## ABSTRACT

**Background:** Population screening for renal cell carcinoma (RCC) using ultrasound has the potential to improve survival outcomes; however a cost-effectiveness analysis (CEA) has yet to be performed. Due to the lack of existing evidence, we performed structured expert elicitation to derive unknown quantities to inform the CEA.

**Objectives:** To elicit the cancer stage distribution (proportion of individuals with each stage of cancer) for different RCC screening scenarios and the annual transition probabilities for undiagnosed disease becoming diagnosed in the NHS.

**Methods**: The study design and reporting adhered to the Reporting Guidelines for the Use of Expert Judgement in Model-Based Economic Evaluations. The elicitation was conducted face-to-face or via telephone between each individual expert and the facilitator, aided by online material. For multinomial data, Connor Mosimann and modified Connor Mosimann distributions were fitted for each expert and for all experts combined using mathematical linear pooling.

**Results**: A total of 24 clinical experts were invited, and 71% participated (7 urologists, 6 oncologists, 4 radiologists). The modified Connor Mosimann distribution provided the best fit for the majority of elicited quantities. Greater uncertainty was noted for the elicited transition probabilities compared to the elicited stage distributions.

**Conclusion**: We performed the first expert elicitation of RCC screening parameters, crucial information which will inform the CEA of screening. Additionally, the elicited quantities may enable future health economic evaluations assessing the value of diagnostic tools and pathways in RCC.

## HIGHLIGHTS

What is already known?

- A cost-effectiveness of analysis of screening for renal cancer has yet to be performed
- Structured expert elicitation may inform cost-effectiveness models, as values of interest may either be unknown or not directly measurable through research

What does the paper add to existing knowledge?

- We have performed the first structured expert elicitation of renal cancer screening parameters, recruiting a multidisciplinary group of 17 experts (urology, radiology and oncology specialists)
- Our approach, consisting of one-to-one interviews between experts and the facilitator which can be delivered online or in person, facilitate data collection, maximise study feasibility and limit incomplete data collection
- This crucial information which will inform a cost-effectiveness analysis

**MANUSCRIPT**

## Introduction

Renal cell carcinoma (RCC) is the 7th most common malignancy in the UK [1]. Nearly 30% of patients have metastases at presentation, and five year relative survival in these patients is only 6% [2]. As such, there is an increasing interest in the role of screening for RCC using focused renal ultrasound as a potential method to downstage the disease and save lives [3, 4]. One of the perceived barriers to population screening remains the relatively low prevalence of the disease in unselected cohorts. It has been estimated that population screening of asymptomatic individuals would lead to the identification of between 1 and 3 patients with RCC, for every 1000 individuals screened [5]. Therefore, a randomised controlled trial (RCT) sufficiently powered to detect an impact on survival would need to recruit hundreds of thousands of participants. Additionally, there is currently insufficient evidence regarding the benefits of screening for RCC and it is therefore not surprising that such a RCT has yet to be undertaken. Small observational studies, recruiting in the order of 10,000 participants, have been performed but have collected only limited data [6, 7]. Classically health economic evaluations are performed following the publication of clinical studies in order to aid decisions regarding the value of implementing new strategies in the health service. Instead, we propose that a cost-effectiveness analysis (CEA) and value of information analysis (VOI) of screening for RCC may be undertaken using the limited available evidence, prior to a large trial being undertaken, aiming to determine the value of investing future funds into a large-scale trial. Indeed, VOI has been used to examine uncertainty surrounding the optimal screening strategy for colorectal cancer and therefore prioritise future research efforts [8].

CEAs often require information which is either not (yet) known or not directly measurable through research. There has been a growing body of work regarding expert elicitation to derive information for health economic evaluation [9]. In the case of screening for RCC, there are a number of unknown quantities, which if determined through expert elicitation rather than a large-scale trial, could inform a preliminary cost-effectiveness model and VOI. Expert elicitation is defined as the quantitative process of deriving a point estimate of an unknown value and, critically, a probability distribution around that value, pooling knowledge from one or more experts [10]. The probability distribution is subjective as it represents the strength of an expert's belief that an unknown quantity of interest lies between two values.

A number of methods have been described to conduct expert elicitation, gather data and aggregate results, each with different advantages and disadvantages. Perhaps the most renowned method is the SHeffield ELicitation Framework (SHELF) [9, 11]. This describes the process of elicitation through a face-to-face group discussion amongst a small group of experts. Recognised limitations include: the time-consuming nature of the exercise, high cost and logistical and geographical constraints in assembling national and international experts together in one location [10]. A number of alternative methods have therefore been described to conduct expert elicitation to bypass these limitations, including independent face-to-face assessment of each expert and online assessments [12, 13]. We performed a modification of existing expert elicitation methods that facilitate data collection and maximise study feasibility.

The aim of this study was to perform a structured expert elicitation exercise to derive unknown quantities with which to inform a CEA and VOI of screening for RCC. The results

of the elicited quantities are of interest to the reader, as they may enable future health economic evaluations in RCC.

## Methods

The design of the study and reporting of results adhered to guidelines described by Iglesias *et al* [10]. Ethics approval was not required [14]. Data were anonymised and no financial incentive was provided for the experts. The full expert elicitation exercise, training material and evidence dossier are found in the Appendix.

### *Research rationale and quantities of interest*

A decision model was developed, adopting a UK National Health Service (NHS) perspective to compare the incremental costs and quality adjusted life years of a hypothetical cohort of asymptomatic individuals from the general population, undergoing screening via focused renal ultrasound ("one off" screening similar to established abdominal aortic aneurysm screening) compared to the standard of care (no screening). The model was populated with data identified from a review of the literature. We identified a number of parameters for which no data were available, for which we sought expert opinion. The focus of this manuscript is on the expert elicitation process and results. The decision-model based economic analysis will be reported elsewhere.

We elicited the cancer stage distribution (i.e. the proportion of individuals in each stage) for individuals with undiagnosed RCC, individuals with RCC who did not attend screening and individuals with RCC who were false negatives at screening. The former encompasses all

patients who have RCC but are still undiagnosed in the current NHS (in the absence of screening). These individuals are asymptomatic and there is a potential for diagnosis if screening were implemented with a perfect test (i.e. no false negatives). Data regarding individuals with RCC who are screening non-attenders is unknown and by definition cannot be measured even in the context of a prospective trial as these individuals do not attend. Data regarding the stage distribution of false negatives at screening is sparse. A previously performed systematic review and meta-analysis identified seven studies that delivered ultrasound for the identification of RCC in a screening paradigm [5]. Only two studies reported false negative rates. In the study by Filipas *et al.,* in order to identify false negatives at initial screening ultrasound, participants were re-invited to attend for a second ultrasound one year later, in addition to monitoring cancer registry data [6]. In the study by Mizuma *et al.*, false negatives were only identified by monitoring registry data rather than repeat imaging [15]. In these two studies combined, despite screening 25,983 individuals, only three false negative cases were identified. This may either represent the high sensitivity of ultrasound or inherent weakness in registry data. Of the three cases, two were stage I (T1a), while the third patient had a RCC 2cm in size but had concomitant metastases and therefore was stage IV. These data were presented to the clinical experts and it was felt not to represent the true stage distribution of false negative disease; the small sample size meant that the proportion of false negatives with stage IV RCC (33%) was over-estimated by the current data. As such, the decision was made to include this parameter in the expert elicitation. Furthermore, we elicited expert opinion regarding the probability of undiagnosed RCC becoming diagnosed in the NHS in the absence of screening (annual transition probability), as this information cannot be directly measured in a clinical trial.

***Objectives***

1. Determine the stage distribution of:
    a. Individuals with undiagnosed RCC in the current NHS (in the absence of screening)
    b. Individuals with RCC who do not attend screening
    c. Individuals with RCC who are false negatives at screening

2. Determine probability of undiagnosed RCC becoming diagnosed (annual transition probability) in the current NHS, in the absence of screening
    a. By RCC stage
    b. Overall (all stages combined)

***Expert selection and inclusion***

Relevant experts in urology, radiology and oncology were identified by the authors based on their expertise in the diagnosis and management of RCC. Participants were selected from a variety of centres across the UK, including both academic centres and district general hospitals, to capture nationally representative data. A mixture of individuals with both academic and clinical expertise were invited to participate. Exclusion criteria consisted of a known conflict of interest or known unavailability.

***Question framing and piloting***

The expert elicitation exercise was created using Microsoft Excel (Company: Microsoft Redmond, Washington, USA). This contained six questions relevant to RCC. The final two questions asked participants to score their overall confidence in the answers provided and to

score how accurately the data provided represented their beliefs (face validity) on a six-point Likert scale, as previously described [12]. The questions were framed in a specific order (not randomised) to allow the scenarios to build on each other. At the end, participants were asked to review their answers for internal consistency, with both the original and revised answers being recorded. The parameter of interest was the stage distribution of patients with RCC, multinomial data which may be expressed as a proportion, with questions framed as: "Imagine 100 individuals with RCC... how many individuals will be in each stage?" [12]. To elicit the annual transition probability of diagnosis, experts were asked: "Imagine 100 individuals with undiagnosed RCC. In the absence of screening, after one year, how many of these will have been diagnosed?" Utilising a sample size of 100 (rather than any larger number) minimises spurious precision by limiting proportions to two significant figures [12]. The quantile approach to elicit quantities of interest was chosen. First, participants were asked to report 95% credibility intervals (CrI) for each stage, followed by median estimates. Participants were required to generate median estimates for each stage that would sum to a total of 100. No constraints were placed on the 95% CrI.

The elicitation process was piloted on three individuals. Several different graphical options were proposed relating to the stage distribution of RCC (questions 1-4) and the one voted most "user-friendly" by the pilot group adopted in the final study. The Excel document was programmed to automatically update the graphical representation as participants typed in their answers, to allow experts to visualise their answers in real time. Furthermore, based on the pilot group, the decision was made to feedback elicitation results to study participants in a graphical format, as this was judged to be easier to comprehend.

### *The elicitation process*

Experts were invited to participate via email. An evidence dossier summarising known evidence was emailed to participants in advance of the elicitation exercise, and experts were encouraged to submit additional information to be shared with other participants. The evidence dossier was kept brief to avoid expert fatigue, however a full bibliography was provided (Appendix).

The elicitation exercise was conducted either face-to-face or via telephone between each individual expert and the facilitator (SHR), with the aid of online material. Experts received a copy of the training material and the Microsoft Excel elicitation exercise via email. The facilitator delivered one-on-one training regarding the elicitation process to each expert, using a power point presentation and material adapted from previous elicitation training materials and exercises [9, 12]. Each expert was guided through two "practice questions" unrelated to RCC to familiarise them with the process. The experts' performance on the "practice questions" was used to assess whether they had understood the elicitation process. None of the experts required further training or practice questions. Participants were then guided through the elicitation process and encouraged to ask questions, whilst completing the exercise. The training and elicitation process took approximately two hours.

### *Feedback*

Mathematical aggregation was performed to pool data from the experts and results were emailed to each individual, demonstrating their individual raw data as well as the aggregated results in a graphical format. Experts were given two weeks to feedback any desired changes.

***Analysis***

Individual experts reported 95% CrI and median estimates for a hypothetical cohort of 100 patients with RCC. These were converted into proportions prior to analysis. Mathematical linear pooling was performed, with experts weighted equally. For multinomial data, three different distributions were applied (Dirichlet, Connor Mosimann and modified Connor Mosimann distributions) using the modcmfitr package in R, as previously described [16]. Briefly, the Dirichlet is a generalisation of the binomial distribution, and the Connor-Mosimann (CM) and modified Connor Mosimann (mCM) are further generalisations allowing more flexible distributions and thus in theory providing a better fit to elicited data. Fitting the distributions makes use of a directed random search algorithm to find the best fitting distributions for each elicited quantity for each expert. These distributions are then sampled from many times, and the pooled median and 95% CrI calculated. A distribution is then fitted to these pooled quantiles, representing the aggregation of all the experts' beliefs. Results from the Dirichlet, CM and mCM distributions were compared against the elicited quantiles graphically (by plotting medians and credibility intervals) and directly by comparing the goodness of fit measure (SSD, sum of squared deviation between the modelled and the elicited quantiles). For binomial data, beta distributions were applied and results represented graphically by plotting medians and 95% CrI [12].

## Results

A total of 24 experts were invited and 71% participated (17 total: 7 urologists, 6 oncologists, 4 radiologists; Figure 1). Participating experts were based at a variety of institutions in England and Scotland, including major University teaching hospitals and district general hospitals.

For multinomial data, the Dirichlet, CM and mCM distributions were fitted as planned in the prespecified analysis. The Dirichlet distribution provided a very poor fit, therefore results are not shown (available from corresponding author). The CM distribution provided the closest fit for the elicited cancer stage distribution of undiagnosed RCC in the current NHS, and the mCM distribution provided the best fit for the remaining two elicited quantities (Table S1). The parameters of the distributions that provided the best fit are shown in table 1 (parameters of alternative distributions are in Supplemental Tables S2-S4). The modelled median and 95% CrI are shown in figures 2-4.

Figure 2 demonstrates the stage distribution of undiagnosed RCC in the current NHS (in the absence of screening). The pooled data suggested that the majority of patients would be in stage I T1a and stage I T1b. Figure 3 demonstrates the stage distribution of individuals with RCC in screening non-attenders. The elicited data suggested that the stage distribution in screening non-attenders is slightly shifted towards more advanced disease compared to overall undiagnosed RCC in the current NHS, with a marginally smaller proportion in stage I T1a and T1b, marginally larger proportion in stages II and III and equal proportions in stage IV. Figure 4 demonstrates the stage distribution of individuals with RCC who are false

negatives at screening, based on data pooled from all the experts. The median estimate for stage I T1a RCC is over 75%, with a median estimate of <10% for stage IV.

For the binomial data, beta distributions were fitted (Table S5). Figure 5 demonstrates the probability of undiagnosed renal cancer becoming diagnosis in the current NHS (annual transition probability), by RCC stage. Data are demonstrated for radiologists, oncologists and urologists analysed separately (Figure 5A-5C). Pooled estimates for all experts combined are shown in supplemental figure S1.

Generally, there was more uncertainty (i.e. wider credibility intervals) noted for the elicited transition probabilities compared to the elicited stage distributions in figures 2-4.
For all experts, the greatest uncertainty was noted for stage III and IV disease (Figure 5). Urologists consistently estimated a higher probability of localised disease (Stages T1a, T1b and stage II) becoming diagnosed compared to the other expert groups, while radiologists estimated the probability of stage IV disease becoming diagnosed higher than the other expert groups (Figure 5A-C).

Two experts submitted additional information for the evidence dossier. Experts were emailed the results obtained by pooling estimates from all study participants, presented in a graphical format. None of the experts wished to make modifications to the data.

On a six-point Likert scale, where 1 is not confident and 6 is very confident, the mean and median overall confidence scores were 3.8 and 4 respectively (range 2-5). The mean and median face validity scores were 5.25 and 5 (range 4-6).

## Discussion

In summary, we elicited expert opinion regarding the stage distribution of individuals with RCC in a number of scenarios and the annual transition probability from undiagnosed to diagnosed RCC in the absence of screening. Due to the nature of the condition of interest, i.e. undiagnosed disease and/or screening non-attenders, this information cannot be estimated by a clinical trial or prospective cohort study. Therefore, quantifying these data through expert elicitation is paramount to inform a cost-effectiveness analysis.

### *Interpretation*

We report pooled estimates from seventeen RCC experts from across the UK, regarding the stage distribution (i.e. the proportion of individuals in each stage) for individuals with undiagnosed RCC in the absence of screening and individuals with RCC who did not attend screening. The experts felt that the stage distribution in screening non-attenders would be slightly shifted towards more advanced disease compared to screening attenders (2% stage IV RCC in attenders *vs* 8% in non-attenders) [17]. The clinical relevance of this remains to be investigated. Research in other disease areas suggests that screening non-attenders have significantly more risk factors for disease and are more often from lower socioeconomic backgrounds than screening attenders [18]. There is a growing interest in targeted screening of high-risk individuals and indeed a recent pilot study of community based lung cancer screening demonstrates that high pick-up rates can be achieved by targeting individuals in high-risk and deprived areas by maximising convenience of screening (such as minimizing travel time) [19]. This represents an interesting avenue to be explored.

One of the elicited parameters in this study was the stage distribution of individuals with RCC who are false negatives at screening. Indeed, ultrasound enables the detection of only 67-82% of tumours 2-3cm in size, therefore there is a potential for false negatives in small masses [4, 20-22]. Two studies in the literature have reported false negative rates of screening for RCC to date. Due to the small sample size (n=3), these studies over-estimated the proportion of false negatives with stage IV RCC (33%) [6, 15]. The experts felt that the median estimate for stage I T1a RCC would be over 75%, with a median estimate of <10% for stage IV. This reflects the experts' knowledge of the accuracy of ultrasound and the known evidence. This information will allow a more plausible cost-effectiveness model in the context of the limited available data in the literature.

We also report the probability of undiagnosed RCC becoming diagnosed in the NHS, by stage (annual transition probability). Urologists consistently estimated a higher probability of localised disease (Stages T1a, T1b and stage II) becoming diagnosed compared to the other expert groups. This may reflect urologists' clinical exposure and experience investigating and managing localised disease. Furthermore, radiologists estimated the probability of stage IV disease becoming diagnosed higher than the other expert groups, potentially reflecting their imaging expertise. This highlights the importance of including experts from different backgrounds. When elicited parameters vary between expert groups, it may be useful to perform a sensitivity analysis in the CEA to evaluate the impact of heterogeneity of results on the model outcome.

***Strengths and Limitations***

We adopted a modification of existing expert elicitation methods, an approach that has several strengths and limitations that should be considered when interpreting the data obtained [13]. One-on-one interaction between the expert and facilitator, delivered online or in person, was selected to maximise feasibility and practicality. This allowed a larger number of experts to be recruited than conventional group methods, bypassing any inconveniences relating to individual participants' time schedules and geographical constraints [23]. Indeed, SHELF methods recommend between six and ten experts, whereas we were able to recruit seventeen. The process also allows individual experts to remain anonymous and avoids "group thinking" which may occur in the presence of assertive or dominant experts within a group exercise [24], although an experienced group facilitator may be able to skilfully avoid this. In addition, one-on-one interaction with the facilitator ensures that the expert fully understands the elicitation process, engages with the exercise and completes all the components. This avoids incomplete submissions or the exercise being interrupted, which is often noted in purely online group methods previously described [12]. However, a potential limitation is that experts were asked to feedback on pooled data via email. None of our experts wished to change their results. It is unclear if this is due to lack of engagement or whether the pooled data accurately represented their beliefs. Two experts commented on the evidence dossier and one added additional evidence for circulation to the group, suggesting a degree of engagement.

A number of different methods to elicit the quantity of interest have been described. In the histogram and chip-and-bin approaches, experts are asked to place crosses or chips respectively, in equally sized bins at fixed intervals, representing the strength of their belief regarding a specific quantity of interest [9]. This allows the expert to visualise graphically

their probability density function (PDF) [23]. For example, this method has been used in a Health Technology Assessment of Screening for Oral Cancer to estimate the proportion of individuals with undiagnosed oral cancer whose disease will progress to a higher stage within one year [25]. A reported limitation of this method is that experts tend to focus on the shape of the PDF rather than the probabilities [9]. Another commonly used method is the bisection or quantile approach, in which experts are asked to report median estimates for a particular quantity of interest, in addition to other quantiles, usually tertiles or quartiles. This allows the expert to divide their plausible range into equiprobable sections. However, clinical experts who do not have a statistical background often experience difficulty placing tertiles and quartiles, underestimating the proximity of these values to the median [9]. We utilized the quantile approach to elicit parameters of interest. Importantly, participants were asked to report 95% credibility intervals first, followed by median estimates, as this has been suggested to minimize heuristics such as anchoring and overconfidence [26]. We acknowledge that using the median and 95% credibility intervals reduces the available data and may make it more difficult to fit a parametric distribution, however this method is easier to understand for clinical experts through familiarity with 95% confidence intervals and therefore increases internal consistency [27]. In addition, clinicians report higher face validity with the quantile approach [28]. Indeed, our participants reported a mean face validity score of 5.25 out of 6, consolidating these findings. There is a paucity of studies directly comparing the impact of different elicitation methodology on outcomes and further comparative research in this field is paramount [29].

Generally, more uncertainty (i.e. wider credibility intervals) was noted for the elicited transition probabilities compared to the elicited stage distributions. This may reflect the clinical experts' genuine uncertainty as to appropriate values. However, this could also

reflect relatively low familiarity with the concept of transition probabilities and difficulty estimating rate of change, which may impact interpretation of results. To overcome this known challenge, particular attention was paid to training experts and questions regarding transition probabilities were framed as frequencies rather than probabilities, as previously described [12, 29]. It is important to ensure, through adequate training, that experts are expressing uncertainty rather than heterogeneity. In addition, the training session alerted the experts to the concept of cognitive biases such as anchoring and overconfidence, aiming to reduce these. There is no established method to evaluate whether experts have achieved sufficient normative skills, therefore studies often request experts to rate their overall confidence in the expert elicitation process [12]. The mean confidence scores reported by our experts were similar to those previously reported in other expert elicitation exercises, suggesting experts were satisfied with the training received and the answers provided [12].

We performed mathematical linear pooling with equal weighting for all experts as there is uncertainty regarding the identification of suitable seed questions which are clinically meaningful and which accurately represent the experts' knowledge [29]. Furthermore, introducing seed questions may increase the overall number of questions in the elicitation and contribute to expert fatigue [13, 23]. We acknowledge that the absence of seed questions precluded an analysis of expert calibration. Further research into the quantity and quality of seed questions has been identified as a key research priority [29].

## Conclusion

We report pooled data from seventeen RCC experts from across the UK, across three different clinical specialties. The relatively large sample size, compared to other expert elicitation exercises of this nature, increases the reliability and validity of our results. One-on-one interaction between each expert and the facilitator aided data collection, maximised study feasibility and limited incomplete data submissions. We have performed the first expert elicitation of RCC screening parameters, crucial information which will inform a CEA of screening. Additionally, the elicited quantities may enable future CEAs assessing the value of diagnostic tools and pathways in RCC.

**Figure Legends**

Figure 1: Study participants and host institutions

Figure 2: Stage distribution of undiagnosed renal cancer in the current NHS (pooled expert beliefs), including median (dot) and 95% credibility intervals (lines)

Figure 3: Stage distribution of individuals with renal cancer who do not attend screening (pooled expert beliefs), including median (dot) and 95% credibility intervals (lines)

Figure 4: Stage distribution of individuals with renal cancer who are false negatives at screening (pooled expert beliefs), including median (dot) and 95% credibility intervals (lines)

Figure 5: Probability of undiagnosed renal cancer becoming diagnosed in the current NHS (annual transition probability), by renal cancer stage. Median (dot) and 95% credibility intervals (lines) are shown. Data is demonstrated for radiologists (A), oncologists (B) and urologists (C) analysed separately. Data demonstrating values for all experts combined is demonstrated in the supplemental materials (Figure S1).

Table 1: Table 1 demonstrates the pooled data from all 17 experts regarding the stage distribution of RCC in three scenarios: undiagnosed individuals in the current NHS, screening non-attenders and screening false negatives. Parameters are shown for the distribution which provided the best fit, along with modelled medians (Med) and 95% lower and upper credibility intervals (LCrI and UCrI).*

| | Stage distribution of undiagnosed RCC in the NHS | | | | | | | Stage distribution of screening non-attenders | | | | | | | | | Stage distribution of false negatives at screening | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Connor Mosimann distribution | | | | | | | Modified Connor Mosimann distribution | | | | | | | | | Modified Connor Mosimann distribution | | | | | | | |
| Stage | Zed | LCrI | Med | UCrI | a | b | | Zed | LCrI | Med | UCrI | a | b | L | U | | Zed | LCrI | Med | UCrI | a | b | L | U |
| I T1a | 1 | 0.14 | 0.39 | 0.7 | 4.228455 | 6.262833 | | 1 | 0.11 | 0.31 | 0.54 | 3.929694 | 5.093263 | 0.001468 | 0.721504 | | 1 | 0.43 | 0.76 | 0.95 | 6.7170 | 2.4098 | 0 | 1 |
| I T1b | 2 | 0.04 | 0.25 | 0.58 | 2.077874 | 2.643674 | | 2 | 0.11 | 0.29 | 0.49 | 4.030906 | 2.812999 | 0.004625 | 0.721208 | | 2 | 0.01 | 0.09 | 0.44 | 0.3457 | 0.4867 | 0.157376 | 1 |
| II | 4 | 0.01 | 0.08 | 0.32 | 1.140113 | 1.223572 | | 5 | 0.04 | 0.13 | 0.31 | - | - | - | - | | 4 | 0 | 0.01 | 0.14 | 10.00 | 10.00 | 1.00E-05 | 0.99999 |
| III | 3 | 0.02 | 0.10 | 0.28 | 5.321397 | 9.908572 | | 3 | 0.06 | 0.15 | 0.31 | 6.908978 | 9.938337 | 0.002224 | 0.996406 | | 5 | 0 | 0.01 | 0.14 | - | - | 0 | 0 |
| IV | 5 | 0.01 | 0.09 | 0.33 | - | - | | 4 | 0.01 | 0.08 | 0.23 | 1.815232 | 2.702502 | 0.003505 | 0.957281 | | 3 | 0 | 0.04 | 0.32 | 0.6444 | 0.4040 | 0 | 1 |

**Note: Modelled medians do not sum to 1, however means will (data not shown). Ordering of Zed parameters is critical to ensure correct calculation of probabilities, although this order may not be the same as the logical order (stages I-IV).*

**References**


1. Kidney cancer incidence statistics: Kidney cancer incidence in Europe and worldwide. In: UK CR, ed., 2014.

2. Five-Year Relative Survival by Stage, Adults (Aged 15-99 Years), Former Anglia Cancer Network, 2002-2006. Cancer Research UK, 2014.

3. Shea MW. A proposal for a targeted screening program for renal cancer. Front Oncol. 2013; 3: 207.

4. Rossi SH, Klatte T, Usher-Smith J, et al. Epidemiology and screening for renal cancer. World J Urol. 2018.

5. Beldhuis IE, Streng KW, Ter Maaten JM, et al. Renin-Angiotensin System Inhibition, Worsening Renal Function, and Outcome in Heart Failure Patients With Reduced and Preserved Ejection Fraction: A Meta-Analysis of Published Study Data. Circ Heart Fail. 2017; 10.

6. Filipas D, Spix C, Schulz-Lampel D, et al. Screening for renal cell carcinoma using ultrasonography: a feasibility study. BJU Int. 2003; 91: 595-9.

7. Malaeb BS, Martin DJ, Littooy FN, et al. The utility of screening renal ultrasonography: identifying renal cell carcinoma in an elderly asymptomatic population. BJU Int. 2005; 95: 977-81.

8. Hassan C, Hunink MG, Laghi A, et al. Value-of-information analysis to guide future research in colorectal cancer screening. Radiology. 2009; 253: 745-52.

9. O'Hagan A. Uncertain Judgements : Eliciting Experts' Probabilities.: John Wiley & Sons, 2006.

10. Iglesias CP, Thompson A, Rogowski WH, et al. Reporting Guidelines for the Use of Expert Judgement in Model-Based Economic Evaluations. Pharmacoeconomics. 2016; 34: 1161-72.

11. O'Hagan A. The SHeffield ELicitation Framework (SHELF).

12. Wilson ECF, Usher-Smith JA, Emery J, et al. Expert Elicitation of Multinomial Probabilities for Decision-Analytic Modeling: An Application to Rates of Disease Progression in Undiagnosed and Untreated Melanoma. Value Health. 2018; 21: 669-76.

13. 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: 2363-80.

14. Kroll D, Noser L, Erdem S, et al. Application of the 8th edition of the AJCC yTNM staging system shows improved prognostication in a single center cohort of esophageal carcinomas. Surg Oncol. 2018; 27: 100-05.

15. Mizuma Y, Watanabe Y, Ozasa K, et al. Validity of sonographic screening for the detection of abdominal cancers. J Clin Ultrasound. 2002; 30: 408-15.

16. Utter K, Goldman C, Weiss SA, et al. Treatment Outcomes for Metastatic Melanoma of Unknown Primary in the New Era: A Single-Institution Study and Review of the Literature. Oncology. 2017; 93: 249-58.

17. Rossi SH, Hsu R, Blick C, et al. Meta-analysis of the prevalence of renal cancer detected by abdominal ultrasonography. Br J Surg. 2017; 104: 648-59.

18. Holland C, Cooper Y, Shaw R, et al. Effectiveness and uptake of screening programmes for coronary heart disease and diabetes: a realist review of design components used in interventions. BMJ Open. 2013; 3: e003428.

19. Crosbie PA, Balata H, Evison M, et al. Implementing lung cancer screening: baseline results from a community-based 'Lung Health Check' pilot in deprived areas of Manchester. Thorax. 2018.

20. Warshauer DM, McCarthy SM, Street L, et al. Detection of renal masses: sensitivities and specificities of excretory urography/linear tomography, US, and CT. Radiology. 1988; 169: 363-5.

21. Hackert T, Kienle P, Weitz J, et al. Accuracy of diagnostic laparoscopy for early diagnosis of abdominal complications after cardiac surgery. Surg Endosc. 2003; 17: 1671-4.

22. Jamis-Dow CA, Choyke PL, Jennings SB, et al. Small (< or = 3-cm) renal masses: detection with CT versus US and pathologic correlation. Radiology. 1996; 198: 785-8.

23. Grigore B, Peters J, Hyde C, et al. Methods to elicit probability distributions from experts: a systematic review of reported practice in health technology assessment. Pharmacoeconomics. 2013; 31: 991-1003.

24. Singh GG, Sinner J, Ellis J, et al. Group elicitations yield more consistent, yet more uncertain experts in understanding risks to ecosystem services in New Zealand bays. PLoS One. 2017; 12: e0182233.

25. Speight PM, Palmer S, Moles DR, et al. The cost-effectiveness of screening for oral cancer in primary care. Health Technol Assess. 2006; 10: 1-144, iii-iv.

26. Morgan MG. Use (and abuse) of expert elicitation in support of decision making for public policy. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111: 7176-84.

27. Haakma W, Steuten LM, Bojke L, et al. Belief elicitation to populate health economic models of medical diagnostic devices in development. Appl Health Econ Health Policy. 2014; 12: 327-34.

28. Grigore B, Peters J, Hyde C, et al. A comparison of two methods for expert elicitation in health technology assessments. BMC Med Res Methodol. 2016; 16: 85.

29. Soares MO, Sharples L, Morton A, et al. Experiences of Structured Elicitation for Model-Based Cost-Effectiveness Analyses. Value Health. 2018; 21: 715-23.