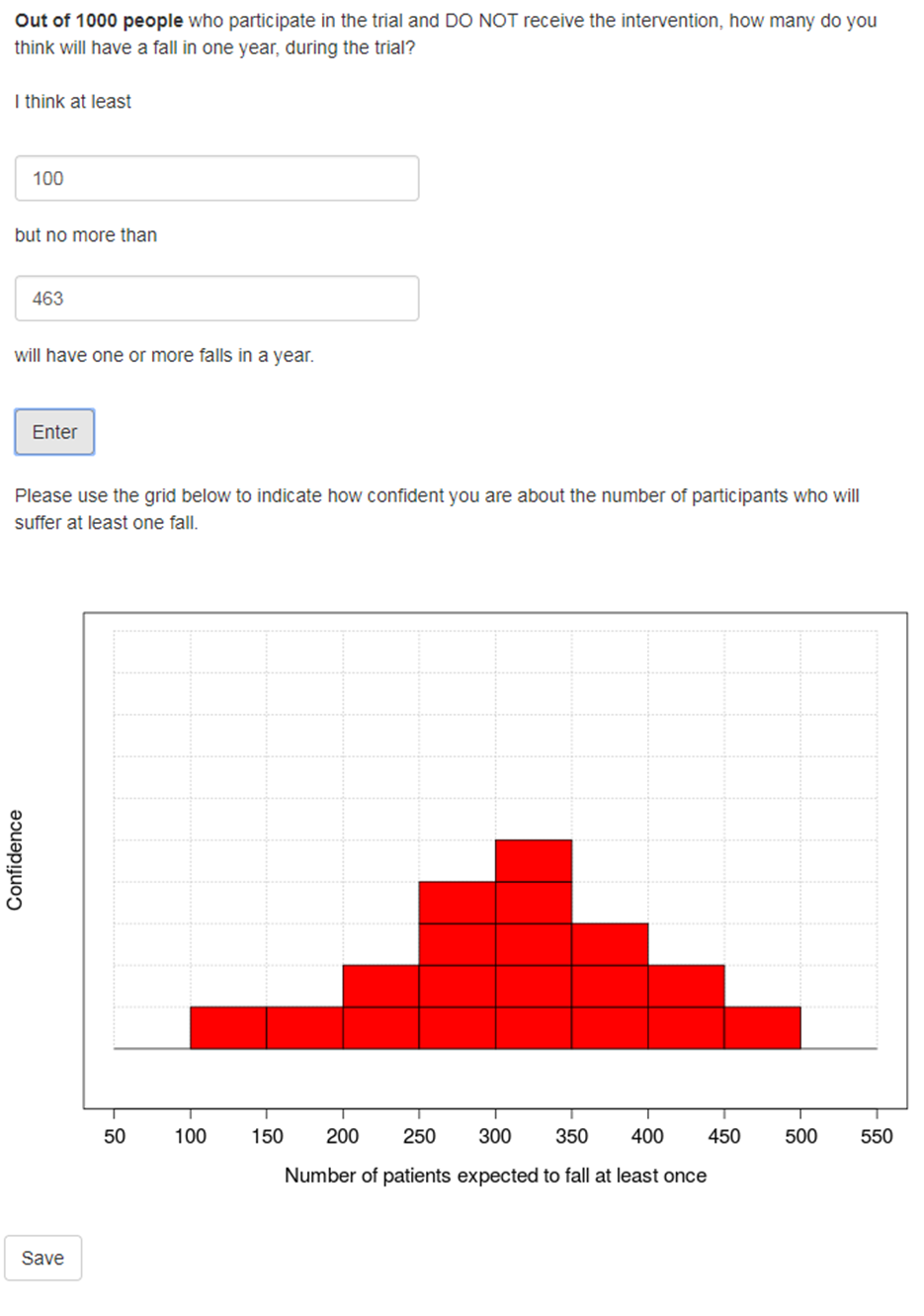
Experts’ beliefs will be elicited using the Chips and Bins (histogram, roulette) method (11).

For each elicited parameter, experts will first be asked to give the minimum and maximum plausible values to avoid overconfidence due to anchoring to the mean (5), and to narrow down the range of values on the chips and bins grid.

The x-axis on the grid will have a range of parameter values determined by the minimum and the maximum suggested by the expert. The bin width will be 1, 2, 5, 10, 20, 50 or 100, whichever of these resulted in as close to 10 bins as possible. Two extra bins outside the experts’ range will be added, unless they are outside the limits of the parameter. The grid will always be 10 bins high, and there is no limit on how many chips could be added to each grid. An example question is shown in Figure A1.2.

When experts entered the minimum and maximum, the program automatically selects the higher of the two values as the maximum and vice versa (even if an expert entered it the wrong way round) to prevent reverse order of values on the grid.

Figure A1.2. Application of Chips and Bins.



## Selecting the Experts

Participants will be identified via the following four avenues.

* Contacting clinicians who have published research in the field of fall prevention, in particular any studies that evaluate the effect of exercise and foot and ankle health on falling behaviour.
* Contacting members of the appropriate professional bodies including, but not restricted to, the Chartered Society of Physiotherapy, ASPIRE (a special interest group for physiotherapists working with elderly patients) and the British Geriatrics Society.
* Contacting individual fall clinics/departments in NHS trusts in England. Trusts will be chosen based on recommendation or geographical location and regional patient characteristics to give a heterogeneous sample of experts, as recommended in the literature (11).
* On recommendation by contacts gained through these bodies and by other research staff at the University of York.

All experts will be contacted via e-mail or phone, using publicly available details or those provided by those contacts who recommend them.

The target sample size is 30 to 50 experts. The sample size was decided to include a representative sample of experts from each profession. The upper limit of 50 was based on feasibility.

Information was collected on experts’ professional experience, including

* Their role. Experts are asked to describe their role in free text in as much detail as possible.
* Years of experience in current role, entered as number of years.
* Research experience, aiming to use experts’ contribution to the field as a potential indicator of expertise. Research experience was captured in two questions: the number of publications (0-3, 4-20 or 20-50 or over 50 publications) and the number of successful research grants co-written (0, 1-5 or more than 5) to explore the effect of different levels of research activity.
* Proportion of working time spent with patients who are at increased risk of falling, either helping prevent falls or treating fall related injuries. Categories were 0-10%, 11-30%, 31%-50% and 50%-100% to reflect different levels of patient contact.
* Awareness of any research into podiatry interventions designed to reduce the risk of falls (yes/no).

## Elicitation process

The elicitation will be conducted with each expert individually, to capture the variation in beliefs between different experts. All experts will be given the choice of completing the exercise on their own or with the help of the investigator, although the latter will be encouraged.

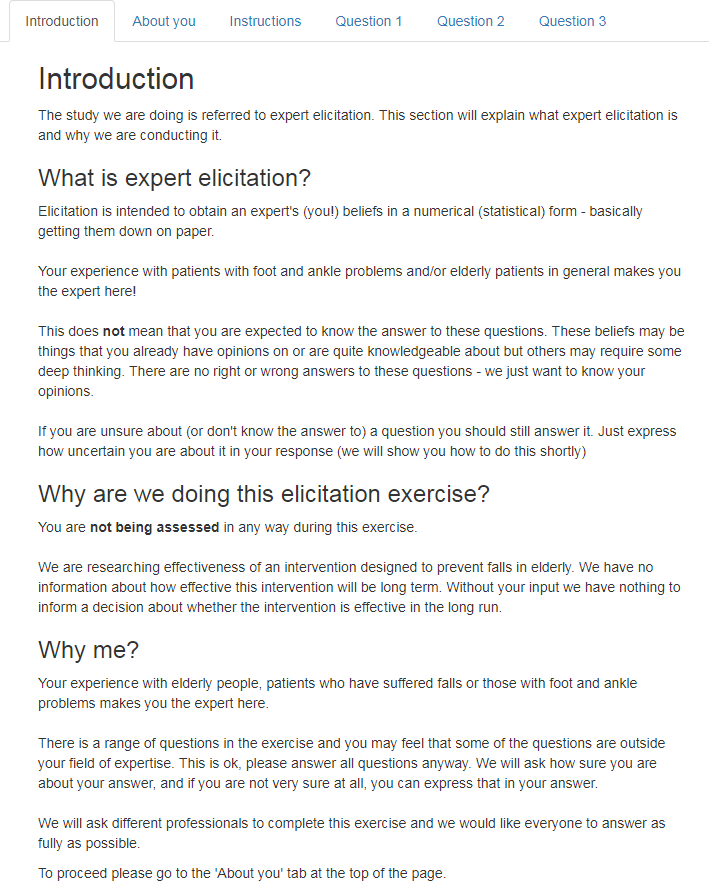
Data for the expert elicitation exercise will be collected using a bespoke web application produced in Shiny package for R (the elicitation-tool) (12), which is an extension of the MATCH code developed by Morris et al. (13). The elicitation-tool was piloted before the main study (see Supplementary file 2).

The elicitation-tool includes motivation and training (section 8), background information about the REFORM trial (section 9), and the elicitation questions (section 4). The elicitation tool guides experts through a series of sequential steps to ensure all questions are answered, with values suggested that were outside the plausible range triggering an error message.

## Motivation and training

Experts will be informed about the purpose of the elicitation exercise, as indicated in Figure A1.3.

Figure A1.3. Information about the project provided to experts.



Training will be provided at the beginning of the exercise and consist of two components - explaining uncertainty and teaching experts to use the chips and bins method.

### Explaining uncertainty

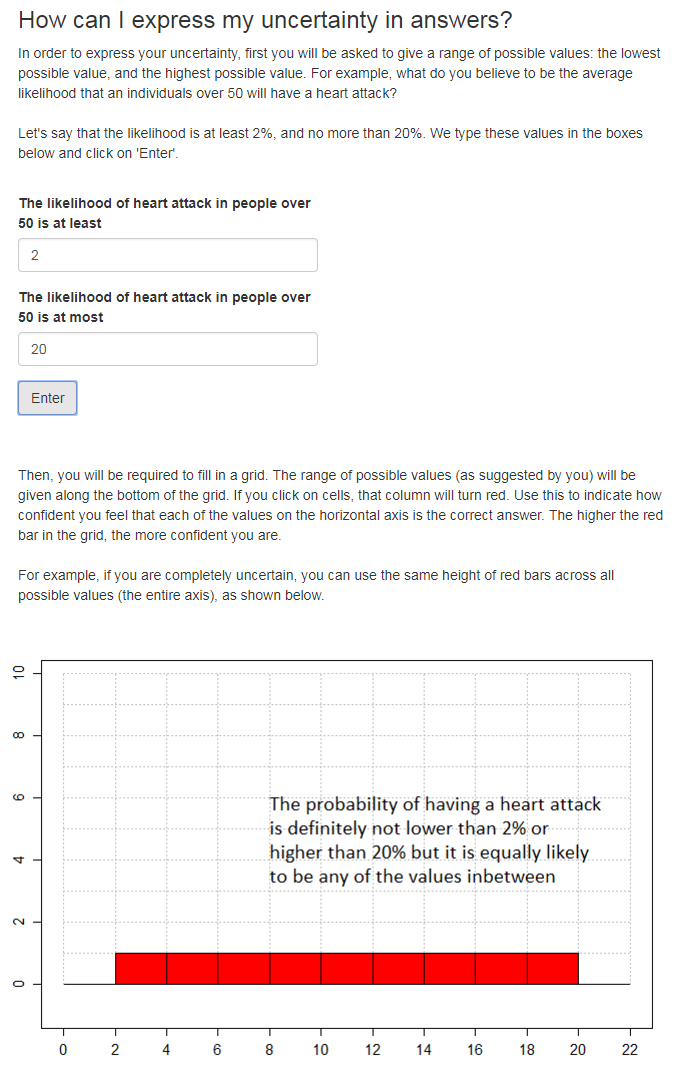
The aim of elicitation is to elicit uncertainty around parameter values, rather than heterogeneity or variability, as recommended by Bojke et al. (14). While this is to some extent affected by the question format, training will include a discussion about the difference between uncertainty and variability to help experts understand the difference. The information provided is shown in Figure A1.4.

|  |
| --- |
| Figure A1.4. Training on the difference between uncertainty and variability. |

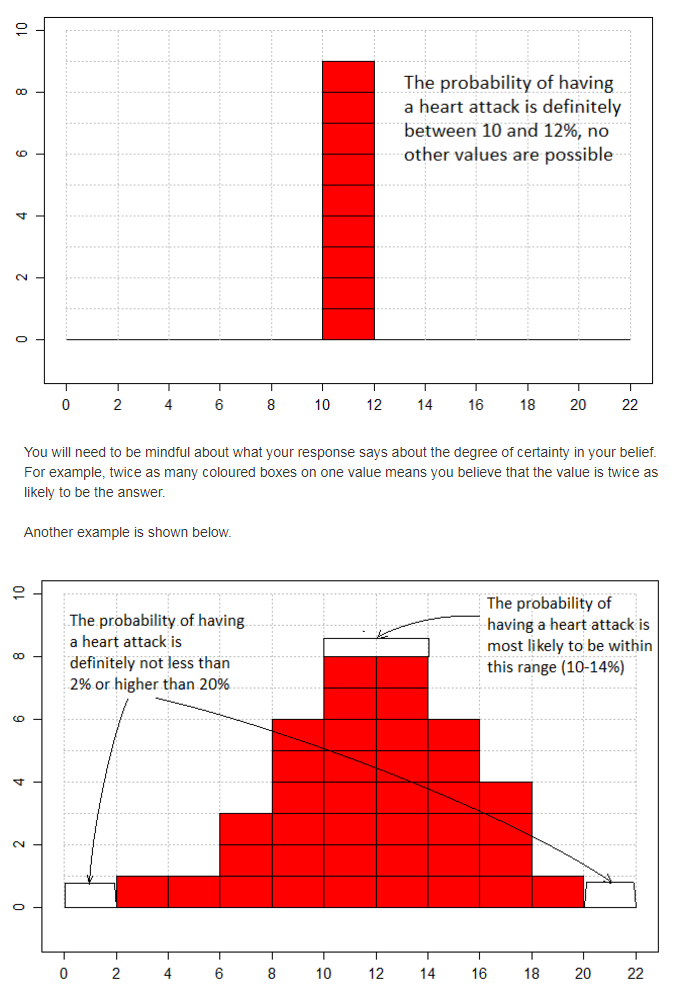
### Training in elicitation technique

Experts will be shown how to use the histogram technique, and were then provided with examples of a uniform distribution, a normal distribution and complete certainty with explanations of what they imply about experts’ beliefs. The experts will be asked to complete a practice example as many times as they required. The experts could not move onto the next section without completing the training. The screen shot of the training material are shown in Figure A1.5.

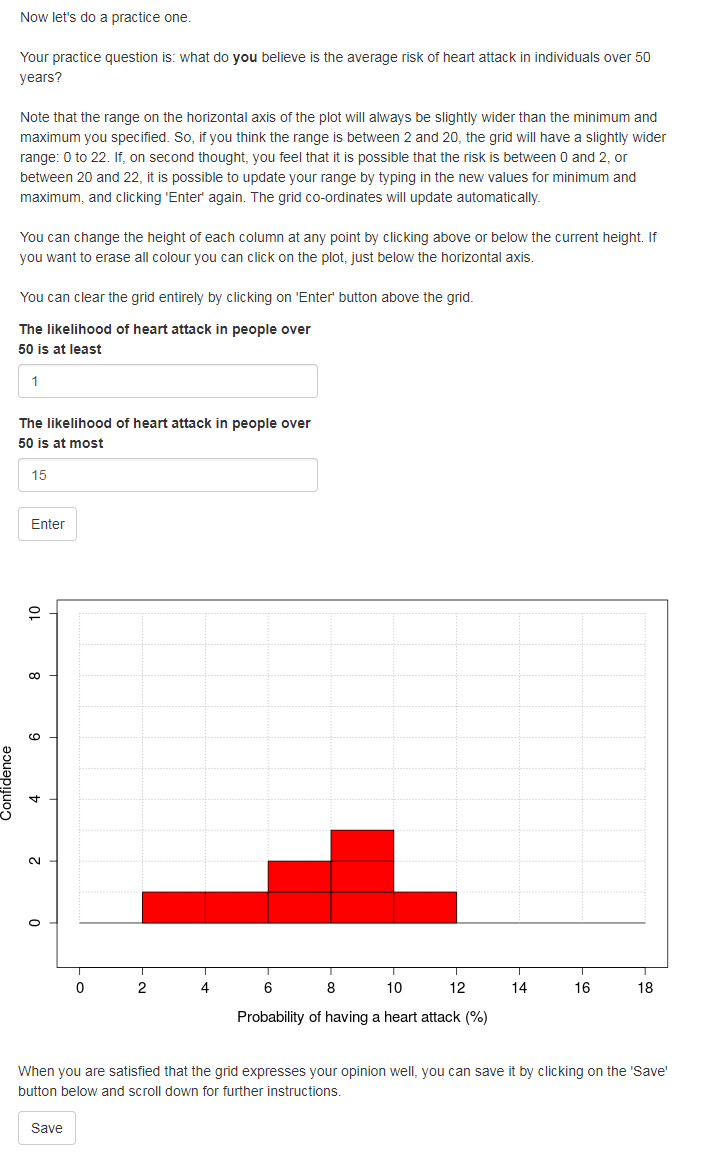
Figure A1.5. Screenshot of the training material.



**Figure A1.5 continued.**



**Figure A1.5 continued.**



## Background Information

This section provides the text that will be provided to experts about the REFORM trial.

### Background

The following two questions are about what you believe would be the outcomes of an intervention designed to reduce the risk of falling by addressing foot and ankle health. We would like you to assume that the intervention has been studied in a randomised controlled trial (RCT) for one year. In this exercise, we are ultimately aiming to capture what you believe would happen if participants continued to receive the intervention after the trial has finished. In order to use your beliefs in our analysis, we first need to put down on paper your beliefs about what is likely to be observed in the trial.

In this section we will provide information about the intervention and the trial, and give instructions on how to proceed with the exercise.

### About the trial

It is a randomised controlled trial (RCT) aiming to evaluate the clinical and cost-effectiveness of a multi-component podiatry intervention for the prevention of falls in patients over the age of 70.

### Why?

There is evidence that foot and ankle problems are associated with increased risk of falls. To read the summary of evidence click on the 'Evidence summary' button below.

|  |
| --- |
| **Evidence summary**  A recent prospective study of 176 older people indicated that ankle flexibility, toe plantarflexor strength and plantar sensation were significant and independent predictors of balance and functional test performance.  A 12-month follow-up of this cohort confirmed that these factors, in addition to foot pain , were significant independent predictors of falls.  A number of studies have assessed footwear in older people who have fallen and suggest that walking barefoot or wearing stockings, increased shoe heel height and smaller sole contact area can all increase the risk of a fall.  A number of other studies have investigated the main features of a shoe thought to affect balance, and found that increased heel height and reduced sole hardness have detrimental effects on balance.  Several studies have suggested that some treatments provided by podiatrists, such as lesion debridement, foot orthoses and foot and ankle exercises, may play a role in improving balance.  Given the emerging evidence that foot problems and inappropriate footwear increase the risk of falls, it has been suggested that podiatry may have a role to play in falls prevention. |

### About the intervention

The intervention aims to address all foot and ankle related risk factors for falling. It consists of four components: provision of foot orthoses, foot and ankle exercise programme, footwear assessment and shoe provision and a fall prevention leaflet.

**- Foot orthoses** are insoles designed to redistribute pressure away from plantar lesions. To read more about the orthoses click on 'Show/hide foot orthoses'.

|  |
| --- |
| Participants will be fitted with a prefabricated insole (Formthotics TM Foot Science) manufactured from a thermoformable cross-linked closed cell polyethylene foam which will be shaped to fit the participant's foot. The orthoses will then be appropriately customised using 3mm thick PPT urethane to redistribute pressure away from any plantar lesions. The orthosis will be supplied either by the podiatrist delivering the intervention or by a manufacturer in response to a prescription from the podiatrist. |

**- The exercise programme** aiming to stretch and strengthen the muscles of the foot and ankle involves 30 minute home based exercise to be undertaken three times per week indefinitely. The exercises will be demonstrated by the podiatrist at the participant's initial visit and will be supplemented by a DVD demonstrating the exercises and an illustrated explanatory booklet showing how to do them at home. To read more about the exercises click on 'Show/hide exercise details'.

**- Footwear assessment** will be carried out. Participants who do not have appropriate footwear will be provided with footwear vouchers and advice on optimal footwear.

**- Falls prevention advice leaflets** will be sent to participants.

### Data collection

They will then complete a questionnaire at 1, 3, 6 and 12 months after randomisation to collect data on compliance with the exercise programme, wearing the foot orthoses and the number of falls they had. They will keep monthly exercise and fall calendars to help them complete the questionnaire.

Follow-up questionnaires will be sent to participants in the post. Participants who provide an email address will be given the opportunity, if they prefer, to complete the questionnaire on-line.

### Who will take part?

Trial participants be

- 70 years of age and over

- Community dwelling

- have had at least one fall in the past 12 months; or one fall in the past 24 months requiring hospital attention.

Participants will be excluded if:

- They are known to have neuropathy.

- They are known to have a neurodegenerative disorder.

- They fail to return all monthly falls diaries over the first three-month period (the pilot) or fail to return the baseline questionnaires.

- They have had a lower limb amputation (including partial foot amputation)

- They are unable to walk household distances (10 metres/32 feet) unaided.

- They are currently wearing a full or 3/4 length in-shoe foot orthotic with the purpose of altering or modifying foot function in order to treat, adjust, and support various biomechanical foot disorders.

- They are known to have dementia.

- They are unable to read or speak English.

- They would be excluded if their usual footwear has been adapted in such a way which would not allow an orthotic to be fitted.

In **question 2** (this entire tab) we will ask about patients in the trial. We will ask a number of questions about those individuals who DO NOT receive the intervention (the control group) followed by a number of questions about those individuals who DO RECEIVE the intervention (the treatment group).

Please click on 'Control group' in the side panel to begin.

You can click back on the 'Instructions' tabs at any point if you wish to remind yourself how to fill in the grid.

## Analysis plan

The analysis plan is provided in detail in the main body of the paper.

## References for Supplementary file 1

1. 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(5):503–12.
2. 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). Quinn TJ, editor. PLoS One. 2017;12(1):e0168712. Available at [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5249075/](about:blank) (accessed 14/12/2020)
3. Sculpher MJ, Claxton K, Drummond M, McCabe C. Whither trial-based economic evaluation for health care decision making? Health Econ. 2006 Jul;15(7):677–87.
4. Bojke L, Soares MFO, Fox A, Jankovic D, Claxton KP, Morton A, et al. Developing a reference protocol for expert elicitation in healthcare decision making. Heal Technol Assess Reports (in Press. 2019).
5. 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 (The Stat). 1998;47(1):3–19.
6. Peterson C, Miller A. Mode, median and mean as optimal strategies. J Exp Psychol. 1964;68(4):363–7.
7. 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 [Internet]. 2011 Jan 16;342:d3411. Available from: [http://www.bmj.com/content/342/bmj.d3411.long](about:blank) (accessed 14/12/2020)
8. 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(1):30.
9. Clemen RT, Fischer GW, Winkler RL. Assessing Dependence: Some Experimental Results. Manage Sci [Internet]. 2000;46(8):1100–15.
10. Gigerenzer G. The Psychology of Good Judgment. Med Decis Mak. 1996;16(3):273–80.
11. O’Hagan A, Buck CE, Daneshkhah A, Eiser JR, Garthwaite PH, Jenkinson DJ, et al. Uncertain Judgements: Eliciting Experts’ Probabilities [Internet]. John Wiley & Sons; 2006.
12. Chang W, Cheng J, Allaire J, Xie Y, McPherson J. shiny: Web Application Framework for R. R package version 1.5.0. [Internet]. 2020. Available from: [https://cran.r-project.org/package=shiny](about:blank) (accessed 14/12/2020)
13. Morris DE, Oakley JE, Crowe JA. A web-based tool for eliciting probability distributions from experts. Environ Model Softw. 2014 Feb;52:1–4.
14. Bojke L, Grigore B, Jankovic D, Peters J, Soares M, Stein K. Informing Reimbursement Decisions Using Cost-Effectiveness Modelling: A Guide to the Process of Generating Elicited Priors to Capture Model Uncertainties. 2017 Jun 15;1–11.

# Supplementary file 2 Piloting the elicitation exercise

## Pilot 1

Pilot 1 was conducted to test different approaches to eliciting the rate of falls: the direct elicitation of the rate of falls, and the indirect method (eventually used in the exercise) described in Supplementary file 1. The participants were two highly quantitative lay participants to minimise the burden of training – they were both postdoctoral Research Fellows in health economics at the University of York. The pilot was delivered using a bespoke Web application comparable to that described in Supplementary file 1.

The summary statistics derived from the two participants are presented in Table A2.1.

Table A2.1. Summary statistics of experts’ priors on the rate of falls elicited directly, and those derived from elicited multinomial distributions.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | Mean | Median | Min | Max |
| Expert 1 | Rate (direct elicitation) | 1.19 | 1.10 | 0.3 | 2.10 |
| Rate (derived from probabilities) | 0.76 | 0.73 | 0.09 | 1.76 |
| Expert 2 | Rate (direct elicitation) | 0.28 | 0.28 | 0.125 | 0.475 |
| Rate (derived from probabilities) | 0.80 | 0.75 | 0.12 | 2.22 |

Directly and indirectly elicited rates were more consistent for Expert 1 than Expert 2.

There was no constancy in effect from using two different methods – for Expert 1 the directly elicited rates were higher than the indirectly elicited rates and their uncertainty was comparable, while for Expert 2 the directly elicited rates were much lower and more certain.

Rates derived using the indirect method appear to be similar for the two experts, whereas directly elicited rates were very different. Furthermore, both experts expressed that they found the indirect method more intuitive and so this was chosen for the exercise.

## Pilot 2

Pilot 2 tested different methods for eliciting uncertainty. The fixed interval method and the histogram technique are the most widely used methods in model-based economic evaluations (MBEE) (1). Both methods were considered for use in this study. The methods were piloted on a lay participant who was a university-educated health professional (pharmacist) with quantitative skills comparable to the experts recruited in the study. While they did not have domain specific expertise, this was not expected to impact the findings from the pilot because the elicited parameter was not field-specific- the participant was asked to express their uncertainty around the number of days it rained in their locality every November. The pilot was conducted remotely using publicly available software MATCH (2); while the investigator guided the participant over the phone.

When testing the fixed interval method the elicited prior followed a U-shaped distribution (the probability density was the highest at the edges or the range) that was judged by the investigator to be implausible. The participant verbally expressed confusion with the method. They found the histogram technique to be more intuitive and the distributions they provided were more plausible (bell-shaped). The histogram technique was thus adopted in the exercise.

## Pilot 3

Pilot 3 involved testing the final version of the exercise to identify any practical challenges with undertaking the exercise. The tool was piloted on seven podiatrists who delivered the intervention in the clinical trial (3). The participants were chosen based on the ease of recruitment (the participants were attending a presentation at the University of York), quick delivery (as they did not require background information on REFORM trial), and the assumption that their quantitative skills were comparable to those of experts in the REFORM elicitation study. The participants were delivered a group training session in a presentation that covered the contents of the ‘Instructions’ tab (described in section 2.5 and Supplementary file 6). The experts used the practice example in the exercise and completed the exercise independently. The pilot was conducted to test the exercise for sense, clarity and ease of use. One technical challenge arose in the pilot: several experts entered the minimum and maximum values in the wrong field (they entered the minimum in the ‘maximum’ field and vice versa), precipitating an error message. As result, the tool was updated to automatically select the smaller value as the minimum and the larger value as the maximum.

**References**

1. Grigore B, Peters J, Hyde C, Stein K. Methods to Elicit Probability Distributions from Experts: A Systematic Review of Reported Practice in Health Technology Assessment. Pharmacoeconomics. 201;31(11):991–1003.

2. Morris DE, Oakley JE, Crowe JA. A web-based tool for eliciting probability distributions from experts. Environ Model Softw. 2014;52:1–4.

3. 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). Quinn TJ, editor. PLoS One. 2017 Jan 20;12(1):e0168712. Available from: <http://dx.plos.org/10.1371/journal.pone.0168712> (accessed 14/12/2020)

# Supplementary file 3. Methods for sensitivity analysis

This file describes the methods used to explore the plausibility of assumptions imposed in the case study using sensitivity analysis.

## Conditional independence between outcomes

The methods for deriving the temporal change in the treatment effect from elicited priors assumed that the 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 assumptions could not be tested directly post-hoc, as they required input from experts to indicate whether they agree with the assumption or not. Instead, correlation coefficients were derived 1) between elicited conditional probabilities to explore whether there was any pattern in the way experts’ beliefs about the probabilities of falling affected their beliefs about the probabilities of falling multiple times; and 2) between baseline outcomes and the treatment effect, to explore whether there was any consistency in the way experts’ beliefs about falling and fractures affected their beliefs about the treatment effect.

## Method for deriving the rate of falls derived from experts’ priors

The method used to derive the rate of falls, assumed that the rate of falls derived from experts’ priors on the probabilities of falling, accurately represented their’ beliefs.

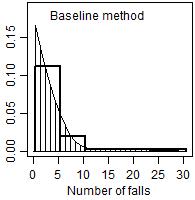
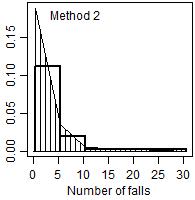
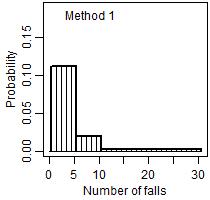
To test whether the derived rates accurately represent experts’ beliefs, first, the probabilities of having 1-5, 6-10 and >11 falls, predicted by the model in Equation A2.1 in Supplementary file 2 were compared to those derived directly from experts’ priors, using Equation 3.1.

|  |  |
| --- | --- |
| *P (a < x ≤ b) = P (x > b) – P (x > a)* | **Equation A3.1** |

The probabilities were compared for five randomly chosen experts from the sample, selected using the random number generator in R.

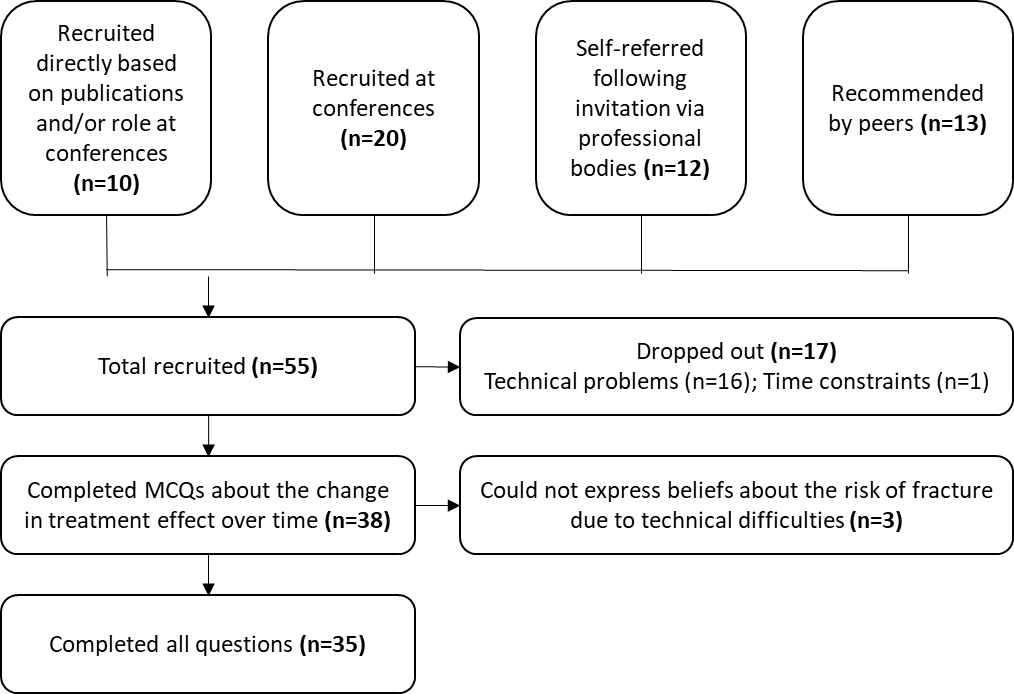
The plausibility of the assumption was further explored by comparing the rates derived using the method described in Supplementary file 2 to those derived using two alternative methods; both methods involved deriving P (0 < x ≤ 5) and P (5 < x ≤ 10) using Equation 3.1. The maximum number of falls was assumed to be 30 (P (x > 30) = 0) as the model for deriving rates predicted that the probability of having more than 30 falls was negligible. The first alternative method assumed that the probabilities of all outcomes within each category (1-5 falls, 6-10 falls and 11-30 falls) were equal, while the second method assumed that the probability of each category (1-5, 6-10 and 11-30 falls) was that stated by the expert, but the probability of each additional fall within that category decreases at a constant rate. Figure A3.1 demonstrates the assumptions made in the two methods graphically, in comparison to that used in the baseline analysis.

**Figure A3.1. Alternative methods for deriving the rate of falls.**



# Supplementary file 4. The expert panel

**Figure A4.1. Recruitment of experts for the structured elicitation exercise**



**Table A4.1. Completion rate of the structured expert elicitation study.**



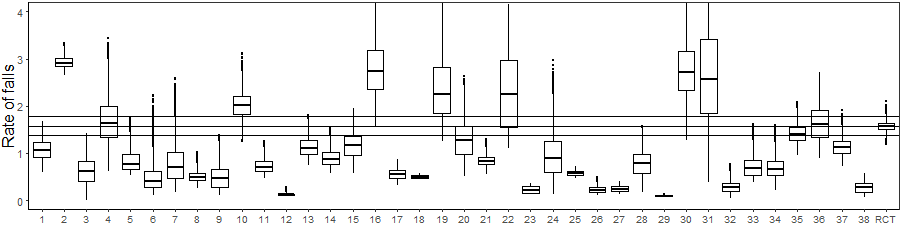
# Supplementary file 5: elicited priors from the sample of experts

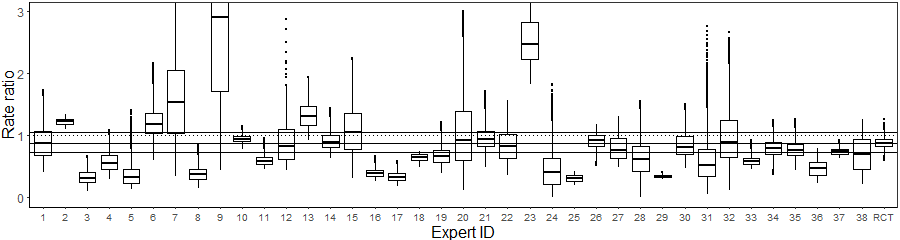
|  |
| --- |
| **Figure A5.1. Experts’ priors on the probabilities of falling in patients who did not receive the intervention and the values observed in the trial (RCT).**    Box = median and interquartile range, whiskers = range. Horizontal lines = mean and 95% confidence interval observed in the trial. |

|  |
| --- |
| **Figure A5.2. Experts’ priors on the relative risk of falling and the treatment effect observed in the trial (RCT).** |

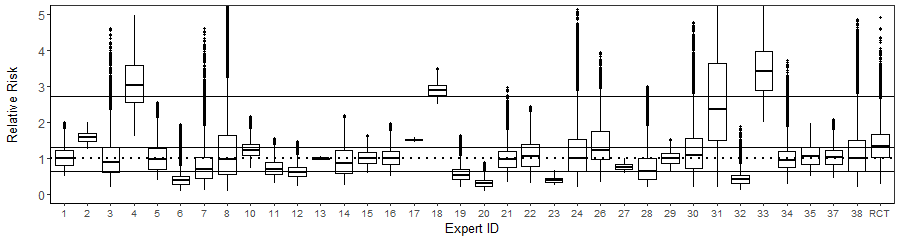
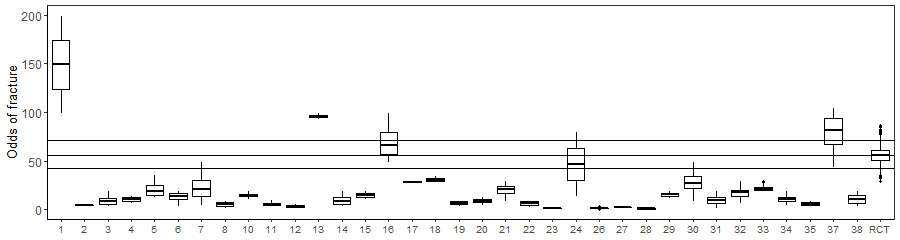
Box = median and interquartile range. Whiskers = range excluding outliers. Points = outliers. Solid horizontal lines = mean and 95% confidence interval observed in the trial.

**Figure A5.3 Experts’ priors on the baseline rate of falls and the rate ratio.**





Box = median and interquartile range, whiskers = range. Horizontal lines = mean and 95% confidence interval observed in the trial. Dotted line = 1 (no treatment effect).

Figure A5.4 Experts’ priors on the baseline odds of fracture, and the treatment effect. 

# Supplementary file 6: results of the sensitivity analysis

### Methods for deriving the rate of falls

The study protocol (Supplementary file 2) explained that the method used to derive the rate of falls required predicting the probabilities of having 1, 2, 3, etc. falls, and that the predicted probabilities of having at least one fall, more than five falls and more than ten falls may deviate from experts’ expressed beliefs. In order to assess whether such discrepancies existed, five experts from the sample were chosen at random, using the random number generator in R. For these five experts the elicited probabilities were compared with those predicted by the regression models detailed in Supplementary file 3. The results are shown in Table A6.1.

Of the fifteen analysed priors (three priors from five experts), only three probabilities differed by more than two percentage points -these are highlighted in the table. The highest difference between predicted and elicited probabilities was 0.07 (or 7 percentage points); Expert 19 believed that the probability of falling between one and five times was 0.217 (mean) whereas the predicted probability was 0.143.

Table A6.1. Elicited and predicted mean probabilities of 1-5, 6-10 and more than 11 falls (range) for five experts chosen at random. Highlighted values differ between elicited and predicted probabilities by more than two percentage points.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | **P(1-5 falls)** | **P(6-10 falls)** | **P(>11 falls)** |
| **Expert 11** | **Elicited** | 0.186  (0.141-0.260) | 0.038  (0.022-0.073) | 0.012  (0.005-0.034) |
| **Predicted** | 0.173  (0.118-0.263) | 0.045  (0.025-0.088) | 0.011  (0.004-0.030) |
| **Expert 15** | **Elicited** | 0.267  (0.201-0.360 | 0.043  (0.022-0.090) | 0.008  (0.002-0.027) |
| **Predicted** | 0.258  (0.171-0.385) | 0.047  (0.022-0.102) | 0.007  (0.002-0.024) |
| **Expert 19** | **Elicited** | 0.217  (0.201-0.250) | 0.010  (0.009-0.011) | 0.007  (0.007-0.008) |
| **Predicted** | 0.143  (0.132-0.166) | 0.026  (0.024-0.031) | 0.004  (0.004-0.005) |
| **Expert 34** | **Elicited** | 0.396  (0.101-0.800) | 0.145  (0.017-0.451) | 0.111  (0.013-0.367) |
| **Predicted** | 0.356  (0.080-0.764) | 0.195  (0.030-0.546) | 0.095  (0.009-0.320) |
| **Expert 36** | **Elicited** | 0.166  (0.101-0.300) | 0.042  (0.021-0.108) | 0.012  (0.004-0.043) |
| **Predicted** | 0.161  (0.090-0.310) | 0.045  (0.022-0.11 ) | 0.011  (0.004-0.040) |

The rate of falls derived using the three different methods are presented in Table A6.2. When the probabilities for each number of falls are assumed to be decreasing, the direct method and predicted probabilities yields very similar results to those obtained using regression – the maximum difference in rates was 0.09 for Expert 34. Assuming an equal probability of all outcomes within a category resulted in higher rates than with the other two methods, as it assigned greater probability to higher number of falls within each category than the other two methods.

**Table A6.2. Rates of falls derived using three different methods.**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Mean rate of falls derived using different methods** | | |
| **Uniform distribution assumes across outcomes of the same category** | **Decreasing probabilities assumed across outcomes of the same category** | **Predicted probabilities of each outcome** |
| **Expert 11** | 0.89  (0.59-1.54) | 0.74  (0.47-1.28) | 0.73  (0.47-1.28) |
| **Expert 15** | 1.11  (0.74-1.85) | 0.90  (0.57-1.56) | 0.90  (0.57-1.55) |
| **Expert 19** | 0.79  (0.63-0.91) | 0.56  (0.52-0.65) | 0.50  (0.46-0.58) |
| **Expert 34** | 3.30  (0.59 -9.12) | 2.73  (0.46-7.7) | 2.82  (0.45-7.76) |
| **Expert 36** | 0.86  (0.46-1.86) | 0.72  (0.39 -1.61) | 0.71  (0.38 -1.57) |

In the baseline analysis, the change in the treatment () effect was assumed to be linear. Sensitivity analysis was conducted to explore the effect of assuming a logarithmic change in the treatment effect, where the change in the treatment effect , and the annual change in the treatment effect were derived using Equation A6.1 and Equation A6.2, respectively.

|  |  |
| --- | --- |
|  | Equation A6.1 |
|  | Equation A6.2 |

Where indicates treatment effect one year after starting treatment, and indicates treatment effect at the second time point.

To get the probability distribution of , 10,000 random samples were drawn from experts’ priors on , and was calculated for each.

### Correlation between conditional probabilities

The correlation coefficients between the proportion of fallers and the conditional probabilities P(x>5|x>0) and P(x>10|x>5) are shown in Table A6.3. The correlation coefficient can take any value between -1 and 1, where positive values indicate positive correlation, and negative values indicate negative correlation. The higher the absolute value the stronger the correlation. The coefficients in Table A6.3 range from -0.11 to 0.27 – these are relatively low values and with varying signs suggesting that the probabilities are not correlated.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table A6.3. Correlation between conditional probabilities of different outcomes (1-5, 6-10 and >10 falls)   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Conditional probabilities** | | | **Correlation coefficient** | | | No intervention (baseline) | P(x>0), P(x>5|x>0) | -0.05 | | | P(x>0), P(x>10|x>5) | -0.05 | | | P(x>5|x>0), P(x>10|x>5) | 0.17 | | | Intervention, one year after starting treatment | P(x>0), P(x>5|x>0) | -0.11 | | | P(x>0), P(x>10|x>5) | 0.1 | | | P(x>5|x>0), P(x>10|x>5) | 0.27 | | | Intervention, at second timepoint | P(x>0), P(x>5|x>0) | -0.07 | | | P(x>0), P(x>10|x>5) | 0.08 | | | P(x>5|x>0), P(x>10|x>5) | 0.12 | | |

### Correlation between baseline outcomes and the treatment effect

The correlation coefficients between conditional probabilities are shown in Table A6.4. The correlation coefficient between outcomes and the treatment effect ranged from -0.35 to -0.06. The coefficient is always negative, indicating the higher the risk of falls and fractures, the lower the treatment effect. However, the coefficient is relatively low, indicating a weak correlation.

**Table A6.4. Correlation between outcomes and the treatment effect on those outcomes.**

|  |  |  |
| --- | --- | --- |
| Treatment effect | | Correlation coefficient |
| Within trial (after 1 year) | RR for P(x>0) | -0.21 |
| RR for P(x>5|x>0) | -0.19 |
| RR for P(x>10|x>5) | -0.32 |
| RtR | -0.10 |
| OR for fractures | -0.16 |
| Treatment effect after the trial | RR for P(x>0) | -0.28 |
| RR for P(x>5|x>0) | -0.32 |
| RR for P(x>10|x>5) | -0.35 |
| RtR | -0.06 |
| OR for fractures | -0.14 |