APPLICATION OF EXPERTS ELICITATION METHODS IN PROBABILISTIC FORM TO COLLECT RESOURCE USE DATA IN HEALTH CARE. THE CASE OF SKIN CANCER

DOI: 10.26341/issn.2241-4002-2019-sv-9

Vilelmine Carayanni
Department of Public and Community Health, School of Health Sciences, University of West Attica, Greece
bilma@otenet.gr

Christos Kitsos
Department of Informatics and Computer Engineering, School of Engineering, University of West Attica, Greece

Simeon-Constantinos Nisiotis
Department of Statistics, School of Information Sciences and Technology, Athens University of Economics and Business

Abstract

Objective: Expert elicitations methods in probabilistic form are practiced in health economics to help reimbursement decisions, as well as to estimate the global burden of diseases. The purpose of this study is to present a modified Delphi method to estimate the resource use for the management of BRAF-mutated melanoma in Greece.

Methods: A modified Delphi method was used for selected variables. Three experts in the field have completed the questionnaire. Descriptive statistics to summarize the elicited median values and a variety of graphs were used, to evaluate the best fitted individual and consensus distributions. The rate of (D-RS) surprises was successively calculated.

Results: All questions achieved consensus. After the second round and the revisions the rate of surprises was less than 10%, indicating perfect calibration. Tertile method with feedback seems to be an effective method to reduce heuristics and biases.

Conclusions: As data synthesis studies are the most common type of evaluation studies, future work might concentrate on further improvements to these types of elicitation design in order to provide a stronger support to researchers and decision makers and a better allocation of limited resources. In particular, research might investigate strategies to limit the anchoring and adjustment heuristic. The appropriate adjustment will referred in a compact line of thought.

Keywords: *Traditional settlements, protection of natural and cultural environment.*

Introduction

Very few economic evaluations are entirely based on primary research. As data syntheses analyses dominates in the field of health economics because of unreliable, inadequate or unavailable stochastic data, decision analysts use this technique in order to construct decision models. The data synthesis approach uses and synthesizes data from different sources such as literature, experts, consensus panels, and clinical trials.

As policy makers are often confronted with having to make decisions even when evidence is scarce or lacking; delaying the decision until more evidence is available carries a risk of utility loss, and is not always possible, due to legal constraints, or the fact that such evidence may never become available. With the trend towards probabilistic decision-analytic models in HTA, there is a need that the data informing the model parameters is available as distributions.

In such cases, expert opinion can be used to characterize the different types of model uncertainty; this can also be used in value of information analyses to help identify future evidence needs (including the type of study design, sample size) for reducing the decision uncertainty.

An expert elicitation is a method of eliciting subjective probability distributions over some key parameters from experts.

Expert elicitations methods in probabilistic form are practiced in health economics to help reimbursement decisions, as well as to estimate the global burden of diseases such as cancer.

Melanoma is a malignant tumour arising from melanocytes, typically in the skin. Until recently, treatment options for patients with metastatic melanoma were limited, and prognosis was poor, with median survival of less than one year for patients receiving dacarbazine (DTIC) (Mihajlovic et al., 2012). Melanoma was the ninth most common cancer in 2012 in Europe, with 50% approximately of melanomas harbors activating BRAF mutations (Boursault et al., 2013).

The objective of this study is to apply and evaluate modified Delphi method to measure resource use related to advanced BRAF-mutated melanoma.

Materials and Methods

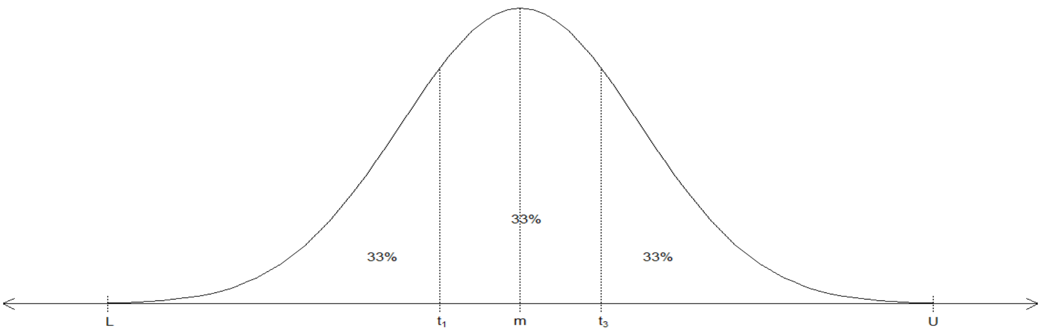
Healthcare resource use (HCRU) related to advanced BRAF-mutated melanoma is based on the information collected by experts on the disease from major Medical Departments around Greece, where 1000 approximately patients with advanced melanoma are treated per year. A modified EFSA Delphi method was used. This corresponds to the first phase of a SHELF elicitation, using the tertile method assessment of individual judgments. A tertile is a division of a set of observations into three defined intervals such that each range contains 33 per cent of the total observation. Each expert should specify their upper and lower tertiles by considering the range from L to U and dividing it into three equally likely intervals. (O'Hagan and Oakley, 2014; CFWI, 2015).

In place of the second phase – the group judgments – Delphi iteration process took place. The experts' judgments and rationales are relayed anonymously back to the experts and they are asked to provide revised judgments**.**

After the two rounds, the experts' individual probability distributions are averaged to provide the final aggregate distribution (CfWI, 2015).

The questions analyzed in this paper are the following:

1. What proportion (%) of patients has had brain metastases since the start of follow-up? (Data 3.1)
2. What proportion (%) of patients has had in transit metastases since the start of follow-up? (Data 3.2)
3. What proportion (%) of patients has had visceral metastases since the start of follow-up? (Data 3.3)
4. What proportion (%) of patients has had bone metastases since the start of follow-up? (Data 3.4)



where L to t_1 the tertile 1, t_1 to t_2 the tertile 2 and t_2 to U the tertile 3

Figure 1 Illustrating elicitation of tertiles

We have used behavioral aggregation. In this approach the experts interact to achieve consensus with the presence of a facilitator in order to ensure that individual and group biases do not detract from the benefits of pooling knowledge and sharing multiple perspectives. (O'Hagan, 2014).

Descriptive statistics to summarize the elicited median values and a variety of graphs were used. For each question, we aggregated individual experts' distributions into a cumulative distribution with equal weights. We computed the arithmetic mean of the elicited percentile values as follows:

$$\alpha_{yk} = \frac{1}{n_k} \sum_{i=1}^{nk} q_{ayik}$$

Where:

i : Expert, $i \in \{1,2,3\}$.

k : Question, $k \in \{1,2,3,4\}$. We use 4 questions to compare modes.

α_y : α^{th} percentile. α_1 is the 5^{th} percentile ($\alpha_1 = 5$), α_2 is the 50^{th} percentile ($\alpha_2 = 50$) and α_3 is the 95^{th} percentile ($\alpha_3 = 95$).

$q_{\alpha_y ik}$: The value of the elicited α_y^{th} percentile for expert i, question k.

The rate of (D-RS) surprises was successively calculated.

We defined a surprise as the event that the observed value that lies outside the 5-95 range (Budescu and Du, 2007):

$$c_{ik} = \begin{cases} 0 & if q_{\alpha_1 ik} < T_k < q_{\alpha_3 ik} \\ 1 & otherwise \end{cases}$$

Where:

T_k : Observed value (true value) for question k.

If the rate of surprises is above 10%, the judgements have a tendency towards overconfidence (Morgan, 2014). R language (package "SHELF") was used.

Results

The process of averaging the density functions is known as the linear opinion pool (with equal weights). We use it in SHELF simply as a benchmark.

Figures 2, 3, 4 and 5 present the individual distribution of each expert as well as the linear pooling. By inspecting the sum of squared differences between elicited and fitted probabilities), we see that the normal distribution fits best for Expert A, and the beta distribution fits best for Expert B and Expert C in the question 1. In the question 2, the normal distribution fits best for Expert B whereas beta fits best for Experts A and B.

Also, in question 3, the beta fits best for all Experts and in question 4, the normal distribution fits best for expert A whereas beta fits best for the other two experts.

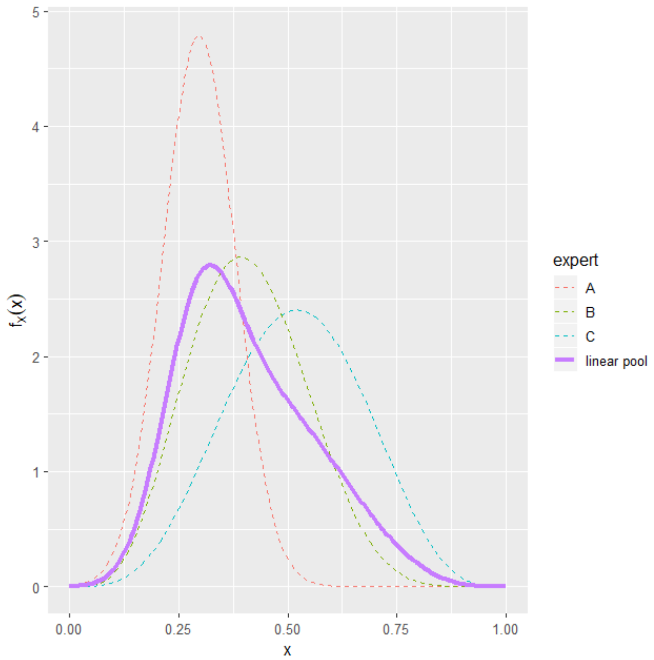


Figure 2 The three fitted distributions (Data 3.1) and an equal-weighted linear pool

	normal	t	gamma	lognormal	logt	beta
expert.A	0.01772982	0.02313336	0.02193681	0.02425278	0.02862050	0.01981696
expert.B	0.05962538	0.06482441	0.05913336	0.05980999	0.06528759	0.05813211
expert.C	0.03429388	0.03901704	0.03543026	0.03623263	0.03876315	0.03290195

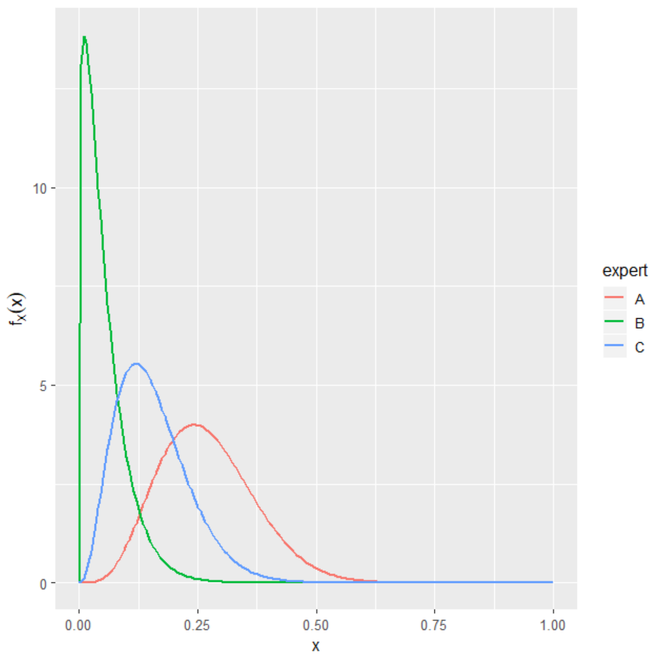


Figure 3 The three fitted distributions (Data 3.2) and an equal-weighted linear pool

	normal	t	gamma	lognormal	logt	beta
expert.A	0.008760332	0.01654749	0.01026118	0.01276256	0.01891287	0.008627447
expert.B	0.032373169	0.04100287	0.03762287	0.04909830	0.05363617	0.036418012
expert.C	0.030184295	0.03925314	0.03092973	0.03465147	0.04168243	0.029548647

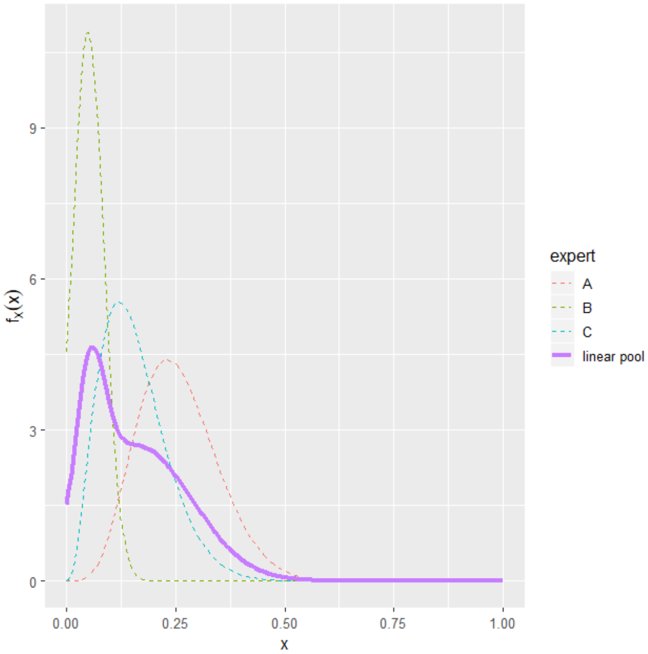


Figure 4 The three fitted distributions (Data 3.3) and an equal-weighted linear pool

	normal	t	gamma	lognormal	logt	beta
expert.A	0.03018429	0.03925314	0.03042662	0.03164643	0.04006583	0.02777849
expert.B	0.05124939	0.05898265	0.05506505	0.05743399	0.06501462	0.03132069
expert.C	0.03465518	0.04305953	0.03531969	0.03584508	0.04443082	0.03232569

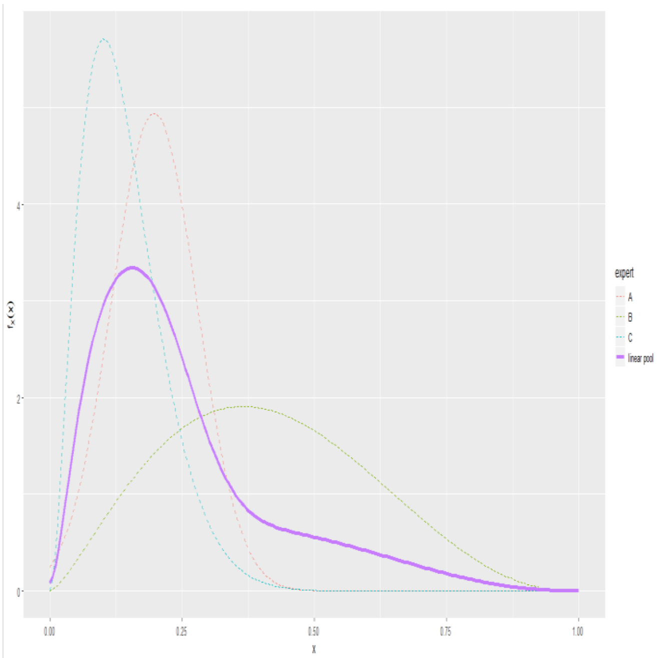


Figure 5 The three fitted distributions (Data 3.4) and an equal-weighted linear pool

	normal	t	gamma	lognormal	logt	beta
expert.A	0.01989472	0.02615544	0.02414562	0.02656792	0.03148572	0.02193545
expert.B	0.04359840	0.05175423	0.04294921	0.04379677	0.05215101	0.04142783
expert.C	0.03717433	0.04256893	0.03803183	0.03882708	0.04186374	0.03561027

EFSA Delphi method

In the second round the group agrees on consensus judgements (EFSA, 2014). Now, we elicit a single 'consensus' distribution from the experts. Experts are invited to revise their original judgements having seen what the other experts think**.** The group judgements are used as a basis for fitting a probability distribution, which is the outcome of the elicitation process, and so must be selected carefully and with full approval of the experts (O'Hagan, 2018).

The final fitted distributions by question are shown in Figures 6, 7, 8 and 9 and the tables compare the three group judgements with the probabilities implied by this distribution.

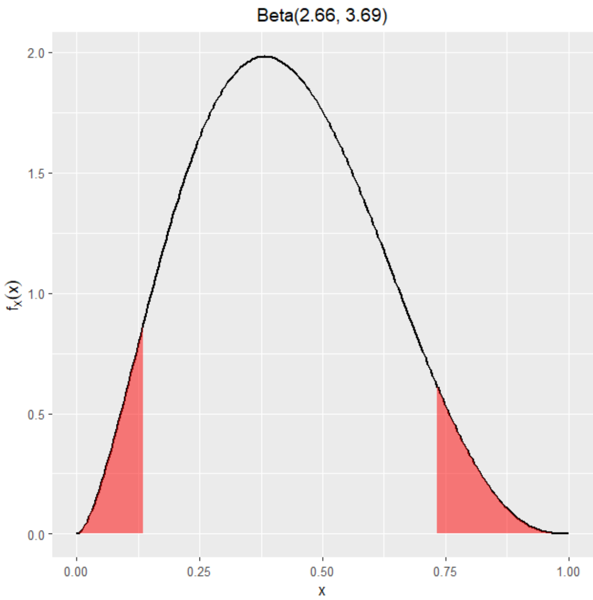


Figure 6 The fitted consensus distribution (Data 3.1)

	fitted probabilities						
	elicited	normal	t	gamma	lognormal	logt	beta
0.33	0.33	0.344	0.344	0.339	0.337	0.337	0.335
0.40	0.50	0.478	0.478	0.485	0.489	0.489	0.490
0.50	0.66	0.670	0.670	0.667	0.666	0.666	0.669

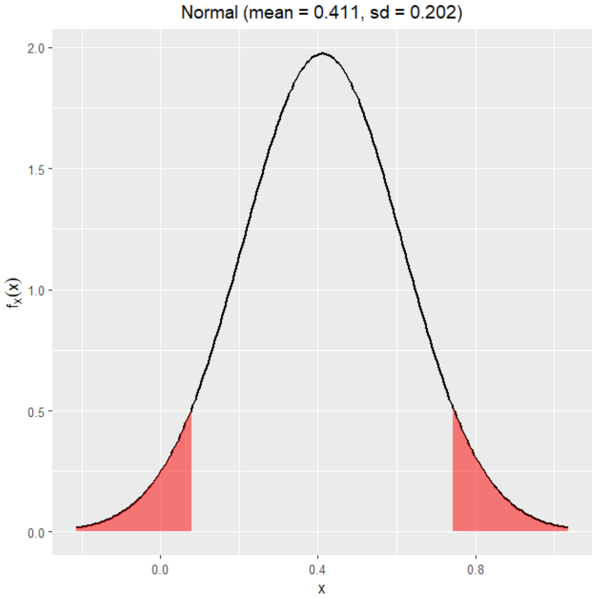


Figure 7 The fitted consensus distribution (Data 3.2)

fitted probabilities							
	elicited	normal	t	gamma	lognormal	logt	beta
0.09	0.33	0.332	0.332	0.324	0.322	0.322	0.325
0.15	0.50	0.497	0.497	0.512	0.518	0.518	0.509
0.21	0.66	0.662	0.662	0.653	0.648	0.648	0.654

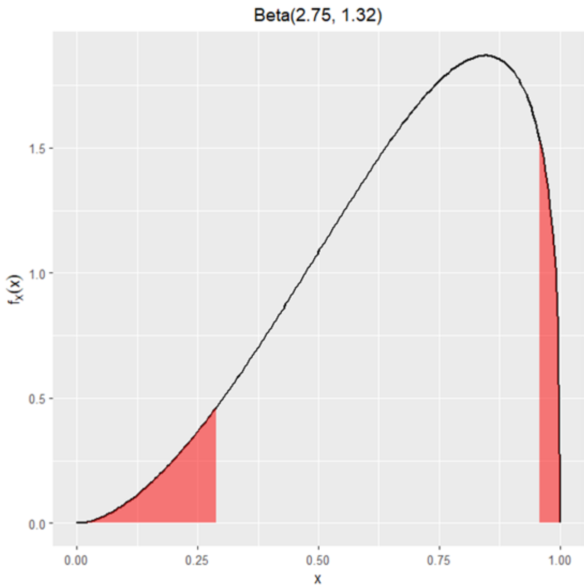


Figure 8 The fitted consensus distribution (Data 3.3)

fitted probabilities							
	elicited	normal	t	gamma	lognormal	logt	beta
0.6	0.33	0.332	0.332	0.329	0.328	0.328	0.336
0.7	0.50	0.497	0.497	0.502	0.504	0.504	0.489
0.8	0.66	0.662	0.662	0.659	0.658	0.658	0.665

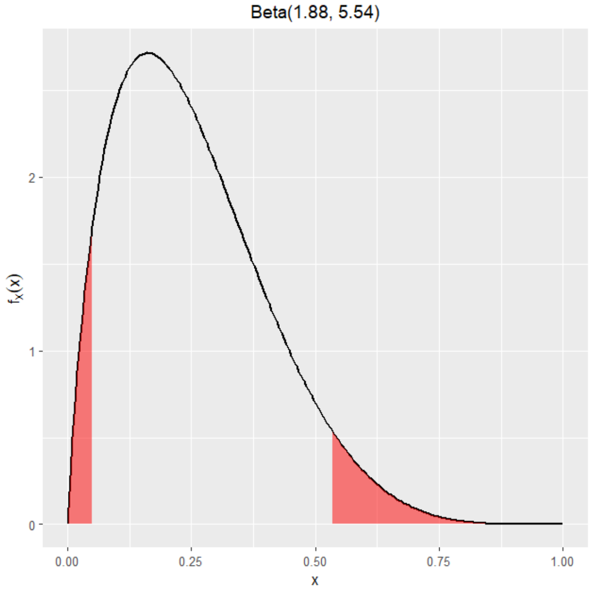


Figure 9 The fitted consensus distribution (Data 3.4)

fitted probabilities							
	elicited	normal	t	gamma	lognormal	logt	beta
0.17	0.33	0.335	0.335	0.329	0.326	0.326	0.330
0.23	0.50	0.492	0.492	0.502	0.507	0.507	0.499
0.30	0.66	0.664	0.664	0.659	0.656	0.656	0.660

Heureistics and Bias

There is evidence that tertiles are elicited more accurately than quartiles, as experts have to double check that the range from t_1 to t_2 is equally likely, as well as from t_1 to m and m to t_2. Therefore, tertiles do not suffer as much from overconfidence and anchoring (O'Hagan and Oakley, 2014). The results of our study are in the same direction: the surprise rate does not surpass 10% after the second round and after revisions of the initial estimates of medians, (D-RS =8%), indicating perfect calibration.

Conclusions

The modified Delphi Method using tertiles with feedback to achive consensus seems to be an effective method to reduce heuristics and biases. Providing people with more feedback seemed to help reducing overconfidence, since feedback serves as a tool allowing people to correct their errors (González-Vallejo and Bonham, 2007).

As data synthesis studies are the most common type of evaluation studies, future work might concentrate on further improvements to these types of elicitation design in order to provide a stronger support to researchers and decision makers and a better allocation of limited resources. In particular, research might investigate strategies to limit the anchoring and adjustment heuristic which can substantially degrade the quality of an economic evaluation in healthcare. The appropriate adjustment will referred in a compact line of thought.

References

Boursault, L., Haddad, V., Vergier, B., Cappellen, D., Verdon, S., Bellocq, J.-P., … Merlio, J.-P. (2013). Tumor Homogeneity between Primary and Metastatic Sites for BRAF Status in Metastatic Melanoma Determined by Immunohistochemical and Molecular Testing. *PLoS ONE*, *8*(8), e70826. https://doi.org/10.1371/journal.pone.0070826

Budescu, D. V., & Du, N. (2007). Coherence and Consistency of Investors' Probability Judgments. *Management Science*, *53*(11), 1731–1744. https://doi.org/10.1287/mnsc.1070.0727

CfWI (2015). "Elicitation methods: applying elicitation methods to robust workforce planning", *Centre for Workforce Intelligence*. Technical Paper Series No. 0011. online.

González-Vallejo, C., & Bonham, A. (2007). Aligning confidence with accuracy: Revisiting the role of feedback. *Acta Psychologica*, *125*(2), 221–239. https://doi.org/10.1016/j.actpsy.2006.07.010

Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment. (2014). *EFSA Journal*, *12*(6), 3734. https://doi.org/10.2903/j.efsa.2014.3734

Mihajlovic, M., Vlajkovic, S., Jovanovic, P., & Stefanovic, V. (2012). Primary mucosal melanomas: A comprehensive review. *International Journal of Clinical and Experimental Pathology*, *5*(8), 739.

Morgan, M. G. (2014). Use (and abuse) of expert elicitation in support of decision making for public policy. *Proceedings of the National Academy of Sciences*, *111*(20), 7176–7184. https://doi.org/10.1073/pnas.1319946111

O'Hagan A., Oakley J. (2014). "SHELF 2.0", online. Downloadable: http://www.tonyohagan.co.uk/shelf/SHELFv2_0.zip

O'Hagan, A. (2014). Eliciting and using expert knowledge in metrology. *Metrologia*, *51*(4), S237–S244. https://doi.org/10.1088/0026-1394/51/4/S237

O'Hagan, A. (2019). Expert Knowledge Elicitation: Subjective but Scientific. *The American Statistician*, *73*(sup1), 69–81. https://doi.org/10.1080/00031305.2018.1518265

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., … Rakow, T. (2006). Uncertain Judgements: Eliciting Experts' Probabilities. In *Uncertain Judgements: Eliciting Experts' Probabilities*. https://doi.org/10.1002/0470033312

Reference to the above article should be made as follows:

Carayanni, V., Kitsos, C., Nisiotis, S-C., (2019). Application of experts elicitation methods in probabilistic form to collect resource use data in health care. The case of skin cancer. *Sustainable Development, Culture, Traditions Journal, Special Volume in Honor of Professor George I. Theodoropoulos*, 82-91. https://doi.org/10.26341/issn.2241-4002-2019-sv-9