Expert elicitation of multinomial probabilities for decision analytic modelling: an application to rates of disease progression in undiagnosed and untreated melanoma

Ed Wilson[1], Juliet Usher-Smith[2], Jon Emery[2,3], Pippa Corrie[4], Fiona M Walter[2]

1. Cambridge Centre for Health Services Research, Institute of Public Health, University of Cambridge
2. Primary Care Unit, Department of Public Health & Primary Care, University of Cambridge
3. Department of General Practice, Centre for Cancer Research Faculty of Medicine, Dentistry and Health Science, Victorian Comprehensive Cancer Centre, University of Melbourne.
4. Cambridge Cancer Centre, Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge, CB2 0RE, UK

ed.wilson@medschl.cam.ac.uk

Abstract: 243/250

Body: 3995/4000

Tables+Figures: 6/6

Acknowledgements

The authors would like to thank the experts in melanoma for taking part in this study and providing their most valuable feedback. We would also like to thank members of the Melatools steering committee for their help and comments in development of this work (www.melatools.org/team.html), James Brimicombe, IT manager, Primary Care Unit, University of Cambridge for assistance and advice with technical aspects of the project and Becky Lantaff, Research Assistant, Primary Care Unit, University of Cambridge, for coordinating webinars and participants.

This work is an unfunded extension to the Melatools programme, which provided participant honoraria. The Melatools programme is funded by a National Institute for Health Research (NIHR) Clinician Scientist Award (RG 68235). EW is funded by the NIHR Cambridge Biomedical Research Centre. JUS was funded by an NIHR Clinical Lectureship. JDE is funded by an NHMRC Practitioner Fellowship.

Views expressed in this publication are those of the authors and not necessarily those of the NHS, NIHR or Department of Health.

Highlights

What is already known about the topic?

- Decision making will almost always require expert opinion to fill gaps in the evidence due to an absence of, or impossibility of conducting, relevant prospective studies.
- A structured expert elicitation process quantifying experts' uncertainty in belief is axiomatically preferable to less formal approaches such as uninformed point estimates with arbitrarily wide credibility intervals to represent parameter uncertainty.
- Previous decision models have used an arbitrary 10% annual probability (approximately 5.1% per six months) of progression from one melanoma disease stage to another, with wide credibility intervals.

What does this paper add?

- We successfully elicited and pooled the beliefs of a number of UK, Australia and New Zealand experts in melanoma regarding the probability of disease progression from 12 discrete melanoma disease stages. The pooled distribution has been included in a decision analytic model estimating the cost-effectiveness of melanoma screening.
- The transition probabilities differ substantially from the previous arbitrary estimates and their credibility intervals represent a more appropriate characterisation of uncertainty.

Abstract

Background: Expert elicitation is required to inform decision making where relevant 'better quality' data either do not exist or cannot be collected. An example of this is to inform decisions as to whether to screen for melanoma. A key input is the counterfactual, in this case the natural history of melanoma in patients who are undiagnosed and hence untreated.

Objectives: To elicit expert opinion on the probability of disease progression in patients with melanoma that is undetected and hence untreated.

Methods: Bespoke webinar-based expert elicitation protocol administered to 14 participants in the UK, Australia and New Zealand, comprising 12 multinomial questions on the probability of progression from one disease stage to another in the absence of treatment. A modified Connor-Mosimann distribution was fitted to individual responses to each question. Individual responses were pooled using a Monte Carlo simulation approach. Participants were asked to provide feedback on the process.

Results: A pooled modified Connor-Mosimann distribution was successfully derived from participants' responses. Feedback from participants was generally positive with 86% willing to take part in such an exercise again. However, only 57% of participants felt this was a valid approach to determine the risk of disease progression. Qualitative feedback reflected some understanding of the need to rely on expert elicitation in the absence of 'hard' data.

Conclusion: We successfully elicited and pooled the beliefs of experts in melanoma regarding the probability of disease progression in a format suitable for inclusion in a decision analytic model.

Keywords: expert elicitation, decision modelling, melanoma

Introduction

Evidence based medicine is defined as the "conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients… [which] means integrating individual clinical expertise with the best available external clinical evidence from systematic research".[1] These principles apply equally to population level decision making, such as whether a healthcare payer should provide reimbursement for a new drug, treatment pathway or screening programme.

Decision analytic models are frequently used by agencies such as the National Institute for Health and Care Excellence (NICE) in England as a framework to structure all current best (i.e. relevant, quality-assessed) evidence to estimate the overall costs and consequences of alternative treatment strategies over an appropriate time horizon.[2 3] A judgment is then made to decide whether the added benefit of a treatment exceeds its opportunity cost. Evidence to populate a model is ideally obtained exclusively from good quality systematic reviews and meta-analyses of RCTs or other relevant study designs as appropriate. However, due to data limitations, evidence is typically obtained from various sources including routine databases and observational studies. Where no suitable prior data exist decision makers are required to rely on expert opinion to bridge the evidence gaps.

Such an evidence gap is the natural history of undetected and hence untreated melanoma.

The Melatools programme (www.melatools.org) is an NIHR-funded programme based in the UK to improve the early diagnosis of melanoma to reduce associated mortality and morbidity. This includes investigation of the feasibility and cost-effectiveness of introducing a risk-based surveillance programme using a self-completed assessment tool.[4]

To address this we developed a decision model to estimate the most cost-effective cut-offs for various intervention policies.[5] However, a key component of this is the counterfactual, in other words, the natural history of untreated melanoma in the absence of medical intervention. Good quality data exist on prognosis post diagnosis and subsequent treatment,[6] but there are no data on untreated individuals. Obtaining such data from a prospective study by withholding treatment from newly diagnosed patients would clearly be deeply unethical. Therefore, the only way to estimate the probability of an undiagnosed and hence untreated patient progressing from one disease stage to another is to garner expert opinion.

In this paper, we apply a method to elicit multinomial probabilities from experts regarding their beliefs about the rate of progression from different melanoma disease stages (in situ disease to stage IV) to any other stage or death. The primary purpose of the analysis was to use the resulting multinomial distributions in our decision model to predict the cost-effectiveness of a self-completed risk assessment tool and subsequent surveillance programme. However, the distributions themselves are of interest as they represent a summary of expert opinion and belief.

Method

Research Problem

There are four main types of cutaneous melanoma (superficial spreading, lentigo maligna, acral lentiginous and nodular),[7] which current guidelines categorise into nine stages of invasion.[6] All but one are also described with a pre-invasive (*in situ*) phase; nodular melanoma is by definition invasive. We wished to elicit expert opinion on the rate of progression from each stage to any other. We simplified a possible set of 39 questions into 12 by assuming that invasive disease would progress at the same rate irrespective of primary melanoma subtype; we allowed the rate of progression from *in situ* disease to vary by subtype (Table 1). Each question is a multinomial problem: the quantities to be elicited are probabilities, but there are more than two outcomes. For example, after a defined time period, a patient with Stage 1a disease may remain in Stage 1a, or progress to 1b or 2a and so forth. The sum of the probabilities must equal one.

Elicitation Protocol

The protocol and associated materials are in Appendix 1. The protocol was designed with the following constraints in mind:

- We wished to elicit opinion from experts of more than one country. We chose the UK, and Australia and New Zealand (hereafter ANZ) as areas of relatively high melanoma prevalence. Arranging a single workshop event in the same place at the same time would be prohibitively expensive and extremely difficult to schedule. Therefore, an online webinar approach that could be repeated to suit availability of participants was desired.
- Due to demands on experts' time, the webinar could not exceed two hours in length.

Ethics

Ethical approval was not required for this study.[8] Invitation letters explained to participants that their responses would be anonymised, with the only details being their broad job title and country (UK or ANZ).

Identification and recruitment of experts

Inclusion criteria were that participants had to be located in the UK, Australia or New Zealand with an academic or clinical background in either dermatology, oncology, plastic surgery, or epidemiology, with a particular interest and expertise in melanoma. A list of potential participants was identified by two of the investigators (FW & JE), based on known expertise and relevant publications in the field. Participants were invited to take

part via email, and several sessions were scheduled to allow flexibility to maximise recruitment. Participants were paid an honorarium of £200 / Aus$400 for their time.

Background materials

We circulated background materials to participants prior to the webinars, including confirmation of date and time, an explanation of the overall purpose of the exercise, a user guide explaining how responses would be recorded (on a specifically designed Microsoft Excel spreadsheet) and relevant background literature. The only relevant literature identified was the current AJCC staging recommendations for melanoma, which includes survival curves by disease stage at diagnosis.[6]

Pre-elicitation training

Each webinar began with a 30-minute presentation introducing the concept of elicitation and example questions, followed by a live demonstration of how to use the Excel spreadsheet.

Elicitation method

Questions were asked in the format of "Imagine a cohort of 100 patients with Stage X undiagnosed and hence untreated disease. After 6 months, the patients may be in any of the following stages:". At this point participants could select from a drop-down list any stage they think it is possible for patients of the cohort to be in. They then ranked these in order of likelihood, from most likely to least likely (screenshot in Figure 1a). Once participants were happy with their selections they clicked 'Update chart' which populated a chart with the selected stages, ordered from most to least likely (Figure 1b), with a default equal spread of probabilities. Edits to the chart could be made by selecting cells in the table below the chart and either clicking increase/decrease or by simply entering an appropriate number. The chart updated instantly providing visual feedback to the participant.

Participants were asked to first adjust the medians according to their beliefs, working from least to most likely (right to left of the chart, or bottom to top of the table, in the example shown starting from Stage 3B, then 3A etc). A restriction was placed that the medians had to sum to 100; the software would not permit participants to finalise their answer unless this was satisfied. Once the participant was happy that the selected medians did indeed represent their median beliefs, the lower and upper 95% credibility bounds were set for each stage, again working from right to left of the chart. There was no restriction on the sum of the lower and upper bounds: participants could set them such that the interval contained within them represented the strength of their belief

about plausible values. As before the chart updated immediately allowing the participant to visualise his/her responses. Once happy, the participant clicked 'submit'. The responses were stored and timestamped and the spreadsheet automatically moved on to the next question.

The webinar was designed such that it could be repeated with relative ease to accommodate availability of participants. The facilitator (EW) would remain online whilst they completed the exercise to attend to any problems. This also served to ensure participants completed the exercise at the allocated time and not rushed in their own time. The timestamp on responses allowed monitoring of the length of time spent on each question.

Fitting distributions to elicited data

A modified Connor-Mosimann distribution[9-11] (a generalisation of the Dirichlet distribution) was fitted to the elicited median and 95% credibility intervals for all dimensions for each participant. These summary distributions were then sampled from many times (Monte Carlo simulation). The empirical median and upper and lower 95% credibility bounds (ie. 2.5th, 50th and 97.5th centiles) were then calculated from the samples and a modified Conor-Mosimann distribution fitted to these overall figures, representing an aggregate of the individual participants' beliefs. This was conducted at a national level (UK and ANZ), and for all participants together. Distributions were fitted using R.[12] A copy of the code is available on request from the corresponding author.

Piloting

The protocol was piloted amongst Melatools steering committee members on three separate occasions, resulting in several modifications to the initial plans. These related to the (1) specification and ordering of questions, (2) length of the webinar, and (3) mode of elicitation.

Specification and ordering of questions

As described above, to avoid respondent fatigue, progression of melanoma was simplified into 12 stages (and a death state): 3 *in situ* subtypes, and progressive from Stage 1a to 4 (as per American Joint Committee on Cancer [AJCC] definitions[6]). The training session included an explanation to participants that if they felt there was a substantial difference in rate of progression from one stage to another by sub-type (e.g. Stage IIa superficial spreading vs Stage IIa lentigo maligna), then they should consider a weighted average risk based on their experience of case-mix. Finally, to further minimise the impact of respondent fatigue, questions would be asked in random order.

Two further issues were the time horizon over which participants were asked to estimate changes and the number of patients in the cohort. This was originally proposed to be 1000 patients over one month. However, it was considered that insufficient patients would have progressed over this time, and that asking participants to allocate 1000 patients may lead to spurious precision. Therefore, the time horizon was set at six months and participants were asked to allocate a cohort of 100 patients.

Length of the webinar

The webinar was originally proposed at three hours. However, concerns were raised that it would be difficult to recruit participants for a three-hour session, and so the timing was reduced to two. This proved to be a reasonable estimate in both piloting sessions.

Mode of elicitation

Two main modes of elicitation are the quantile and roulette modes.[13] In the roulette mode, participants place 'chips in bins' representing the relative strength of their belief about different values for a parameter. Whilst commonly used to elicit binomial probabilities and continuous quantities, we considered this less suitable for eliciting multinomial probabilities. Therefore, we opted for a quantile approach, where a minimum of 3 points along the distribution are elicited, typically the median and tertiles. The tertile method was originally proposed: the median expressed as the value X at which the participant would place a 50:50 bet on the 'true value' being greater or less than X, and the lower and upper tertiles being the values at which a participant would place a 2:1 bet.

This was rejected by the steering group on the grounds that clinicians would not be familiar with tertiles and objections to the comparison with gambling to explain the method. Despite the same rules of probability governing both clinical outcomes and games of chance, explanations in terms of betting odds were removed and tertiles (33% credibility intervals) rejected in favour of 95% credibility intervals, defined in terms of "it is possible for X to be greater than this value, but I would be extremely surprised if this were to be the case". The numeric quantity (1 in 40 or 2.5% probability) was also stated as the likelihood associated with this situation.

Format of Results

This manuscript is written to conform as closely as possible with recommendations for reporting of expert judgement,[14] although this work was designed and conducted prior to publication of these guidelines. We

report details of the participants, present tables and figures of aggregate distributions, and analysis of the feedback forms from participants, including time to complete.

Results

Participants

Sixteen participants agreed to take part in the exercise from a pool of 39 invited experts. Of those, 13 successfully completed the entire exercise, one completed questions on invasive disease only, and two did not complete any questions/withdrew their participation (Figure 2). Elicitation from UK participants took place over four webinars scheduled between November 2015 and January 2016. Elicitation from ANZ participants also took place over four webinars between January and February 2016.

Elicited probabilities

The results comprise a modified Connor-Mosimann (mCM) distribution for each question for each participant, plus a summary distribution representing the aggregation of all responses to each question (all UK, all ANZ and both combined). Density plots proved somewhat unclear when visualising these data. Therefore, we present box and whisker plots showing medians, interquartile and 100% ranges for the fitted distributions for UK and ANZ participants, and all combined (Figure 3). Parameters of the respective mCM distributions and resulting medians and 95% credibility intervals (CrIs) for UK, ANZ and all combined are in Appendix 2.

The broad rankings of disease progression are consistent between UK and ANZ, but there is some variation: in general, there was least disagreement in point-estimates (i.e. medians) for *in situ* and stage 1 disease. There was most disagreement in medians for stages 2B, 2C and 4. A crude mean of the credibility intervals for each stage suggests there is greatest overall uncertainty in progression from stages 3B, 3C and 4, and the least from *in situ* and stage 1B disease, although the 95% Credibility Interval varied between almost zero and one for the risk of remaining in an *in situ* stage. These extremes are driven by the ANZ participants, with tighter 95%CrIs in the UK. Finally, the results suggested a greater consensus in some of the transitions: regression of disease to an earlier stage was not generally considered possible (one of the ANZ participants believed there was a very small possibility for this), and progression straight to more extreme stages from *in situ* or early stage was considered less likely, with appropriately low medians and narrow 95% CrIs.

Time to completion and analysis of feedback forms

Time stamps reporting a time per question greater than 30 minutes were excluded (this was common for the first question asked where participants had opened the spreadsheet early. Details of exclusions are in Appendix 3). After excluding these, the mean time to completion per question was 5m09sec (SD 3m16sec), with a total

time of 59m20sec (SD 21m19sec). There was a downward trend with question number suggestive of either a learning effect and/or respondent fatigue (Figure 4).

A summary of the feedback is in Table 2. Four participants (29%) had heard of the concept of expert elicitation prior to this study with involvement in Delphi panels. One participant had not taken part in one before but was familiar with the hierarchy of evidence and that sometimes expert opinion is the only source of relevant data. The mean (median) response to ease of understanding was 2.14 (2.00) on a 6-point scale, where 1=Easy and 6=Difficult, with mean self-reported time to completion of 73.2 minutes (median 60). This is slightly longer than the measured mean of 59m20sec. However, the measured mean excludes outliers >30 minutes in length.

Mean confidence in responses was 3.93 (median 4.00) on a six-point scale (1=Not at all confident, 6= very confident). Free text explanations stated that whilst the answers may reflect a participant's belief, the participant had concerns in the limitations of their beliefs due to lack of 'hard' evidence. Other comments focused around the complexity of both melanoma as a disease, and of the staging recommendations. The background of the participants was also mentioned, with dermatologists likely to be more confident in early stage progression and oncologists more familiar with later stage disease. Due to small numbers we were unable to identify whether this was reflected in our data. Finally, one respondent commented that the 6-month time horizon was too short and would have been more confident making a 1 or 2 year prognosis.

Only 57% (8/14) participants felt this was a valid approach to answering the study question. Some free text responses acknowledged that in the absence of 'better' evidence, expert elicitation was the only option. Other participants suggested a cluster RCT of screening versus no-screening, or that the approach may be valid given a 'large enough' sample size (suggesting 100 respondents).

All but two participants indicated they would be willing to take part in such an exercise again, although the participants are a self-selected group: a further two participants who initially agreed to take part declined to provide answers and did not provide feedback forms. Most free-text comments suggested participants found it interesting with a desire to see the final results of the study.

Discussion

Summary of results

We elicited parametric distributions representing 12 unknown multinomial probabilities describing experts' beliefs about the rate of progression of an individual with untreated melanoma from one stage to another over a 6-month time horizon. The resulting distributions are in a format suitable for incorporation in a decision model. The exercise revealed where there was varying confidence both within and between individuals in the rate of progression. For example, the probability of progressing from 1A to 2C, 3A or 3B had medians of 1% to 3%, with 'tight' 95%CrIs of 0% to 17%. Other areas were highly uncertain: the 95%CrI of patient with *in situ* acral lentiginous melanoma remaining in that state ranges between 0% and 99%. However, this is far from a uniform distribution (representing complete ignorance) as the median is 81%.

Comparison with other studies

Probably the most well-known structured elicitation technique is the SHELF tool (the SHeffield Elicitation Framework).[13 15] This is a consensus-based approach requiring participants to agree on a final summary distribution representing their belief about plausible values for a single parameter. Recently this has been extended to elicit multinomial parameters using a Dirichlet distribution.[16] Our analysis here extends this by fitting elicited data to a Connor-Mosimann distribution[10 11] modified to allow greater flexibility, thus providing a much better fit to the data.

The SHELF approach, whilst considered best practice, suffers from several practical limitations. Firstly, the consensus approach requires a face-to-face workshop bringing together all relevant participants into the same room at the same time. As well as being somewhat logistically challenging, this approach limits the number of participants to 6-8 at most to facilitate conversation, and the workshop requires an experienced facilitator to ensure even representation of all views. Finally, the consensus approach also limits the number of questions that can realistically be asked in one session to four or five at most.

Other approaches have used computer-based methods. For example, a study eliciting the opinion of nurses on the effectiveness of different bandages for severe pressure ulcers used a bespoke spreadsheet in Microsoft Excel/VBA.[17] The nurses completed the task together in a computer suite, ensuring sufficient attention was paid to answering the questions as well as providing technical support in case of difficulties.

Previous decision modelling studies requiring an estimate of the risk of progression in undiagnosed and untreated melanoma[18 19] had used notional 10% annual probabilities with a 'wide' 95%CrI of 0.0001% to 54.87% (a Beta(0.3,2.7) distribution). This represents a mean of approximately 5.1% per 6-months. The transition probabilities presented in this analysis are very different from these previous estimates. It would be of value reiterating those previous models with these new parameter estimates to explore the impact on their conclusions.

Practical issues associated with the elicitation protocol

Piloting of the protocol was an extremely important component in this project, leading to several changes in approaches and indeed delayed the entire project by several months to ensure the internet-based workshop was as valuable as possible. However, the rejection of tertiles in favour of 95% credible intervals may have resulted in loss of precision: eliciting tertiles requires the participant to consider relative odds explicitly whereas eliciting 95% credibility intervals requires the participant to consider 'almost certainty', which is somewhat vague. We also relied on verbal confirmation that experts had fully understood the nature of the task, were willing to honestly report their subjective uncertainty, and in particular understood the difference between means and medians. Furthermore, in retrospect it may have been preferable to request estimates of prognosis over one- or two-year time horizon, not six months.

Hosting the sessions as online webinars most likely increased the overall response rate, allowing scheduling to fit around the diaries of the participants. However, we were only able to elicit responses from 14 participants (7 UK and 7 Aus/NZ) across eight facilitated sessions, although this may be a function of relative rarity of expertise in the area. The numerous repeats of the webinar also required quite a number of input hours from the facilitator (EW), and a risk of inconsistency between webinars leading to systematically different results. The effort could have been eased and consistency issues partially addressed by pre-recording the presentation component with subsequent opportunity for questions, although this reduces the interactivity of the session.

We had originally intended to use a bespoke web-based tool for participants to record their answers. This would be platform independent and automatically upload answers to a study database. However, coding such a platform proved troublesome and we found it expedient to use a macro-enabled spreadsheet written in Microsoft Excel. This led to its own problems, with participants requiring appropriate security settings on their

computers to allow macros to run. In particular, the macros would only run on Windows PCs, not Macs. Participants with Macs therefore had to source Windows PCs in order to take part in the study.

There was also a lack of clarity and apparent inconsistency between participants regarding the 'dead' state. The intention of the facilitator was that this would be specifically melanoma-mortality as the decision model into which the results will be entered already includes background mortality. However, this was not made explicit to the participants. Thus allowing a non-zero probability of death from an earlier stage implies that over that six months a participant believes that the patient could progress through all stages of the disease and die. This is unlikely from Stage 1A, therefore the two participants that allowed non-zero values for death here may have been considering background mortality. Making this explicit is a lesson for future studies of this nature.

Justification for seeking expert opinion to inform parameter estimates

The response of participants to the use of expert elicitation in decision making was very varied, ranging from complete acceptance to extreme scepticism, with only 57% considering this a valid approach. It is therefore important to consider the alternative: decisions to adopt or reject new technologies must be made irrespective of the evidence available at the time. A decision to remain with the status quo pending further evidence is still a decision not to adopt and so risks an opportunity loss. The purpose of a decision model is to assemble all evidence there is, critically appraise it and structure it in a way to assist with a decision. For example, short term effectiveness from an RCT combined with epidemiological data on long term progression. There will always be gaps in this evidence, or areas where the 'best available' data may fall short of the 'best conceivable'. Decision modelling highlights this, and expert elicitation provides one means of plugging it where better quality evidence either does not yet exist or cannot exist/be collected (eg due to physical impossibility or ethical concerns as in the melanoma example presented here). The alternative to this is subjective, informal discussion of decision makers. A structured consultation process with relevant 'experts' focused on carefully eliciting their epistemic uncertainty may be considered superior *a priori*. It should be noted that where data can be, and are, subsequently collected the model should be updated accordingly. Where those data cannot be collected, the best that can be achieved is to present the results of the elicitation process fully and transparently, allowing readers to decide whether they feel the values are plausible. If so, then all else being equal, the results generated from a decision model using those values must also be plausible.

Conclusion

We successfully developed an online structured process that succeeded in eliciting and fitting a multinomial distribution, representing an aggregation of experts' beliefs about the risk of progression of untreated melanoma from one stage to another. The parameters of the overall distribution have been inserted into a decision model to estimate the cost-effectiveness of various screening and monitoring strategies to identify those at high risk of melanoma in a UK setting.[5] Critically, the uncertainty in belief of experts about the rates of progression has been captured and when combined with uncertainty in other parameters, translated into decision uncertainty as to which strategies are likely to be the most cost-effective.

Table 1: Study questions

In situ superficial spreading melanoma
In situ lentigo maligna melanoma
In situ acral lentiginous melanoma
Stage Ia
Stage Ib
Stage IIa
Stage IIb
Stage IIc
Stage IIIa
Stage IIIb
Stage IIIc
Stage 4

Participants were asked for their beliefs about the probability of progression from each of the 12 stages stated to any other stage and death.

Table 2: Summary of quantitative feedback from participants

	Mean (SD)	
	Mean	Median
Had you heard of the concept of expert elicitation prior to this study?	29% Yes	
How easy or difficult did you find the concepts to understand? (1=Easy, 6=difficult)	2.14	2.00
How long did it take you to complete the questions (mins)?	73.21	60.00
How confident are you that your answers reflect your belief about the risk of progression from one stage to another? (1=not at all confident, 6=very confident)	3.93	4.00
Do you think this is a valid approach to determining the risk of progression from one stage to another?	57% Yes	
Would you take part in one of these exercises again?	86% Yes	

Figures 1a (left), b (mid) & c (right): Screenshots of Spreadsheet tool

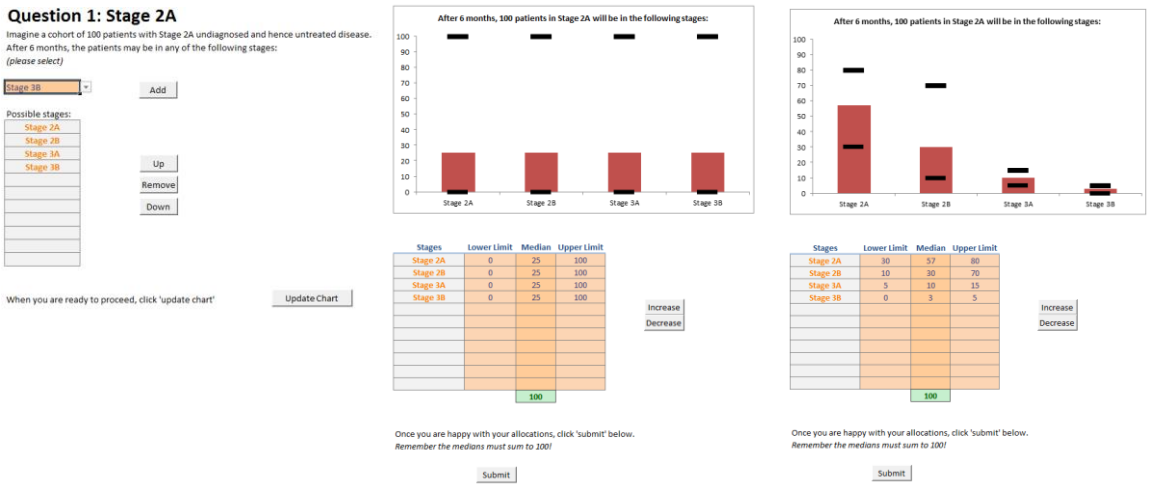
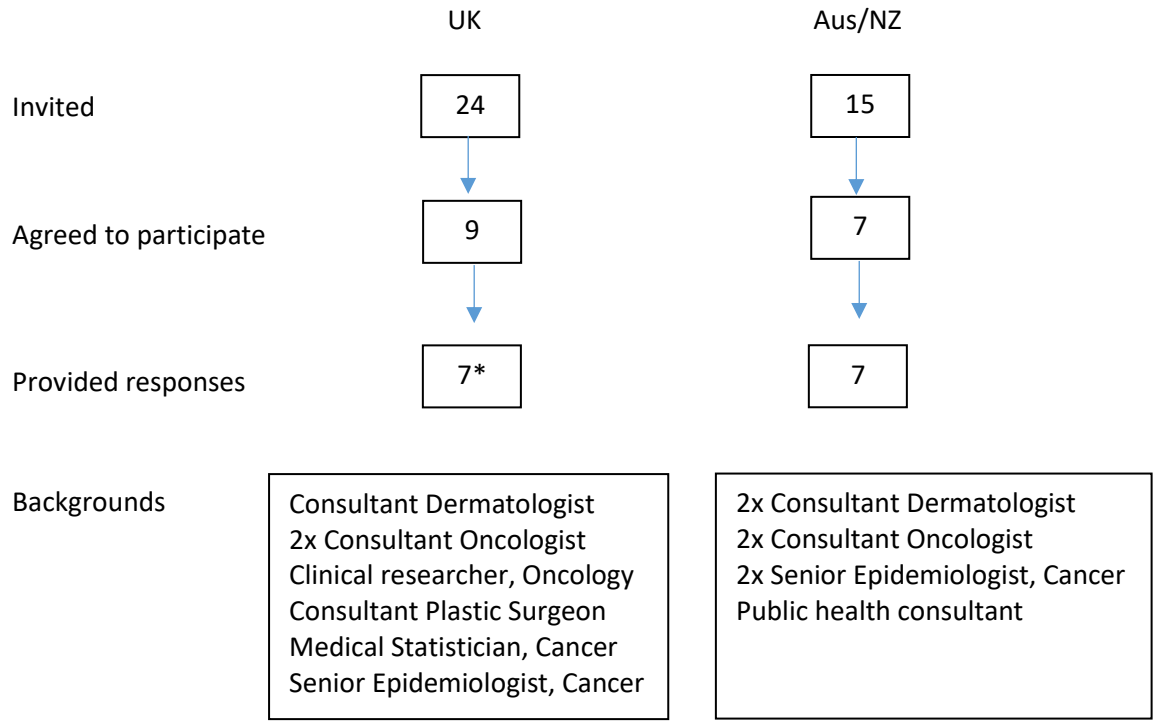
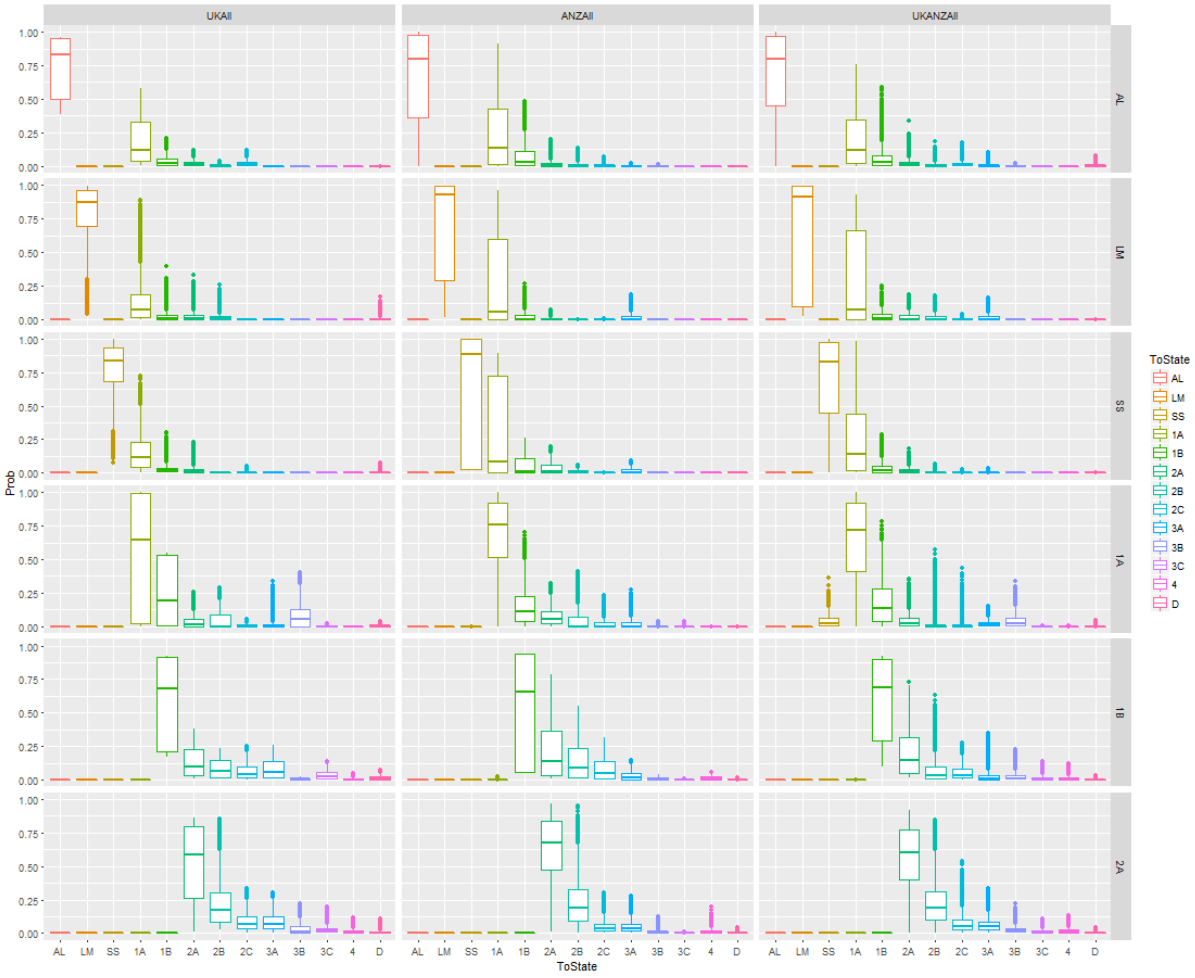


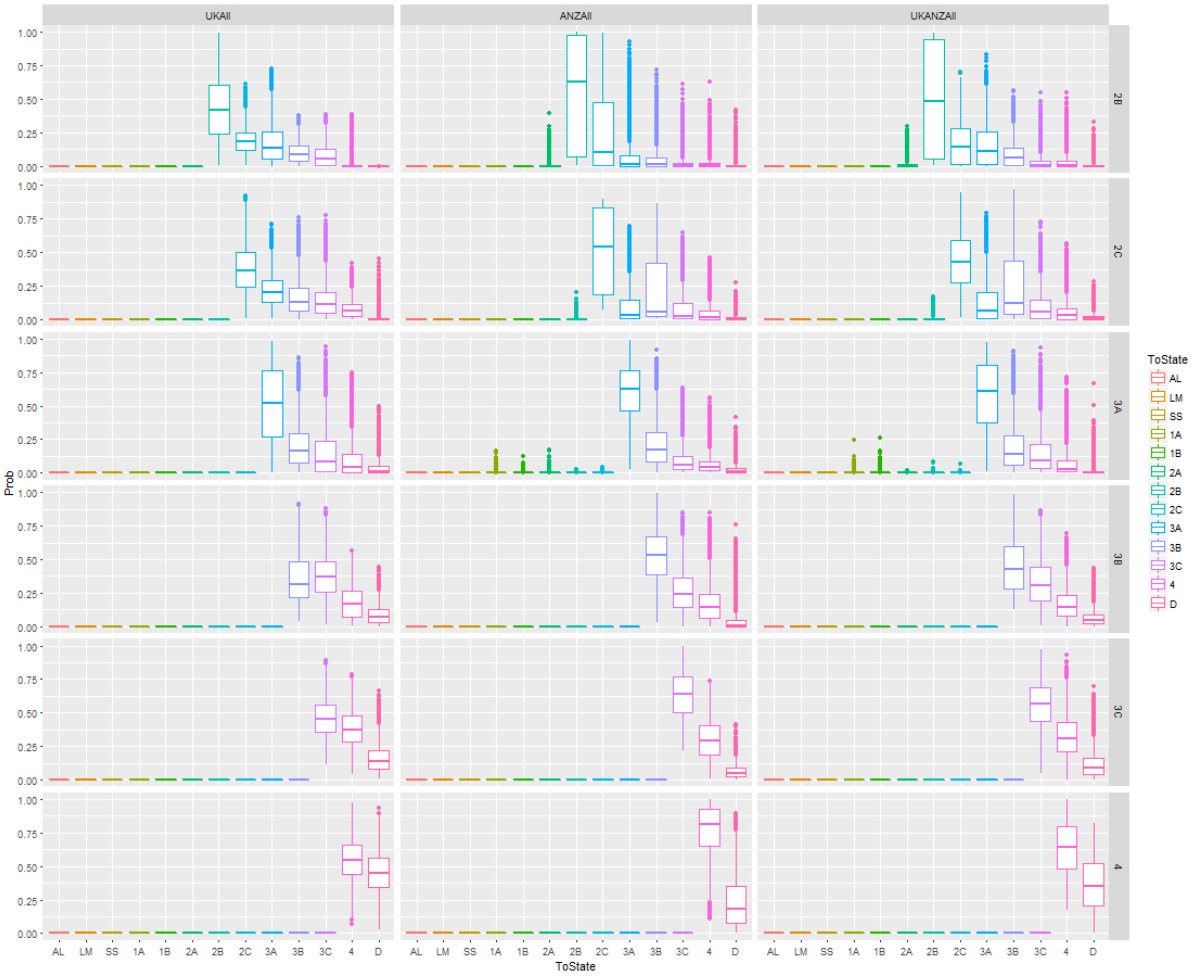
Figure 2: Invitations and participants



**6 provided full answers, 1 provided answers to questions on invasive disease stages only.*

Figures 3a&b: Summary results of UK, ANZ and both combined

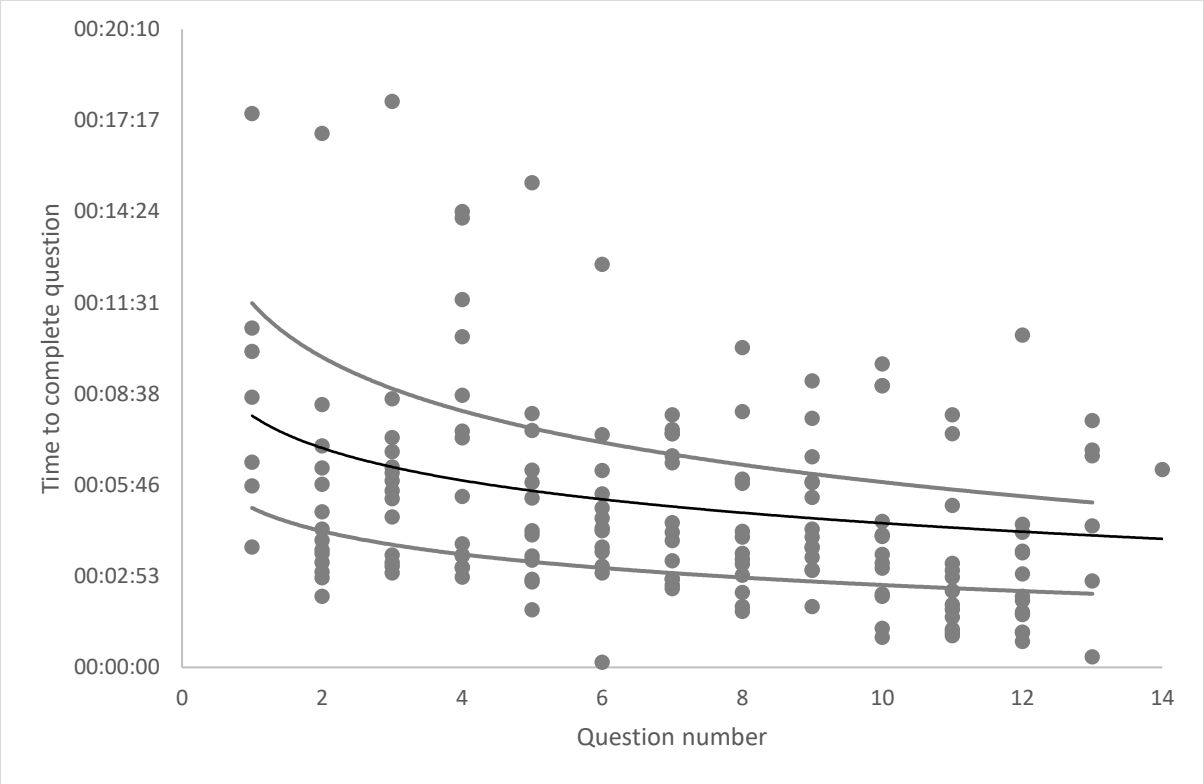




Fitted modified Connor-Mosimann distributions for combined results of UK and ANZ participants and all combined. Each row represents the starting stage. Thus, the top left cell of Figure A2.1 shows the summary of all UK participants' beliefs about the probabilities of a patient with in situ acral lentiginous melanoma (AL) transitioning to any other state after 6 months. In this case, the median probability of remaining in the in situ AL stage is 82%, 6% probability of transitioning to stage 1A, 5% to 1B, 2% to 2A and 1% to stage 2B or death. The IQRs (boxes) and ranges (whiskers) show the overall uncertainty in belief amongst the participants.

AL: in situ acral lentiginous melanoma; LM: in situ lentigo maligna melanoma; SS: in situ superficial spreading melanoma; 1A – 4: invasive melanoma of stage 1A to 4 respectively; D: dead.

Figure 4: Time taken to respond to questions



Points represent each participant's response to each question. Participants reporting more than 12 questions include repeated answers to previous questions. Lines show fitted mean and associated 95% confidence interval. Note question number is the chronological order rather than a specific question: the order of questions was randomised.

References

1. Sackett DL, Rosenberg WM, Gray JA, et al. Evidence based medicine: what it is and what it isn't. BMJ 1996;**312**(7023):71-2
2. Drummond M, Sculpher M, Claxton K, et al. *Methods for the Economic Evaluation of Health Care Programmes*. 4th ed. Oxford: Oxford University Press, 2015.
3. Briggs AH, Sculpher M, Claxton K. *Decision Modelling for Health Economic Evaluation*. Oxford: Oxford University Press, 2006.
4. Williams LH, Shors AR, Barlow WE, et al. Identifying Persons at Highest Risk of Melanoma Using Self-Assessed Risk Factors. J Clin Exp Dermatol Res 2011;**2**(6) doi: 10.4172/2155-9554.1000129.
5. Wilson E, Usher-Smith J, Emery J, et al. A modelling study of the cost-effectiveness of a risk stratified screening programme for melanoma in the UK. [Submitted for publication]
6. Balch CM, Gershenwald JE, Soong SJ, et al. Final version of 2009 AJCC melanoma staging and classification. J Clin Oncol 2009;**27**(36):6199-206 doi: JCO.2009.23.4799.
7. Liu V, Mihm MC. Pathology of malignant melanoma. Surg Clin North Am 2003;**83**(1):31-60, v doi: 10.1016/S0039-6109(03)00003-3.
8. Medical Research Council, NHS Health Research Authority. Do I need NHS REC approval? Decision tool. Last Accessed 21/11/16. http://www.hra-decisiontools.org.uk/ethics/.
9. Wilson E. A method to fit a flexible multinomial distribution to quantities elicited from experts for decision analytic modelling. [submitted for publication]
10. Connor RJ, Mosimann JE. Concepts of Independence for Proportions with a Generalization of the Dirichlet Distribution. Journal of the American Statistical Association 1969;**64**(325):194-206 doi: 10.2307/2283728.
11. Elfadaly F, Garthwaite P. Eliciting Dirichlet and Connor–Mosimann prior distributions for multinomial models. TEST 2013;**22**(4):628-46 doi: 10.1007/s11749-013-0336-4.
12. R Core Team. R: A language and environment for statistical computing. https://www.R-project.org. Vienna, Austria: R Foundation for Statistical Computing, 2016.
13. Oakley JE, O'Hagan A. SHELF: The Sheffield Elicitation Framework (version 2.0). Sheffield, UK: Schoo of Mathematics and Statistics, University of Sheffield, 2010.
14. Iglesias CP, Thompson A, Rogowski WH, et al. Reporting Guidelines for the Use of Expert Judgement in Model-Based Economic Evaluations. PharmacoEconomics 2016:1-12 doi: 10.1007/s40273-016-0425-9.
15. O'Hagan A, Buck CE, Daneshkhah A, et al. *Uncertain Judgements: Eliciting Experts' Probabilities*. Chichester, West Sussex: Wiley, 2006.
16. Zapata-Vázquez RE, O'Hagan A, Soares Bastos L. Eliciting expert judgements about a set of proportions. Journal of Applied Statistics 2014;**41**(9):1919-33 doi: 10.1080/02664763.2014.898131.
17. Soares MO, Bojke L, Dumville J, et al. Methods to elicit experts' beliefs over uncertain quantities: application to a cost effectiveness transition model of negative pressure wound therapy for severe pressure ulceration. Stat Med 2011;**30**(19):2363-80 doi: 10.1002/sim.4288.
18. Losina E, Walensky RP, Geller A, et al. Visual screening for malignant melanoma: a cost-effectiveness analysis. Arch Dermatol 2007;**143**(1):21-8 doi: 10.1001/archderm.143.1.21.
19. Wilson EC, Emery JD, Kinmonth AL, et al. The cost-effectiveness of a novel SIAscopic diagnostic aid for the management of pigmented skin lesions in primary care: a decision-analytic model. Value Health 2013;**16**(2):356-66 doi: 10.1016/j.jval.2012.12.008.

Appendix 1: Elicitation Protocol

Background & Preparation

A previously developed decision model[19] is to be adapted to a new decision question to estimate the cost-effectiveness of a self-completed checklist as a potential screening tool for malignant melanoma. An input into this model is the risk of progression of malignant melanoma from one stage to the next in patients whose melanoma is undiagnosed and hence progresses without treatment. For ethical reasons such data on the natural history of melanoma do not exist, although prognosis post diagnosis and treatment is well known.[6] The original version of the model relied on a crude assumption with arbitrarily wide confidence interval to represent the uncertainty in risk of progression. We wish to improve this for the revised model by conducting an elicitation exercise amongst group(s) of experts to elicit their opinion as to plausible risks.

There are four main types of cutaneous melanoma (superficial spreading, lentigo maligna, acral lentiginous and nodular),[7] which current guidelines categorise into nine stages of invasion.[6] All but one are also described with a pre-invasive (*in situ*) phase; nodular melanoma is by definition invasive. We wish to elicit expert opinion on the rate of progression from each stage to any other. We have simplified a possible set of 39 questions into 12 by assuming that invasive disease would progress at the same rate irrespective of primary melanoma subtype; we allow the rate of progression from *in situ* disease to vary by subtype (Table A1). Experts will be asked their opinions on the probability of moving from each of the twelve stages to any other (and death).

Table A1: Stages for study questions

In situ superficial spreading melanoma
In situ lentigo maligna melanoma
In situ acral lentiginous melanoma
Stage Ia
Stage Ib
Stage IIa
Stage IIb
Stage IIc
Stage IIIa
Stage IIIb
Stage IIIc
Stage 4

Identification and recruitment of experts

Potential participants will be identified by two of the investigators (FW & JE), based on known expertise and relevant publications in the field. They will be located in either the UK, Australia or New Zealand, and be of backgrounds including practicing and academic consultant dermatologists, oncologists, epidemiologists, plastic surgeons, and statisticians with a particular interest and expertise in melanoma.

Experts will be invited to take part via letter emailed with a reply form (files 1 and 2) to enable scheduling of elicitation sessions. These will be repeated according to availability of experts. Participants will be paid an honorarium of £200 / Aus$400 for their time.

Elicitation process

Experts willing to take part in the process will be sent a welcome pack at least one week prior to their scheduled session. This gives a brief introduction to the concepts of expert elicitation, introduction to the questions that will be asked of them, and relevant background materials (File 3 plus a copy of Balch et al. 2009[6]).

The process itself will take the form of a two-hour webinar comprising a 30-minute training session and 90 minutes to complete the exercise. The presentation (file 4) covers the material provided in the welcome pack: an introduction to the concept of expert elicitation, details of the method used and examples, and a worked example of how to use the Excel spreadsheet to record responses. The facilitator will then remain online over the next 90 minutes while participants complete the spreadsheet. This is for two reasons: firstly, to resolve any issues or questions participants may have and secondly to encourage completion of the exercise in 'real time' and discourage rapid completion at a later date.

Structuring

All questions are of the format "Imagine a cohort of 100 patients with Stage X undiagnosed and hence untreated disease. After six months, the patients may be in any of the following stages:". A bespoke Microsoft Excel spreadsheet was created to ask questions in turn and record responses (File 5). The spreadsheet randomises the order of the questions to minimise any impact of respondent fatigue and/or learning effects.

Experts are invited to select all possible stages from a drop-down list. Once the possible stages are determined, experts rank them from most to least likely. Clicking a button on the spreadsheet enters these stages into a chart.

Experts assign probabilities using the quantile method, where median and upper and lower 95% credible limits are elicited:

Starting with the least likely stage (i.e. working from the right-hand side of the chart), the expert adjusts a slider (or simply enters a number into the table below) representing their median belief about the number of patients expected to be in that stage after 6 months. They then move on to the next least likely and assign their median beliefs. The expert repeats this until the medians have been entered for all possible stages. The sum of the medians must equal 100: all 100 patients must be assigned to a state. The chart gives instant visual feedback to the expert who adjusts the medians until he/she is satisfied that his/her median beliefs are adequately represented.

The expert now assigns the lower and upper 95% credible limits to their beliefs. Again, commencing with the least likely stage the expert enters the relevant numbers. Numerically, the experts should set the limits such that they believe there is a 1 in 20 chance that the 'true values' would be outside the range, that is a 1 in 40 chance of exceeding the upper limit and a 1 in 40 chance of deceeding the lower limit. However, this is explained to the experts in the training that it should represent a value where the expert 'believes it is possible to be outside this range, but would be extremely surprised if that turned out to be the case'. Other approaches were considered, for example eliciting tertiles (33% credibility intervals), but following piloting, 95% credibility intervals were considered the least taxing method due to general familiarity with 95% confidence intervals.

Once the expert is happy with his/her response, he/she presses the 'submit' button. This timestamps and records the data within the Excel workbook, saves the spreadsheet on the participant's hard drive and moves on to the next question.

Once all questions are completed, the expert is offered the opportunity to repeat any of the questions. If not, the expert is asked to email the spreadsheet back to the facilitator and complete a feedback form (file 6).

Fitting distributions to elicited beliefs

A modified Connor-Mosimann (mCM) distribution[9 16] will be fitted to the elicited quantities from each expert. The mCM is a generalisation of the Connor-Mosimann distribution, itself a generalisation of the Dirichlet distribution, which defines a multinomial distribution. The advantage of the generalisations is the added flexibility to provide a better fit to elicited quantities. Wilson[9] provides R code[10] to fit the mCM distribution. This code will be used to fit the multinomial distribution to each expert's response to each question. To generate an aggregate distribution representing the spread of beliefs of all experts, these distributions will be sampled from many times, and the median and 95% credibility limits calculated. The combined mCM will then be fitted to these quantities.

Appendix 2: Parameters of mCM distributions

			Parameters of mCM distribution				Associated medians, upper and lower 95% credibility limits		
Experts	From	To	a	b	L	U	LL	MEDIAN	UL
ANZ	AL	AL	0.59743	0.304534	0	0.999995	0.01	0.79	1.00
		1A	9.631271	6.309027	0.221636	0.99999	0.00	0.14	0.75
		1B	2.311285	0.977214	1.00E-05	0.937206	0.00	0.03	0.29
		2A	2.983218	0.268063	0.016502	0.488581	0.00	0.01	0.08
		2B	0.487804	9.39963	0.554548	0.806302	0.00	0.00	0.06
		2C	10	8.041188	0.542353	0.771784	0.00	0.00	0.03
		3A	9.127489	0	1.00E-05	0.613719	0.00	0.00	0.01
		3B	0	0	0	0	0.00	0.00	0.01
	LM	LM	0.283675	0.131301	0.018904	0.997543	0.02	0.93	1.00
		1A	8.204178	2.194021	0.359029	0.989642	0.00	0.06	0.90
		1B	8.714595	9.670652	0.017422	0.993675	0.00	0.00	0.12
		3A	9.34394	2.577571	0.579118	0.745076	0.00	0.00	0.09
		2A	0.478753	8.624764	0.901361	0.982634	0.00	0.00	0.03
		2B	8.549465	8.683784	0.062673	0.22357	0.00	0.00	0.00
		2C	0	0	0	0	0.00	0.00	0.00
	SS	SS	0.030154	0.02798	0.028879	0.99999	0.03	0.81	1.00
		1A	10	9.000856	0.496504	0.988517	0.00	0.15	0.82
		1B	10	9.812086	1.00E-05	0.983211	0.00	0.02	0.18
		2A	10	7.72739	0.010062	0.989516	0.00	0.01	0.11
		2B	9.913193	0	0.318397	0.364845	0.00	0.00	0.03
		2C	9.904355	9.93199	0.009053	0.06772	0.00	0.00	0.00
		3A	0	0	0	0	0.00	0.01	0.06
	1A	1A	1.504461	0.667352	0	1	0.12	0.76	1.00
		1B	10	10	0	0.945677	0.00	0.11	0.45
		2A	0	10	0.439209	0.99999	0.00	0.06	0.22
		2B	0	0	1.00E-05	0.99999	0.00	0.00	0.25
		3A	10	10	1.00E-05	0.99999	0.00	0.00	0.13
		2C	1.487797	10	0.81765	1	0.00	0.00	0.11
		3B	0	0	0	1	0.00	0.00	0.02
		3C	10	0	1.00E-05	0.99999	0.00	0.00	0.02
		4	0	10	0.99999	1	0.00	0.00	0.00
		Death	0	0	0	1	0.00	0.00	0.00
		SS	0	0	0	0	0.00	0.00	0.00
	1B	1B	0.102911	0.090419	0.059234	0.936614	0.06	0.65	0.94
		2A	7.176432	8.869692	1.00E-05	0.968591	0.02	0.14	0.57
		2B	8.336562	8.602221	0.117337	0.880202	0.01	0.09	0.38
		3B	9.820876	1.205356	0.053221	0.077521	0.00	0.01	0.03
		3A	8.884178	8.545433	0.091749	0.382907	0.00	0.02	0.09
		2C	2.75241	1.327872	0.801846	0.827324	0.01	0.05	0.22
		3C	0	7.331239	0.136902	0.949866	0.00	0.00	0.01
		4	1.98482	0.072729	0.054859	0.897669	0.00	0.01	0.04
		Death	8.294663	9.220662	0.008618	0.781475	0.00	0.00	0.00

		1A	0	0	0	0	0.00	0.00	0.00
	2A	2A	1.951325	1.017888	0.009653	0.968314	0.15	0.68	0.96
		2B	4.140928	2.445235	0	1	0.03	0.19	0.61
		3A	0.622836	9.76566	0.341722	0.935004	0.00	0.04	0.17
		2C	1.553755	0.320453	0.018295	0.733605	0.00	0.04	0.18
		3B	4.479816	0	0.168717	0.344376	0.00	0.01	0.04
		4	10	8.088065	0.710776	0.895263	0.00	0.01	0.07
		3C	2.524135	10	1.00E-05	0.814571	0.00	0.00	0.00
		Death	0	0	0	0	0.00	0.00	0.01
	2B	2B	0.248326	0.207882	0.002298	0.997483	0.00	0.62	1.00
		2C	0.620873	0.470587	0.004969	0.99472	0.00	0.10	0.97
		3A	1.006865	1.779137	0.008102	0.99712	0.00	0.01	0.48
		3B	2.864799	3.488157	0.006589	0.998085	0.00	0.01	0.35
		4	1.162021	2.386737	0.005065	0.99534	0.00	0.00	0.17
		Death	0.483908	1.378194	0.004843	0.992812	0.00	0.00	0.10
		3C	1.975598	0.544843	0.012471	0.99692	0.00	0.00	0.19
		2A	0	0	0	0	0.00	0.00	0.06
	2C	2C	0.397347	0.355192	0.076073	0.894542	0.08	0.53	0.89
		3B	0	0	0.06112	0.939323	0.01	0.06	0.87
		3A	1.409543	1.820839	0.095126	0.822627	0.00	0.03	0.50
		3C	1.440143	1.137282	0.229924	0.841103	0.00	0.02	0.41
		4	2.898793	0.574577	0.099681	0.848899	0.00	0.01	0.24
		Death	2.482492	1.133833	0.117518	0.918665	0.00	0.00	0.07
		2B	0	0	0	0	0.00	0.00	0.04
	3A	3A	2.995168	1.930706	0.006841	0.994313	0.21	0.63	0.93
		3B	1.27867	1.159076	0.015874	0.990167	0.01	0.17	0.58
		3C	1.906755	2.013663	0.025001	0.976564	0.00	0.06	0.33
		4	2.31061	1.267642	0.025228	0.978739	0.00	0.04	0.24
		Death	1.358594	0.311688	0.038384	0.977857	0.00	0.01	0.13
		1A	0.430853	0.923095	0.02233	0.984153	0.00	0.00	0.02
		1B	1.066876	1.622707	0.017447	0.95757	0.00	0.00	0.01
		2A	5.490861	1.836449	0.032182	0.973223	0.00	0.00	0.02
		2B	2.352341	2.808732	0.021257	0.951185	0.00	0.00	0.00
		2C	0	0	0	0	0.00	0.00	0.00
	3B	3B	3.175643	2.840464	0	1	0.17	0.53	0.87
		3C	2.283026	1.809225	0	1	0.04	0.24	0.62
		4	0.764168	0.177426	0	1	0.01	0.14	0.50
		Death	0	0	0	0	0.00	0.00	0.27
	3C	3C	2.358183	1.996845	0.199707	1	0.30	0.64	0.94
		4	9.774438	2.525477	0.103695	1	0.05	0.29	0.60
		Death	0	0	0	0	0.01	0.06	0.20
	4	4	2.509242	0.853222	0.067355	1	0.30	0.81	1.00
		Death	0	0	0	0	0.00	0.19	0.70
UK	AL	AL	0.369688	0.229379	0.3855	0.959007	0.39	0.82	0.96
		1A	9.598789	4.070474	1.00E-05	0.99999	0.02	0.12	0.50
		1B	2.492956	0	0.432166	0.451401	0.00	0.02	0.12
		2A	8.435852	1.683278	0	0.537005	0.00	0.01	0.07
		2B	10	4.326686	0	0.297963	0.00	0.00	0.02

	2C	4.741977	8.422931	0.99999	1	0.00	0.01	0.07
	Death	0	0	0	0	0.00	0.00	0.00
LM	LM	2.301555	0.558191	0.00289	0.998852	0.29	0.88	1.00
	1A	2.62531	1.557074	0.020133	0.994679	0.00	0.07	0.51
	1B	4.816311	9.578459	0.017625	0.974256	0.00	0.01	0.12
	2A	9.505588	9.616541	0.005425	0.97701	0.00	0.01	0.12
	2B	9.897045	5.116054	0.006307	0.992072	0.00	0.01	0.08
	Death	0	0	0	0	0.00	0.00	0.04
SS	SS	3.279439	0.882013	0.022255	0.99999	0.37	0.84	1.00
	1A	2.12226	0.230381	0.01257	0.822724	0.00	0.12	0.49
	1B	10	1.000246	1.00E-05	0.578567	0.00	0.02	0.11
	2A	3.886608	0.829852	1.00E-05	0.880147	0.00	0.01	0.07
	2C	0	10	0.394195	1	0.00	0.00	0.01
	Death	0	0	0	0	0.00	0.00	0.02
1A	1A	0.174328	0.146953	0	1	0.00	0.65	1.00
	1B	0	10	0.546081	1	0.00	0.19	0.55
	2A	2.360595	10	0	0.775024	0.00	0.02	0.13
	2B	0	0	0	0.635571	0.00	0.00	0.27
	3A	0.564205	4.175201	1.00E-05	0.99999	0.00	0.00	0.11
	3B	10	10	0.73863	1	0.00	0.05	0.35
	2C	10	10	0.144976	1	0.00	0.00	0.03
	3C	10	10	0.030998	0.706957	0.00	0.00	0.01
	4	0	10	0	1	0.00	0.00	0.00
	Death	0	0	0	0	0.00	0.00	0.02
1B	1B	0.223691	0.182173	0.171267	0.920287	0.17	0.71	0.92
	2A	7.92191	3.742915	0	0.472585	0.02	0.09	0.33
	2B	6.852427	5.311878	0.188886	0.364835	0.01	0.06	0.19
	2C	3.170922	3.883282	0.013786	0.625196	0.01	0.04	0.18
	3A	0	8.025806	0.563521	0.708222	0.01	0.05	0.20
	3B	3.350625	8.537315	0.106292	0.129229	0.00	0.00	0.02
	3C	7.109377	4.703316	0.16711	0.865774	0.00	0.02	0.09
	Death	3.219595	3.193568	0.357788	1	0.00	0.01	0.05
	4	0	0	0	0	0.00	0.00	0.03
2A	2A	0.659129	0.447355	0.010228	0.861544	0.02	0.58	0.86
	2B	1.436617	3.423216	0.215948	1	0.04	0.18	0.59
	2C	5.873514	2.755437	1.00E-05	0.496491	0.01	0.07	0.25
	3A	8.252809	0	1.00E-05	0.493687	0.02	0.07	0.22
	3B	0	0	0.067978	0.735867	0.00	0.01	0.15
	4	7.87938	10	1.00E-05	0.657204	0.00	0.01	0.06
	3C	10	3.688372	0	1	0.00	0.02	0.11
	Death	0	0	0	0	0.00	0.01	0.04
2B	2B	1.534297	2.030859	0.00039	0.999846	0.05	0.42	0.88
	2C	2.097513	8.208028	0.169449	0.998589	0.04	0.18	0.42
	3A	0.817723	0.97243	1.00E-05	0.999088	0.00	0.14	0.52
	3B	0.896715	0	0.003156	0.494168	0.00	0.09	0.27
	3C	0.012378	0.004207	0.003793	0.999202	0.00	0.05	0.27
	4	9.980448	0.012976	0.999302	0.99999	0.00	0.00	0.22
	Death	0	0	0	0	0.00	0.00	0.00

	2C	2C	2.507996	4.105559	0.006301	0.991172	0.09	0.37	0.74
		3B	1.289735	3.476961	1.00E-05	0.977609	0.01	0.14	0.48
		3C	1.095318	2.461759	1.00E-05	0.99999	0.01	0.11	0.44
		3A	9.937615	4.300662	0.022216	0.98025	0.04	0.20	0.48
		4	0.130387	0.030762	0	0.967015	0.00	0.07	0.24
		Death	0	0	0	0	0.00	0.00	0.16
	3A	3A	1.071636	0.966509	0.005961	0.985589	0.04	0.53	0.97
		3B	2.55411	3.56696	0.006701	0.983128	0.01	0.16	0.57
		3C	0.285266	0.255283	0.010596	0.998774	0.00	0.08	0.58
		4	4.801688	2.091451	0.016015	0.993641	0.00	0.04	0.41
		Death	0	0	0	0	0.00	0.01	0.21
	3B	3C	2.969611	4.448676	0.002981	0.933115	0.10	0.37	0.69
		3B	0.349255	0.45477	0.292156	0.946387	0.12	0.31	0.75
		4	8.199244	4.628028	0.086768	1	0.02	0.17	0.42
		Death	0	0	0	0	0.01	0.07	0.25
	3C	3C	3.290765	4.27166	0.096936	0.933029	0.21	0.45	0.74
		4	4.453119	1.807437	0.010596	0.983427	0.14	0.37	0.65
		Death	0	0	0	0	0.03	0.14	0.40
	4	4	5.276575	4.389197	0.002909	0.99769	0.24	0.55	0.83
		Death	0	0	0	0	0.17	0.45	0.76
ALL	AL	AL	0.839126	0.385843	1.00E-05	0.99999	0.03	0.79	1.00
		1A	1.492014	0.247581	0	0.759439	0.00	0.13	0.70
		1B	10	10	0	0.969459	0.00	0.03	0.23
		2A	6.602994	8.077545	1.00E-05	0.776528	0.00	0.01	0.09
		2B	8.717574	9.295483	0	0.671483	0.00	0.01	0.05
		2C	9.633907	10	0.246869	0.810438	0.00	0.01	0.06
		3A	8.704724	9.195737	0	0.972577	0.00	0.00	0.02
		3B	0.708039	10	0.086467	1	0.00	0.00	0.00
		Death	0	0	0	0	0.00	0.00	0.02
	LM	LM	0.167053	0.10176	0.022832	0.999198	0.02	0.92	1.00
		1A	9.623585	2.133245	0.07268	0.955987	0.00	0.06	0.88
		1B	9.884843	9.987646	0.038889	0.604451	0.00	0.00	0.11
		3A	2.411535	9.960102	0.100902	0.994081	0.00	0.00	0.07
		2A	9.963524	3.22278	0.01894	0.664518	0.00	0.00	0.09
		2B	9.810062	9.969743	0.649155	0.964291	0.00	0.00	0.07
		2C	8.907087	8.113418	0.958967	0.987231	0.00	0.00	0.02
		Death	0	0	0	0	0.00	0.00	0.00
	SS	SS	0.739052	0.319184	0.002824	0.995898	0.03	0.83	1.00
		1A	2.068897	2.033196	0.610485	1	0.00	0.14	0.82
		1B	9.054651	6.031458	0.011876	0.973273	0.00	0.02	0.15
		2A	9.687131	2.788773	0.033901	0.902645	0.00	0.01	0.08
		2B	7.133308	9.546354	0.264258	0.989727	0.00	0.00	0.02
		2C	0.44956	0.136839	0.038445	0.621446	0.00	0.00	0.01
		3A	9.631037	0.264069	0.004024	1	0.00	0.00	0.01
		Death	0	0	0	0	0.00	0.00	0.00
	1A	1A	1.034381	0.568252	1.00E-05	0.99999	0.05	0.72	1.00
		1B	9.733502	9.973478	0.004373	1	0.00	0.14	0.53
		2A	2.931828	9.951031	0.001945	1	0.00	0.03	0.16

	2B	0.051924	0.232846	0.009194	0.984818	0.00	0.00	0.26
	3A	9.86484	9.992441	1.00E-05	0.365153	0.00	0.01	0.08
	2C	0.010515	0.125603	0.030811	0.989357	0.00	0.00	0.10
	3B	9.867243	9.902834	0.02078	0.993498	0.00	0.02	0.17
	3C	0	9.994807	0.013573	0.961755	0.00	0.00	0.00
	4	0	9.971261	0.009402	0.99999	0.00	0.00	0.00
	Death	0.160857	10	0	0.991819	0.00	0.00	0.01
	SS	0	0	0	0	0.00	0.02	0.16
1B	1B	0.448233	0.294108	0.098371	0.923842	0.10	0.69	0.92
	2A	10	9.914674	0	0.957017	0.03	0.14	0.52
	2B	0.892612	1.658899	1.00E-05	0.962568	0.00	0.03	0.33
	2C	0	9.695914	0.392525	0.909939	0.00	0.03	0.18
	3A	0.232188	0.398362	1.00E-05	0.939368	0.00	0.01	0.18
	3B	9.841029	9.398021	0	0.93445	0.00	0.01	0.12
	3C	9.525846	9.76894	0.064622	0.967901	0.00	0.01	0.07
	4	9.808781	0.844963	0.081298	0.931483	0.00	0.01	0.06
	Death	0.787374	10	0.965123	0.99999	0.00	0.00	0.01
	1A	0	0	0	0	0.00	0.00	0.00
2A	2A	1.687969	1.004313	1.00E-05	0.91826	0.11	0.61	0.91
	2B	2.725174	2.357424	0.02302	0.962066	0.03	0.18	0.60
	3A	8.994734	9.106172	0	0.628603	0.01	0.05	0.19
	2C	0.523682	0.622619	0.219633	0.883525	0.01	0.05	0.25
	3B	5.261794	8.58761	0.265845	0.691082	0.00	0.02	0.10
	4	0.077742	1.900017	0.469569	0.81829	0.00	0.01	0.07
	3C	9.482078	2.686904	0.658586	0.739639	0.00	0.01	0.05
	Death	0	0	0	0	0.00	0.00	0.02
2B	2B	0.259866	0.257236	0.009117	0.994492	0.01	0.52	0.99
	2C	4.960413	10	0.007869	0.986863	0.00	0.14	0.49
	3A	2.068522	2.495441	0.027659	0.977224	0.00	0.11	0.52
	3B	4.565419	5.172593	0.019409	0.991251	0.00	0.06	0.31
	3C	0.139852	0.23038	0.009796	0.990614	0.00	0.00	0.24
	4	0.387311	0.262258	0.009867	0.984233	0.00	0.00	0.21
	Death	6.786227	8.605502	0.00828	0.981074	0.00	0.00	0.08
	2A	0	0	0	0	0.00	0.00	0.11
2C	2C	1.977441	2.396082	0.014196	0.953622	0.09	0.44	0.83
	3B	0.129428	0.174102	0.057696	0.98636	0.01	0.14	0.82
	3A	1.036033	1.319266	0	0.942916	0.00	0.07	0.52
	3C	1.813304	1.456001	0.030214	0.931945	0.00	0.05	0.41
	4	4.445778	1.758624	0.061265	0.927437	0.00	0.03	0.26
	Death	4.173727	1.342888	0.074551	0.922388	0.00	0.01	0.10
	2B	0	0	0	0	0.00	0.00	0.04
3A	3A	1.392516	0.992749	0.012544	0.976772	0.08	0.61	0.96
	3B	1.36006	1.643439	0.014562	0.986543	0.01	0.15	0.63
	3C	0.983344	0.586824	0.010825	1	0.00	0.09	0.54
	4	1.690844	0.340197	0.012112	0.991131	0.00	0.03	0.35
	Death	0.437478	0.071505	0.026964	0.992584	0.00	0.00	0.10
	1A	0.034552	0.218092	0.021633	0.993463	0.00	0.00	0.00
	1B	4.673151	0.697417	0.020752	0.968636	0.00	0.00	0.01

	2A	0.032524	0.32729	0.007318	0.989644	0.00	0.00	0.00
	2B	0.05443	0.035732	0.014095	0.992498	0.00	0.00	0.00
	2C	0	0	0	0	0.00	0.00	0.00
3B	3B	1.25305	2.102565	0.12553	0.994641	0.15	0.43	0.87
	3C	2.722162	1.931761	0.003412	0.99726	0.05	0.30	0.68
	4	2.21584	1.484447	0.339041	0.996612	0.02	0.14	0.44
	Death	0	0	0	0	0.00	0.04	0.21
3C	3C	3.785295	2.926405	0.007746	0.988342	0.21	0.57	0.87
	4	4.02447	1.405257	0	1	0.08	0.31	0.66
	Death	0	0	0	0	0.01	0.09	0.36
4	4	1.841319	1.456665	0.170922	0.99999	0.26	0.64	0.96
	Death	0	0	0	0	0.04	0.36	0.74

Medians do not sum to 100% due to rounding. AL = in situ acral lentiginous melanoma; LM = in situ lentigo maligna melanoma SS = in situ superficial spreading melanoma

Appendix 3: Excluded data from time to completion analysis

Of 14 respondents, 7 reported a time to completion of the first question of greater than 30 minutes. We assumed these were due to respondents opening the spreadsheet prior to commencement of the exercise. Two respondents also reported time stamps of 13 and 19 hours, and one of these respondents subsequently reported a time of 1hr8mins for one question. The 13 and 19 hours were cases where respondents had requested adjournment and completion of the exercise the following day. The reason for the 1h08 time over one question is unknown, but is consistent with a break from the exercise (it was on the third question following overnight adjournment).