EXTERNAL SCIENTIFIC REPORT

Contract OC.EFSA.AMU.2014.03 – CT2

Training Course on Steering an Expert Knowledge Elicitation

Final Report

Andy Hart*, Anthony O'Hagan, John Quigley***, Fergus Bolger******

*Fera Science Limited, **Professor A O'Hagan Ltd.,

***University of Strathclyde, ****Dr Fergus Bolger

ABSTRACT

EFSA's scientific expertise and capacity consists of the members of the Scientific Panels, the Scientific Committee, their Working Groups, and the Authority's own scientific staff as well as the scientists in Member State institutions working with EFSA.

The overall objective of this project was to organize and deliver high quality training courses to meet the needs identified by EFSA to implement Expert Knowledge Elicitation (EKE) approach for quantifying uncertainty in food safety risk assessment.

As outcome of the project a training course was developed on 'Steering an Expert Knowledge Elicitation'. The course covered two working days and was conducted three times during the year 2015. The three courses had 73 participants in total, whereof 17 EFSA experts, 50 EFSA Staff and 6 Network members.

This report contains a summary of the project, a technical description of the training, the final curriculum, the training materials, results from evaluation of the course by the participants, and recommendations for future training on this subject.

KEY WORDS

Expert knowledge elicitation, Probability judgements, Uncertainty, Parameters, Risk assessment

DISCLAIMER

Any enquiries related to this output should be addressed to amu@efsa.europa.eu

Suggested citation: Andy Hart, Anthony O'Hagan, John Quigley, Fergus Bolger, 2015. Training Course on Steering an Expert Knowledge Elicitation. Final Report. EFSA supporting publication 2016:EN-1009, 235pp.

Available online: www.efsa.europa.eu/publications

SUMMARY

EFSA's scientific expertise and capacity consists of the members of the Scientific Panels, the Scientific Committee, their Working Groups, and the Authority's own scientific staff as well as the scientists in Member State institutions working with EFSA.

The overall objective of this project was to organize and deliver high quality training courses to meet the needs identified by EFSA to implement Expert Knowledge Elicitation (EKE) approach for quantifying uncertainty in food safety risk assessment.

As outcome of the project a training course was developed on 'Steering an Expert Knowledge Elicitation'. The course covered two working days and was conducted three times during the year 2015. The three courses had 73 participants in total, whereof 17 EFSA experts, 50 EFSA Staff and 6 Network members.

This report contains a summary of the project, a technical description of the training, the final curriculum, the training materials, results from evaluation of the course by the participants, and recommendations for future training on this subject.

TABLE OF CONTENTS

BACKGROUND AS PROVIDED BY EFSA

In 2011 EFSA requested the Scientific Assessment Support (SAS) Unit, now Assessment and Methodological Support Unit (AMU), that a Working Group on Guidelines for Expert Knowledge Elicitation (EKE) in food and feed safety risk assessment be set (M-2011-0234). The objective was to develop guidance for EFSA-on the use of expert knowledge and by this to complement the EFSA Guidelines on systematic review methodology. In result Guidelines for expert knowledge elicitation in food and feed safety risk assessment were produced, tested in case studies by EFSA, and discussed via public consultation and a workshop.

In June 2014 the Guidance was published. According to the mandate one task of the Guidance was to give practical advice on how to conduct an expert knowledge elicitation in the context of EFSA's risk assessments. The working group considered this by incorporating three concrete protocols into the Guidance. Nevertheless the working group noticed also that a written Guidance alone is not sufficient to put a new methodology into the practice of an institution. The procurements resulting from the present projects should therefore support EFSA in the implementation of the new Guidance by the development of curricula on "Steering an Expert Knowledge Elicitation (EKE)".

It is intended for scientists closely working for EFSA, which shall be realized in a series of in-house trainings.

This contract was awarded by EFSA to:

Contractor: Food and Environment Research Agency (FERA, now Fera Science Limited) with partners Professor A O'Hagan Ltd., University of Strathclyde and Dr. Fergus Mark Innes Bolger (private person).

Contract title: Training courses on Expert Knowledge Elicitation

Contract number: OC.EFSA.AMU.2014.03 – CT2

INTRODUCTION AND OBJECTIVES

EFSA's Guidance on expert knowledge elicitation (EKE) was published in June 2014 (EFSA, 2014). The Guidance defines EKE as "systematic, documented and reviewable process to retrieve expert judgements from a group of experts in the form of a probability distribution."

This project is one of two commissioned by EFSA to develop training for EKE: a web-based training in making probability judgements and (this project) face-to-face training courses in steering an EKE.

Objective

The objective of this project was to develop and conduct three in-house training courses on "Steering an Expert Knowledge Elicitation", especially for experts from the EFSA Scientific Committee, EFSA panels and their Working Groups, as well as EFSA scientific staff.

The purpose of the training courses is:

- to enable the understanding and practical implementation of Expert Knowledge Elicitation amongst Panel/Scientific Committee members, EFSA scientific staff and MS, and
- to strengthen the dissemination of guidance on expert knowledge elicitation amongst Panel/Scientific Committee members and EFSA scientific staff, and promote and facilitate its uptake.

The course is intended primarily for EFSA staff and experts who will be involved in steering EKE studies, i.e. as a member of the 'Steering Group' or 'Elicitation Group'.

Intended learning outcome

EFSA's specification for the project required that, on completing the course, participants shall be able to:

- Explain probabilistic expert judgements
- Recall the characteristics of Expert Knowledge Elicitation (EKE)
- Identify tasks in risk assessment applicable for EKE, e.g. identify priorities for EKE
- Reason the use of EKE in risk assessment
- Frame a problem for EKE
- Identify, select, and motivate experts for an elicitation
- Discuss and select the appropriate elicitation method
- Define the elicitation protocol, incl. realistic resources, adaptations and selection of elicitors
- Produce background information for an elicitation
- Decide on training needs for the experts
- Recall typical protocols using different elicitation methods (evt. software),including the Cooke method, the Delphi method and the Sheffield method
- Document and interpret results; discuss and handle risks of elicitations
- Produce a complete documentation of an EKE
- Discuss handling of confidentiality during an EKE
- Discuss issues of repeatability of an EKE.

EFSA and other guidance documents and opinions

The principal focus for the training was the EFSA Guidance on Expert Knowledge Elicitation (EFSA, 2014).

Examples from six major areas of EFSA's work were used as case studies in some of the practical sessions:

- Chemical risk assessment: bisphenol A (BPA), http://www.efsa.europa.eu/en/efsajournal/pub/3978.htm
- Biohazards: Ebola virus in bushmeat, http://www.efsa.europa.eu/en/efsajournal/doc/3884.pdf
- Nutrition: Dietary Reference Values for cobalamin (vitamin B12), http://www.efsa.europa.eu/en/efsajournal/pub/4150.htm
- Plant health: citrus black spot disease, http://www.efsa.europa.eu/en/efsajournal/pub/3557.htm
- Environmental risk assessment: Exposure of protected species of Lepidoptera to pollen from genetically modified Bt-maize, http://www.efsa.europa.eu/en/efsajournal/pub/4127.htm
- Animal health and welfare: Rift Valley Fever, http://www.efsa.europa.eu/en/efsajournal/doc/3180.pdf

References to additional documents used in the course are listed in the final slide of each lecture (see Training Material, below).

MATERIALS AND METHODS

Summary of project

The project started in December 2014 and ended in December 2015. It was undertaken by a consortium of 4 partners (Fera Science Limited, Professor A O'Hagan Ltd., University of Strathclyde, Dr Fergus Bolger) supported by 7 subcontractors (see Appendix A for a full list of partners and subcontractors and their roles in the project). All contributed to the development of the training materials. Three of the partners led the delivery of the training courses, with one of the subcontractors (Warwick University) providing backup in case of illness or non-availability. The other six subcontractors were experts from different areas of EFSA's work, who contributed to developing the case study materials for the training. All partners and subcontractors attended a rehearsal of the course in May 2015, providing feedback to improve the course design and materials.

The three courses were held in June, August and September of 2015. The course design and materials were further improved after each course, based on detailed feedback from the course participants and from EFSA staff overseeing the project. The final version of the course materials is provided in Appendix B to this report.

Training methodology

The course was designed to provide a balanced mix of lectures with practice-oriented exercises. The practical sessions were divided into individual work, small group work and plenary discussions designed to reinforce the learning from the lectures, link it to the EFSA work area of each participant and provide individual feedback. The course content was delivered in a timetable designed to promote participants' engagement and concentration by alternating different teaching modalities (lectures, practicals, discussion) and by including timely and adequate breaks.

Technical description of training

The final course timetable comprised 14 hours of teaching time in 4 parts, covering the 4 sets of learning objectives specified by EFSA, plus meal and refreshment breaks. It was organised in 4 half day sessions and can be delivered in two full days, or over three days from lunch time to lunch time: the latter option was used for all 3 courses. The final version of the curriculum is shown in the following section.

Participants were asked to make the following preparations before attending the course:

- Identify which of the case studies tailor-made by the Consortium (chemical risk assessment, biohazards, environmental risk assessment, human nutrition, animal health and welfare assessment, plant health assessment) is most relevant to their own area of EFSA work and read a short briefing document on it, to be provided by the course organisers in advance.
- Bring an example from their own area of EFSA to the course: preferably a risk assessment they had recently completed, or one that was currently in progress.
- Make arrangements to be available for the entire duration of the course (e.g. arrange childcare, avoid other commitments).

At the start of the course, participants were provided with a complete set of printed course materials including the course timetable, handouts of all presentations for lectures and practicals, and templates, a spreadsheet and background information needed for the practical exercises and case studies. All the course materials were also made available to participants electronically, and the EFSA EKE Guidance is publicly available on EFSA's website.

Final curriculum

The final version of the curriculum, including improvements based on feedback from the three courses, is shown below.

SESSION I. AFTERNOON OF DAY 1.

PART 1. Problem definition: role of the Working Group

13:30	WELCOME. Course objectives and agenda
13:40	LECTURE 1. Introduction – reasons and roles for the use of EKE in EFSA risk assessments
14:05	PRACTICAL 1. Examples of expert judgement in EFSA's work
14:35	LECTURE 2. Key principles for EKE
14:55	PRACTICAL 2 - plenary. Discussion of key principles
15:10	LECTURE 3. Probabilistic expert judgements
15:35	PRACTICAL 3. Probabilistic expert judgements - work individually
15:55	*Break*
16:25	LECTURE 4. Identifying priority parameters for EKE

17:00	PRACTICAL 4 - breakout groups. Identifying priority parameters for EKE: sensitivity analysis
17:40	PLENARY DISCUSSION Feedback from practical
18:00	SESSION ENDS

HOMEWORK – consider how what you've learned on day 1 would apply to an example assessment from your own area of work.

SESSION II. AFTERNOON OF DAY 1.

PART 2. The pre-elicitation phase: role of the Steering Group

09:00	LECTURE 5. Framing parameters for EKE
09:30	PRACTICAL 5 - breakout groups. Framing problems for EKE
09:55	PLENARY DISCUSSION - report back from breakout groups
10:15	LECTURE 6. Identifying, selecting, motivating and training experts for an elicitation
10:45	*Break*
11:15	PRACTICAL 6 - breakout groups. Identifying, selecting, motivating and training experts for an elicitation
11:40	PLENARY DISCUSSION - report back from breakout groups
12:00	LECTURE 7. The evidence dossier
12:15	LECTURE 8. Sheffield Method
13:00	*Lunch*

SESSION III. AFTERNOON OF DAY 2.

PART 3. The elicitation phase: role of the Elicitation Group

14:00	PRACTICAL 7 - breakout groups. Key aspects of steering the Sheffield method
14:30	LECTURE 9. Delphi Method
15:05	PRACTICAL 8 - breakout groups. Key aspects of steering the Delphi method
15:35	*Break*
16:05	LECTURE 10. Cooke Method
16:50	PRACTICAL 9 - breakout groups. Key aspects of steering the Cooke method
17:20	PLENARY DISCUSSION - report back from breakout groups

18:00 SESSION ENDS

HOMEWORK – consider how what you've learned on day 2 would apply to an example assessment from your own area of work.

SESSION I. MORNING OF DAY 3.

09:00 LECTURE 11. Selecting the appropriate elicitation method

09:25 PRACTICAL 10 - breakout groups. Selecting the appropriate elicitation method

09:55 PLENARY DISCUSSION - report back from breakout groups

10:25 *Break*

PART 4. The post-elicitation phase

10:55 LECTURE 12. Documentation: repeatability, transparency and confidentiality

11:10 LECTURE 13. Advanced topics in EKE

11:35 LECTURE 14. Steering the elicitation process: review of main points

11:55 PRACTICAL 11 - work individually. Planning EKE for examples from each participant's own area of work

12:35 PLENARY DISCUSSION - opportunities and challenges for uptake in participants' work areas

12:55 COURSE EVALUATION QUESTIONNAIRE.

13:00 COURSE ENDS

Course tutors

The course tutors are shown Appendix A.

Training materials

Specific training materials were provided at each course. The training materials included the Course programme, Hand-outs of PowerPoint presentations and Materials for practical exercises. The training material was provided both on paper and electronically by email. The final version of the training materials is provided in Appendix B.

Course attendance certificates

Each participant received a course attendance certificate after the course that included the name of the participant, name of the course, dates of the course and names of the tutors. The certificates were designed by the EFSA staff overseeing the project. In future the course attendance certificates should also include learning outcomes of the course.

RESULTS

Course participation

Recruitment of participants was carried out by EFSA. The three courses had 73 participants in total: 18 in June, 31 in August and 24 in September. The participants comprised 17 EFSA experts, 50 EFSA Staff and 6 EFSA Network members (see Figure 1).

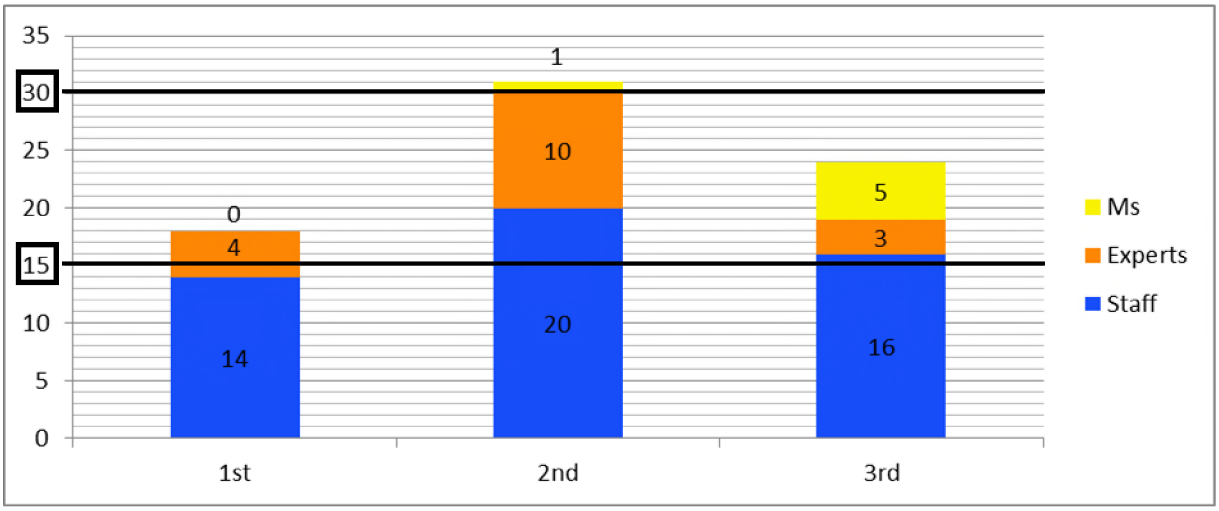


Figure 1. Breakdown of participants by course (Ms = EFSA Network members).

Overall evaluation

Participants were asked by the trainers to complete a detailed questionnaire and leave it behind at the end of the course. This feedback included scores on different aspects of the training, which are summarised below in Table 1. Feedback was optional, and anonymous unless the participant wished to identify themselves in their comments.

The questionnaire also invited participants to offer text comments and/or suggestions for improving the course, which were reviewed in detail by the tutors immediately after each course. Additional feedback was obtained from participants via EFSA's training system and from EFSA staff overseeing the project.

All feedback was taken into account when revising and improving the training design and materials after each course. Overall there was a tendency for scores to increase over the three courses (Table 1). This is thought to reflect the efforts made by the consortium to improve the balance between lectures, practicals and discussion time, and to refine the content in all parts of the course.

Table 1: Summary of participant evaluation of the three courses, obtained via in-course questionnaire. Scores are averages on a scale from 1 (low) to 5 (high). Detailed results including two additional questions (2.2 and 2.7) are in Appendix C.

Question	June	August	September
1.1 Did the course fully meet your expectations and requirements?	4.0	4.3	4.6
1.2 Have you reached the intended learning outcomes of the course?	4.2	4.0	4.5
1.3 Has the course facilitated your future work for EFSA?	4.3	4.0	4.2
2.1 Are you satisfied with the content of the course?	4.1	4.4	4.5
2.3 Are you satisfied with the balance of practical sessions versus lectures?	3.8	4.0	4.2
2.4 Was sufficient time allocated for discussions with fellow participants and tutors?	3.5	3.7	4.1
2.5 Are you satisfied with the teaching ability of the tutors?	4.7	4.7	4.8
2.6 Are you satisfied with the professional and technical competence of the tutors?	4.8	4.8	4.8
3.1 Did the overall organisation and administration associated with the course, prior to and during the training, meet your requirements?	4.5	4.5	4.7
3.2 Did the venue and training facility provided meet your requirements?	4.4	4.5	4.4
3.3 How relevant and user friendly were the training materials/hand outs?	4.5	4.3	4.6
3.4 How suitable was the scheduling, including duration, of the training?	3.9	3.5	4.1

Conclusions and Recommendations

Covering all the topics requested by EFSA within a 2 day course required a full agenda. Several participants commented on this and suggested increasing the proportion of time allowed for discussion. Substantial adjustments were made to moderate the intensity of the course and achieve a good balance between lectures, practicals and discussion. It is understood that if future courses are given on this subject, participants may be required to complete an e-learning module on probability judgements before attending the training. If so, consideration could be given to replacing lectures 2 and 3 of the course with a single shorter lecture, designed to refresh participants' memory of the material from the e-learning. It would be difficult to make further decreases in course time without removing technical content that would be needed by participants when steering an EKE study.

References

EFSA, 2014. Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment. EFSA Journal 2014;12(6):3734. http://www.efsa.europa.eu/en/efsajournal/pub/3734.htm

APPENDICES

Appendix A. List of project partners and subcontractors

Organisation	Role	Personnel and main responsibilities
Fera Science Ltd. (formerly the Food and Environment Research Agency)	Lead Partner	Dr Andy Hart – course developer, tutor and project manager Imogen Foster/Paul Lansell – co-project managers
Professor A O'Hagan Ltd.	Partner	Professor Anthony O'Hagan – course developer and tutor
University of Strathclyde	Partner	Professor John Quigley – course developer and tutor
Fergus Bolger (private person)	Partner	Dr Fergus Bolger – course developer and tutor
Warwick University	Subcontractor	Professor Simon French – course developer and backup tutor (substituted for Professor Quigley in the August course)
Norwegian Institute of Public Health	Subcontractor	Dr Trine Husoy – case study expert, chemical risk assessment
University of Florida	Subcontractor	Professor Dr Ir Arie H Havelaar – case study expert, biohazards
Alterra	Subcontractor	Dr Theo Brock – case study expert, environmental risk assessment
RIVM	Subcontractor	Professor Dr Hans Verhagen – case study expert, human nutrition
Dr Hans-Herman Thulke (private person)	Subcontractor	Dr Hans-Herman Thulke – case study expert, animal health and welfare
Wageningen University	Subcontractor	Dr Wopke van der Werf, Associate Professor – case study expert, plant health

Appendix B. Final version of training materials

PRE-COURSE EMAIL TO PARTICIPANTS

TITLE: Training Course on Steering an Expert Knowledge Elicitation [INSERT DATES HERE]: Agenda and joining instructions

Dear colleagues

Thank you very much for registering for the EFSA training course on Steering an Expert Knowledge Elicitation (EKE), which will be held on [INSERT DATES HERE].

The course times are:

1330-1800 on [INSERT DATE HERE] (PLEASE NOTE START TIME IS 1330)

0900-1300 and 1400-1800 on [INSERT DATE HERE]

0900-1300 on [INSERT DATE HERE]

Please arrive in good time as we have a lot to get through and will start each session promptly. Also, please make any necessary arrangements to enable you to stay until the end of the sessions at 1800 on Monday and Tuesday.

The agenda for the course is attached. We will provide a printed booklet of all the lectures and practicals at the start of the course.

IMPORTANT: Before the course, please DECIDE which of the risk assessment examples you would like to work with during the practical sessions, PRINT a copy of the handout for that example, READ it before the course and BRING it with you to the course. There are 6 examples in all, and the handouts are in the zip file attached to this email. Here's a list of the practical examples:

- Chemical risk – dermal exposure to bisphenol A
- Nutrition – Vitamin B12 requirement
- Environmental risk – GM pollen transport
- Plant health – Citrus Black Spot entry pathway
- Animal Health – Rift Valley Fever
- Biohazard – Ebola in bushmeat

The handouts include links to the original EFSA assessments for these examples, but it is not necessary for you to look at those.

ALSO – please identify another example of a risk assessment in your own area of work, which you are personally familiar with. Ideally this would be a current assessment, or one you are about to start, but previous assessments are also okay. If you are new to EFSA, ask a colleague to suggest a relevant example. It is important that your example includes at least some quantitative aspects (e.g. an exposure calculation, a no-effect level, etc.) Please bring to the course (electronically or on paper) any documentation you need to remind yourself of the details of your example.

The only other preparation we ask you to make is to BRING A LAPTOP with you if possible, with internet access. This will be needed in some of the practicals. If this is not possible for you, you will need to work with another participant who has a laptop during those practicals.

The course is based on the EFSA EKE Guidance Document, which you can find at the following link:

http://www.efsa.europa.eu/en/efsajournal/pub/3734.htm

There is no need for you to read the Guidance before the course (it is 278 pages!). If you have a printed copy it might be useful to bring it with you, but this also is not essential. If you have read it and have any questions about it, we will be happy to discuss them during the course.

Your tutors for this course will be [INSERT NAMES AND AFFILIATIONS HERE].

We look forward to meeting you at the start of the course – please arrive promptly! If anyone has any queries before then, please email me.

Best wishes

[INSERT NAME AND CONTACT DETAILS HERE]

ATTACHMENTS:

[ATTACH COURSE AGENDA HERE]

[ATTACH ZIP FILE OF CASE STUDY HANDOUTS HERE]

Chemical Risk Example

- EFSA's risk assessment of bisphenol A (BPA)
- http://www.efsa.europa.eu/en/efsajournal/pub/3978.htm
- **Short title**: Bisphenol A

The following slides contain a simplified assessment for use only as training example for the EFSA Training Course on Steering an Expert Knowledge Elicitation

Description of Risk Assessment

- EFSA was asked to:
  - evaluate the toxicity of BPA for humans, including specific (vulnerable) groups of the population
  - carry out an exposure assessment from dietary sources and non-dietary sources, and characterise the human health risks of BPA taking into account specific groups of the population
- One of the non-dietary sources was exposure to BPA in thermal paper used for receipts.
- In the toxicity assessment, available studies on different types of effects were considered, including effects on the mammary gland.
- Information on toxicokinetics was used to convert dosing in rodents to equivalent human oral doses.

List of parameters

Parameter name	Description	Units
Number of fingers ($n_{fingers}$)	Average number of fingers that touch receipts during handling	fingers
Number of receipts ($n_{receipts}$)	Number of thermal paper receipts a person handles each day	receipts/day
Skin deposit (d_{skin})	The amount of BPA that remains on the skin after touching thermal paper	µg BPA per finger
Skin penetration (p_{skin})	The percentage of BPA on the skin that is absorbed and enters the body and the bloodstream	%

Model

- We consider the *amount of BPA absorbed into the body* from dermal exposure to thermal paper receipts ($E_{dermal.tp}$), in µg BPA/kg bw/day.
- Model:

$$E_{dermal.tp} = \frac{d_{skin} \times n_{fingers} \times n_{receipts} \times p_{skin}}{100 \times bw}$$

- *bw* is set to the EFSA default for adults (70kg)
- 100 is required because p_{skin} is a percentage
- We need to prioritise EKE for d_{skin}, $n_{fingers}$, $n_{receipts}$, p_{skin}

Number of fingers

- There is no evidence or documentation on how many fingers people use to handle receipts
- It is anticipated that the handling of receipts can be very different from person to person
- As most people have 10 fingers (including thumbs), the actual number of fingers that handle individual receipts can be between 2 and 10, but the number required here is the average
- Thermal paper has BPA only on the front (printed) side
- Based on an experiment where 2 persons with inked fingers handled 4 receipts and the area of ink on the receipts was measured, a Danish study states that 'typically approx. 10 cm^2 of the finger pads (on 8 fingers) will be in contact with receipt when it is checked and folded with the front side turning out'

Number of receipts

- There was only one study, from Denmark in 2011, providing data relevant to this parameter.
- This study estimated the number of receipts used in Denmark by two methods:
  - 1220 million thermal paper receipts/year, based on numbers of credit card transactions reported by a large Danish supermarket, the total number of payment cards in Denmark, and estimates of the proportion of receipts using thermal paper
  - 1355-1627 million BPA receipts/year, based on a supplier's estimate of total tonnage of thermal paper delivered, the average weight of 47 sample cash register receipts, and assuming 75% by weight contains BPA.
- The Danish population above age 12 was stated as 4.7 million
- It was roughly assumed that consumers who carefully check the receipts and keep them, on average handles each one 2.5 times
- BPA is also used in thermal paper used for other purposes, e.g. library receipts, queue tickets, labels, parking tickets, boarding passes, etc.
- Based on the above, the Danish study estimated that:
  - The number of BPA-containing receipts per consumer per is 0.7 per day
  - The number handled per day is 1.8 per day (0.7x2.5)
  - Women with children at an age where they do not shop themselves handle about double this number, i.e. 3.6 per day
  - Other uses of thermal paper are roughly assumed to add 1 receipt/ticket per day
  - Different conditions may apply in other EU countries

Skin deposit

- One study available
- Factors affecting the amount of BPA deposited include:
  - duration and pressure of finger contact
  - sweating, skin hydration, oiliness of fingers
  - variable structure and BPA content of thermal papers

'Standard' conditions: dry, slightly greasy fingers, firm pressure, 5 seconds		
Paper source	BPA in paper (g/kg)	Mean per finger, µg (n)
Lab recorder 1	11	0.13 (4)
Lab recorder 2	16	0.6 (4)
Canteen	17	3.3 (2)
Shop 4	17	0.5 (2)
Shop 7	15	1.1 (2)

Effect of finger condition*	
Finger condition	Mean per finger, µg (n)
'Standard'	0.6 (6)
Dry, clean	0.4 (4)
Humid	8.8 (2)
Wet	20.5 (2)
Oily	5.8 (4)

*Paper from Recorder 2 and Shop 4

Effect of holding behaviour*	
Behaviour	Mean per finger, µg (n)
Holding 1 sec	0.2 (2)
Holding 5 sec	0.6 (4)
Low pressure 5s	0.2 (2)
Wipe 5s	0.4 (2)
Holding 60 sec	0.6 (2)
3 new contacts	0.5 (2)
10 new contacts	0.5 (2)

*Paper from Recorder 2

Skin penetration

- Estimates vary considerably between studies (summarised from EFSA opinion)
- It is assumed that the rate constant *k* follows first-order kinetics, but there is no data on this
- Skin in the hands is thicker than the skin used in the experiments

	Study 1	Study 2	Study 4	Study 5
Skin type	pig skin from the flanks	human skin samples from breast surgery	human skin explants from abdominal region	dorsal part of the upper leg from 2 human cadavers
Number of skin sections	6 (?)	11	3	7
Skin viability		non-viable	viable skin	non-viable
Skin Section thickness		800–1000 µm	500 µm	200 µm
Applied surface density		259 µg/cm^2	2.75 µg/cm^2	1.82 µg/cm^2
Temperature	32.0 ± 0.1 °C	≈32 °C	37 °C	30–32 °C
Method	static Franz diffusion cell	static Franz diffusion cell OECD TG 428	organ culture in Transwell cell culture inserts	flow-through Franz cell OECD TG 428
Duration of incubation	24 h	48 h	72 h	24 h
Recovery	84.3 ± 9.0 % at 10 h	82.1 %	92.6 ± 5.8 %	101.5 ± 1.6 %
Skin deposition		24.6 ± 5.8 %	41.5 ± 10.8 %	35.5 ± 6.6 %
Percutaneous penetration (mean ± SD)	**4.1 % at 24 h**	**13.0 ± 5.4 %**	**45.6 ± 6.2 % at 72 h (15.2% when down-scaled to 24 h)**	**8.6 ± 2.1 %**

Your estimates

Parameter name	m	s	Units
Number of fingers ($n_{fingers}$)			fingers
Number of receipts ($n_{receipts}$)			receipts/day
Skin deposit (d_{skin})			µg BPA per finger
Skin penetration (p_{skin})			%

m: approximate central estimate
s: approximate standard deviation
These are for use in sensitivity analysis and will be explained at the course (Lecture 4)

Nutrition example

- Scientific Opinion on Dietary Reference Values for cobalamin (vitamin B12) http://www.efsa.europa.eu/en/efsajournal/pub/4150.htm
- **Short title:** Vitamin B12

The following slides contain a simplified assessment for use only as training example for the EFSA Training Course on Steering an Expert Knowledge Elicitation

Description of Risk Assessment

- EFSA's NDA Panel provides advice on dietary reference values (DRVs) for nutrients, micronutrients and other essential substances
- Vitamin B12 is the generic descriptor for compounds exhibiting qualitatively the biological activity of cobalamin
- Cobalamin/vitamin B12 deficiency can lead to various health problems including megaloblastic anaemia and neurological dysfunction.
- The Panel considered the biomarkers serum cobalamin, holotranscobalamin, methylmalonic acid plus plasma total homocysteine as most suitable to derive DRVs for cobalamin.
- Due to limited data an Average Requirement (AR) and Population Reference Intake (PRI, intake meeting the requirements of 97.5% of individuals in the population) cannot be determined.
- Instead, the Panel derived Adequate Intake (AI) values for cobalamin. An Adequate Intake is the average observed daily level of intake by a population group (or groups) of apparently healthy people that is assumed to be adequate. For practical purposes the AI is treated similarly to a PRI, though considered less 'firm'.

List of parameters

Parameter name	Symbol	Description	Units
Adult Adequate Intake	AI_{adult}	Adequate Intake for adults	µg/day
Infant body weight	bw_{infant}	Body weight of infants	kg
Adult body weight	bw_{adult}	Body weight of adults	kg
Scaling factor	SF	Exponent for estimating the ratio of metabolic requirements from the ratio of body weights, for mature organisms.	dimensionless
Growth factor	GF	Additional daily intake of cobalamin required by an EU infant aged 7-11 months compared to an EU adult, as a proportion of the amount that would be required for the infant based on scaling with body weight alone.	dimensionless

Lot 2 Model

$$AI_{infant} = AI_{adult} \times \left(\frac{bw_{infant}}{bw_{adult}}\right)^{SF} \times (1 + GF)$$

- SF = allometric scaling factor
- GF = growth factor
- For cobalamin, AI_{infant} is estimated from AI_{adult} by allometric scaling, on the assumption that cobalamin requirement is related to metabolically active body mass.
- When scaling from adult to infant, a growth factor GF is included to account for the nutrient required to be incorporated into newly-formed tissue.

Adult Adequate Intake

- There are limited data on relationships between cobalamin intake and biomarkers of cobalamin status, and uncertainty about the cut-off values for cobalamin insufficiency.
- However, there is consistent evidence from observational and intervention studies that a cobalamin intake of 4 µg/day and greater is associated with biomarker levels indicative of adequate cobalamin status.
- Dietary intake of cobalamin was estimated using consumption data for EU countries from the EFSA Comprehensive Database.
  - Estimated averages (µg/day) for *males aged 18-65y* were 5.3 (UK), 5.8 (NL), 6.2 (Italy), 6.4 (Ireland), 6.8 (France & Finland), 8.2 (Sweden).
  - Estimated averages (µg/day) for *females aged 18-65y* were 4.3 (UK), 4.4 (NL & Ireland), 4.9 (Finland), 5.1 (Italy), 5.2 (France & Latvia), 6.1 (Sweden).
  - Averages for *adults over 65y* ranged from 4.2 (women, Italy) to 8.6 (men, Sweden) (µg/day)

Infant body weight

- The Panel based its assessment on weight-for-age values from WHO.
- WHO used a combination of longitudinal and cross-sectional data on 8440 infants and children in Brazil, Norway, Oman, USA and 'affluent neighbourhoods' in Ghana and India.

Age (months)	Median boys, kg	Median girls, kg
7	8.3	7.6
8	8.6	7.9
9	8.9	8.2
10	9.2	8.5
11	9.4	8.7

Adult body weight

- The Panel based its assessment on weights calculated from heights assuming a Body Mass Index of 22 kg/m^2 (right hand column)

Table 4: Median of measured body heights and body masses of 16,500 men and 19,969 women in 13 EU Member States[10] compared to body masses calculated for a BMI of 22 kg/m^2

Age (years)	n	Measured body height (cm) Median	Measured body mass [(a)] (kg) Median	Body mass [(b)] (kg) assuming a BMI of 22 kg/m^2 Median
Men				
18 - 29	2,771	178	75.0	69.7
30 - 39	2,971	178	82.0	69.7
40 - 49	3,780	177	82.0	68.5
50 - 59	3,575	175	82.0	67.4
60 - 69	2,611	174	80.0	66.4
70 - 79	792	172	80.0	65.1
Women				
18 - 29	3,589	164	60.0	59.4
30 - 39	3,866	164	63.8	59.2
40 - 49	4,727	163	66.0	58.5
50 - 59	4,066	162	68.0	57.7
60 - 69	2,806	160	67.0	56.3
70 - 79	915	159	63.5	55.6

(a): n values for this variable slightly differ.
(b): Body masses calculated for individual measured body heights assuming a BMI of 22 kg/m^2.

[10] Data from Bulgaria, Czech Republic, Finland, France, Germany, Ireland, Luxembourg, Poland, Portugal, Slovakia, Spain, The Netherlands, UK

Scaling factor

- As the requirement for cobalamin is associated with metabolic rate, allometric scaling is performed.
- Kleiber (1932,1947) predicted that requirement for nutrients should be proportional to the metabolic body weight for mature organisms at rest in indifferent environmental temperatures.
- If the ratio of weights is R then the average ratio of metabolic requirements is R raised to the power of the allometric scaling factor.
- This rule has never directly been proven with respect to nutrient requirements and discussion whether the scaling factor is nearer 0.67 or nearer 0.75 has not yet been conclusively resolved. Scaling as the 0.75 power of body mass has been widely accepted in nutritional science.
- If requirement for the nutrient is not associated with metabolic rate, isometric scaling is performed (scaling factor = 1). This has been used for e.g. magnesium and fluoride.

Growth factor

- When scaling down from adult to infants, corrections for growth requirements have to be made in order to account for the nutrient amount required to be incorporated into newly formed tissue.
- One way to do this is to add an age-specific growth factor based on the proportional increase in protein requirements for growth. These can be applied to either isometric or allometric scaling.
- If the requirement based on scaling with body weight is D then the average additional requirement for an infant is D times the growth factor.
- The Panel's assessment for the growth factor was based on the following estimates from WHO/FAO/UNU, which apply to both boys and girls:

Age (years)	0.5	1	2	3
Calculated growth factor	0.70	0.44	0.20	0.11

Your estimates

Parameter name	Symbol	m	s	Units
Adult Adequate Intake	AI_{adult}			µg/day
Infant body weight	bw_{infant}			kg
Adult body weight	bw_{adult}			kg
Scaling factor	SF			dimensionless
Growth factor	GF			dimensionless

m: approximate central estimate
s: approximate standard deviation
These are for use in sensitivity analysis and will be explained at the course (Lecture 4)

Environmental risk example

- Exposure of protected species of Lepidoptera to pollen from genetically modified Bt-maize
- Link: http://www.efsa.europa.eu/en/efsajournal/pub/4127.htm
- **Short title**: GM maize

The following slides contain a simplified assessment for use only as training example for the EFSA Training Course on Steering an Expert Knowledge Elicitation

Description of Risk Assessment (GM maize)

- Bt-maize, expressing insecticidal protein of *Bacillus thuringiensis,* is a genetically modified crop grown commercially in the EU.
- Pollen of Bt-maize may be transported to off-field habitats in which protected Lepidoptera (e.g. butterflies) occur.
- EFSA has used mathematical modelling to assess the risk associated with the ingestion of Bt-maize pollen by non-target larvae of Lepidoptera.
- EFSA recently updated its assessment to take account of new data on dispersal of maize pollen over long distances.
- At the same time, EFSA identified additional factors that modify exposure of non-target Lepidoptera. Four such factors are considered in this exercise.

List of parameters

Parameter name	Symbol	Description	Units
Proportion Bt	f1	Proportion of maize which is Bt-maize	dimensionless
Vegetation structure	f2	Ratio of average pollen deposition on leaf surfaces to pollen deposition on a one-sided horizontal surface in the same conditions	dimensionless
Wind and rain	f3	Effect of wind and rain on pollen concentrations, expressed as the *ratio* of average pollen concentration encountered by non-target Lepidoptera of conservation concern when foraging *to* the concentration originally deposited	dimensionless
Degradation	f4	Degradation of Bt-protein in pollen, expressed as the *ratio* of average Bt concentration in pollen encountered by non-target Lepidoptera of conservation concern when foraging *to* the average concentration of Bt in the pollen when originally deposited	dimensionless

Model

$$F = f1 \times f2 \times f3 \times f4$$

- F = multiplicative effect of modifying factors on exposure of larvae of non-target Lepidoptera of conservation concern to Bt maize pollen, in protected habitats at specified distance from the nearest Bt maize field.
- The EFSA 2015 opinion models the effect of eight modifying factors, but for this exercise we focus on four factors (f1 to f4), which are assumed to act independently.
- Exposure varies spatially, temporally and between larvae but it is assumed that the factors f1 to f4 operate equally on all exposures.
- The factors are applied to conservative estimates of exposure that are expected to be over-estimates. Therefore the factors are expected to take values less than one, although the possibility of increases in exposure (factors >1) should be considered if appropriate.

Proportion Bt

- Estimates of exposure assume that all maize pollen deposited in the protected habitat is derived from Bt-maize.
- It is not expected that all maize fields will be planted with Bt-maize. To allow for this it is desired to estimate what proportion of maize fields will be Bt-maize.
- One study reports that the proportion of Bt-maize pollen collected by pollen samplers ranges between 7 and 44% at distances between 5 m and 120 m from a single Bt-maize field.
- There is a requirement that a minimum of 20% of the maize area should be cropped with non-Bt-maize as refuge for insect resistance management.

Vegetation structure

- Exposure assessments estimate pollen deposition assuming a one-sided horizontal surface.
- However, vegetation has a three-dimensional structure. To allow for this, it is proposed to apply a factor extrapolating from deposition on a 1-sided surface to deposition on leaf surfaces in three dimensional vegetation.
- Vegetation structure can be described by the leaf area index (LAI), defined as the one-sided green leaf area per unit ground surface area.
- The LAI varies between vegetation types. In the literature, values of between 2.5 and 8 were measured in crops, 1.6 to 13 for grasslands, 0.5 to 0.8 for rural areas and up to 19 for forests.

Wind and rain

- Pollen on leaf surfaces can be removed by wind and rain.
- Pollen on leaf surfaces can be displaced by rain and wind, leading to accumulation on lower leaves, or on leaf veins and leaf axils. This could lead to a higher exposure of larvae of those species that feed on lower leaves.
- In most cases, larvae do not prefer to feed on leaf veins and leaf axils; this could lead to a lower exposure of larvae of those species.
- Existing exposure estimates ignore these effects of wind and rain. It is proposed to allow for them by applying a multiplicative factor that modifies exposure upwards or downwards to an appropriate degree.

Degradation

- Data on degradation of Bt-protein in pollen is scarce.
- One study examined effect of ultra-violet light on degradation of Bt-protein in maize pollen. There was no significant difference between concentrations in pollen exposed to UV light and control (unexposed) pollen. However, concentrations in both UV-treated and control pollen reduced by about 50% during the 240 hours of the experiment (statistically significant, $P = 0.001$). Flaws in this study include limited data on expression of Bt-protein in the pollen compared with other studies.
- In a different study, it was observed that Bt-proteins in maize pollen were not detectable after 15 or 18 days, depending on the type of Bt-maize (different genetic modifications).

Your estimates

Parameter name	Symbol	m	s	Units
Proportion Bt	f1			dimensionless
Vegetation structure	f2			dimensionless
Wind and rain	f3			dimensionless
Degradation	f4			dimensionless

m: approximate central estimate
s: approximate standard deviation
These are for use in sensitivity analysis and will be explained at the course (Lecture 4)

Plant health example

- Risk of citrus black spot for the EU territory
- Web link: http://www.efsa.europa.eu/en/efsajournal/pub/3557.htm
- **Short title**: Citrus black spot disease

The following slides contain a simplified assessment for use only as training example for the EFSA Training Course on Steering an Expert Knowledge Elicitation

Description of Risk Assessment

- The European Commission asked EFSA in 2013 to make a risk assessment of the fungus *Phyllosticta citricarpa*, causal agent of a plant disease called "citrus black spot" (CBS). *Phyllosticta citricarpa* can be carried on living plants and on citrus fruit, either in trade or with passenger traffic.
- EFSA was asked to identify risk reduction options and to evaluate their effectiveness in reducing the risk posed by CBS to European citrus.
- EFSA was also asked to carry out an evaluation of the effectiveness of the present EU requirements in reducing the risk of introduction of *P. citricarpa* into the EU.
- Europe's regulations with respect to CBS are contested by South Africa, an important citrus exporter. South African experts contend that Europe is not at risk because fruit would not be a pathway for entry of CBS and because Europe's climate would not be conducive to the disease.

List of parameters

Parameter name	Symbol	Description	Units
Tonnes imported	V	Tonnes of citrus imported	tonnes
Weight per fruit	W	Weight of a single citrus fruit	kg
Proportion infected	$P_{infected}$	Proportion of fruits which are infected	dimensionless
Proportion at packing houses	$P_{packing}$	Proportion of fruits sent to packing houses	dimensionless
Proportion exposed	$P_{exposed}$	Proportion of fruits exposed at packing houses	dimensionless
Transfer to citrus	$P_{transfer}$	Proportion of infected fruits with spores transferring to citrus	dimensionless

Model

$$I = \frac{1000 \times V}{W} \times P_{infected} \times P_{packing} \times P_{exposed} \times P_{transfer}$$

- I is the number of infected fruits exposed at Spanish packing houses from which *Phyllosticta citricarpa* spores reach citrus trees
- EFSA based its assessment in part on a quantitative model of the entry pathway into Spain. For this exercise, we use a simplified version of the model, in which regional distribution within Spain is ignored.
- Transfer to citrus was omitted in the EFSA model but is included here.
- The model considers only fruit originating in *countries where P. citricarpa was present in 2015*, defined as Argentina, Australia, Bhutan, Brazil, China, Cuba, Ghana, Indonesia, Kenya, Mozambique, Philippines, South Africa Taiwan, Uganda, the United States, Uruguay, Vanuatu, Zambia and Zimbabwe

Tonnes imported

- Data on import of citrus (except limes) to Spain from CBS-affected countries were extracted from the Eurostat Comext database.
- The median of the annual data from 2007-2011 was 142000 tonnes with a range of 111000 to 192000 tonnes.
- Data for the last three years were all close to the minimum as a result of the economic recession, but the fact that they were close together may suggest that further reduction is unlikely and that imports may return to normal variation as the economy recovers.
- Comext data exclude intra-EU trade, which is described as 'considerable'. Some EU Member States, particularly the Netherlands and the UK, import citrus fruit from CBS-affected third countries and then re-export the fruit within the EU.

Weight per fruit

- Data on unit weights of fruit are used when assessing human exposure to pesticides in food.
- The EFSA Pesticide Residue Intake Model (PRIMo), which is used for these assessments, includes the following:
  - Grapefruit 160g
  - Oranges 160g
  - Lemons 71.8g
  - Limes 67g
  - Mandarins 100g
- The above values are stated in the model to be average unit weights for the edible portion, whereas the model for CBS requires whole fruit weights.

Proportion infected

- EFSA (2015) used two lines of reasoning to assess the likely level of infection with CBS of citrus fruit under current regulations.
- The first line of reasoning gave an estimate of 10 infected fruit per million with a range of 3-35 infected fruit per million. This was based on:
  - A meta-analysis of average infection level in trials in affected countries after the most effective fungicide regimes (2% with 95% CI 0.6 – 7%)
  - Inspection in the country of origin reducing infection by a factor of 100
  - Inspection at the EU border removing badly infested consignments, reducing overall infection by a further factor of 20
  - An implied assumption that infection level varies considerably between consignments
- The second line of reasoning gave an estimate of 7.67 per million with a range of 2.19 – 26.9 per million, based on:
  - Data on inspections of citrus entering the Netherlands in 2012-2013, showing 100 interceptions in 36729 lots
  - Estimation of the efficiency of the inspection procedure, using a Poisson approximation and assuming the proportion infected is constant between lots

Proportion at packing houses

- Fruit is imported mostly by boat, and transported by road in trucks to packing houses, distribution centres for retail and food processing industries.
- Based on a personal communication from a citrus expert at the Valencian Institute of Agrarian Research (IVIA), the allocation of citrus between destinations was estimated as follows:
  - Packing houses: 40 % (plausible range 30–50 %)
  - Retail: 40 % (plausible range 30–50 %)
  - Food processing, predominantly juice making: 20 % (plausible range 15–25 %).

Proportion exposed

- Packing houses receive fruit and repack it before forwarding it to distribution centres for retail. Packing houses process fruit to ensure it fulfils quality regulations imposed by the EU and by retail companies.
- Packing houses purchase fruit, at the quality standard they require, during the season and then apply further checks during the packing process.
- Packing houses produce waste, but they select not specifically for spots on the peel, such as those produced by *P. citricarpa*, but for major blemishes and bruises. Data from FAO and WRAP indicate a waste fraction of 3 % in the grading process followed by a further 0.1–0.5 % in the packing process, with the total loss being quoted as up to 4 %.
- The waste from packing houses is usually mixed with rotten fruit so it cannot be used for juicing. Instead, it is stored in open containers, generally under cover, until full and then spread outdoors in open-air facilities for solar drying.

Transfer to citrus

- All packing houses in Spain are located in the citrus-growing areas because they are associated with local fruit production. Consequently, packing houses are in close proximity to the citrus orchards, with distances between the waste and the nearest citrus trees often in the order of metres.
- Experiments with sweet orange fruit showed that fruit misted to simulate light rainfall continue to exude *P. citricarpa* pycnidiospores from pycnidia for at least one hour.
- In still air conditions, 99.4 % of the splashes produced by single incident rain drop on the fruit were of less than 2 mm diameter, with an average of 1–21 pycnidiospores. Larger but less frequent splashes of 4–5.5 mm diameter contained an average of 308 pycnidiospores.
- In these experiments, the maximum horizontal distance of splash was 70 cm and the maximum height was 47.4 cm. However, when multiple incident rain drops were combined also in still air, splashes were forced higher than occurred in single-drop experiments to over 60 cm.
- In another experiment combining single incident rain drops and wind, splashes from infected fruit were disseminated up to two metres downwind from the target fruit with a 4 m/s wind speed and up to eight metres at a wind speed of 7 m/s, the highest wind speed evaluated, reaching heights up to 75 cm and even higher as a result of fine droplets becoming aerosolised.
- When the rain is combined with a moderate wind (7 m/s), the pathogen can be dispersed at least eight metres downwind from the infected fruit to heights of at least 75 cm. Such conditions occur about 0.5-1% of time over the year in Spanish regions where packing houses occur.
- If rain is combined with stronger wind, small aerosolised droplets formed by a rain splash are expected to be dispersed much further. A study of dispersal of citrus canker in Florida, found that rain-splashed pathogens can travel several kilometres.

Your estimates

Parameter name	Symbol	m	s	Units
Tonnes imported	V			tonnes
Weight per fruit	W			kg
Proportion infected	$P_{infected}$			dimensionless
Proportion at packing houses	$P_{packing}$			dimensionless
Proportion exposed	$P_{exposed}$			dimensionless
Transfer to citrus	$P_{transfer}$			dimensionless

m: approximate central estimate
s: approximate standard deviation
These are for use in sensitivity analysis and will be explained at the course (Lecture 4)

Animal health example

- Scientific Opinion on Rift Valley Fever
- http://www.efsa.europa.eu/en/efsajournal/doc/3180.pdf
- **Short title:** Rift Valley Fever

The following slides contain a simplified assessment for use only as training example for the EFSA Training Course on Steering an Expert Knowledge Elicitation

Description of Risk Assessment

- EFSA was asked to deliver a scientific opinion on the risk of Rift Valley fever (RVF). One of the terms of reference was to assess the risk of introduction of Rift Valley fever virus (RVFV) into the region of concern (RC) through the movements of live animals from countries in East and West Africa where it is endemic. Officially, movements of such animals are banned, and the extent of illegal movements is very uncertain
- The RC was defined as comprising Morocco, Algeria, Tunisia, Libya, Jordan, Israel, the Palestinian Territories, Lebanon and Syria.
- RVF is a disease of cattle, sheep, goats and camels. The virus is transmitted by mosquitoes.
- RVF is typically a mild infection of adult animals, the primary effect and symptom being abortions in pregnant animals.

List of parameters

Parameter name	Symbol	Description	Units
Volume	v	Number of animals to be transported from endemic countries to the Region of Concern.	animals
Prevalence	p	Prevalence of Rift Valley Fever Virus in animals to be transported.	dimensionless
Transport	t	Change of infection during transport, expressed as a ratio. Animals may become non-infectious during transport, but there may also be reinfection.	dimensionless
Entry control	e	Proportion of infected animals that are denied entry to the Region of Concern.	dimensionless

Lot 2 Model

- $N = v \times p \times (1 - d) \times t \times (1 - e)$

- N is the number of animals entering the RC from endemic countries
- d (departure control) is the proportion of animals for export that are denied departure from the endemic country
  - It is assumed equal to 0, so the model becomes
    $N = v \times p \times t \times (1 - e)$

Volume

- Officially, all RC countries have banned live import of animals from the endemic countries. However, there are believed to be large numbers of unofficial animal movements (smuggling, traditional tribal movements, etc.)
- Import from East Africa into the RC is considered larger than from West Africa.
- Sudan is considered to be a main producer of livestock and exporter of animals. There are two major trade flows: from Sudan upwards over the Nile into Egypt, and from the Horn of Africa into Yemen and Saudi Arabia.
- Official animal imports from the Horn of Africa into Saudi Arabia are around 6 million ruminants.
- There is a vast demand for sheep around Eid al-Adha, a Muslim feast when it is traditional for a family to slaughter an animal, which may increase the numbers of undocumented animal movements.
- Control measures at ports and on the Red Sea are severe.
- From the occurrence of diseases with African origin, such as lumpy skin disease in Israel, undocumented movements into Jordan and Israel must exist, but they are hard to quantify.

Prevalence

- This parameter is the proportion of animals for transport that are infective, averaged over the whole year.
- The proportion of animals in an epidemic area that would be infected at some time during an epidemic is between 10% and 40%, based on data from antibody testing after epidemics.
- Prevalence in animals for export over the whole year will be reduced by the following factors:
  - Prevalence is negligible except when an outbreak of RVF occurs.
  - The proportion of animals infected in an epidemic that would be infected at any one time during the infection. This would depend on the length of the epidemic (typically around 12 weeks) and the length of time that an animal is infected (from a few days to two weeks).
  - The proportion of traded animals from the East Africa that would be traded during the epidemic period. In 2014 the Muslim festival Eid al-Adha (when there is increased demand for sheep) was in October, at the end of the rainy season, when both vector population density and age are compatible with the RVF transmission.
  - The proportion of animals traded from the East Africa that will be traded from the epidemic area. For example, a recent large epidemic in West Africa was 20,000km^2 in Senegal, while the total area of the West Africa source is more than 5,000,000 km^2).

Transport

- The majority of transportation is on foot, or a combination of driving and walking.
- Journeys on foot take up to about 2 weeks from the West African source countries. A combination of driving and walking takes about half as long. The average journey length would typically be shorter from East Africa than West Africa.
- Animals that are infective when loaded for transportation may no longer be infective on arrival at the RC. The infective period is thought to be about 30 days.
- New infections of vectors are possible during the journey since RVF vectors are important in some places along the Nile river. However, taking into account the incubation period for RVFV in mosquitoes, infection from mosquitoes that were infected by animals on the same truck seems very unlikely.

Entry controls

- No official import, so all trade is undocumented import and thus no import controls applied.
- The quarantine procedures applied in adjacent countries to the Region of Concern, such as Egypt or Saudi Arabia, do not prevent recently infected animals from entering and spreading the virus further to mosquitoes or susceptible animals, since only serological testing (Immunoglobulin G) is carried out in these quarantines.
- The great majority of movements of live animals will be unofficial/illegal and will only be subject to sporadic policing, if any.

Your estimates

Parameter name	Symbol	m	s	Units
Volume	v			animals
Prevalence	p			dimensionless
Transport	t			dimensionless
Entry control	e			dimensionless

m: approximate central estimate
s: approximate standard deviation
These are for use in sensitivity analysis and will be explained at the course (Lecture 4)

Biohazards example

- Ebola virus in bushmeat http://www.efsa.europa.eu/en/efsajournal/doc/3884.pdf
- **Short title**: Ebola

The following slides contain a simplified assessment for use only as training example for the EFSA Training Course on Steering an Expert Knowledge Elicitation

Risk Assessment Description

- EFSA was asked to review the risk for persons in Europe linked to transmission of Zaire Ebola virus (ZEBOV) via handling and preparation (both carried out by consumers immediately before consumption) as well as consumption of bushmeat illegally imported from Africa.
- The risk is the result of a combination of several necessary steps:
  - the bushmeat has to be contaminated with ZEBOV at the point of origin;
  - the bushmeat has to be (illegally) introduced into the EU;
  - the imported bushmeat needs to contain viable virus when it reaches the person;
  - the person has to be exposed to the virus;
  - the person needs to get infected following exposure.
- The public health consequences of any infection are very serious, given the high lethality and potential for secondary transmission.

List of parameters

Parameter name	Symbol	Description	Units
Amount imported	A	Amount of bushmeat illegally imported into Europe	tonnes
Proportion contaminated	p1	Proportion of bushmeat that is contaminated with ZEBOV	proportion of 1kg units
Processing effect	p2	Effect of processing on viability of ZEBOV virus	proportion of 1kg units
Transport effect	p3	Effect of transport on viability of ZEBOV virus	proportion of 1kg units

Lot 2 Model

- This example assesses the amount of contaminated bushmeat illegally entering the EU
- Evidence from primate studies indicates that ZEBOV is highly infective. Consequently the number of bushmeat portions containing any viable ZEBOV virus will be a major driver of the potential number of human infections.
- Therefore, the model for this example estimates E, *the number of 1 kg units of bushmeat illegally imported to the EU during 2015 that contain viable ZEBOV virus*

$$E = 1000 \times A \times p1 \times (1 - p2) \times (1 - p3)$$

Amount imported

- Spot-checks on luggage of 61335 passengers entering the EU at Vienna airport in 2012-13 found 6 items of bushmeat (1 per 5000 pieces of luggage)
- 5.5 tonnes of meat were seized from air passenger luggage entering Switzerland in 2008-2011, of which 249 kg was bushmeat. The total annual bushmeat inflow for Switzerland was estimated by modelling as 8.6 tonnes (95 % CI 0.8 to 68.8)
- A 2010 study found 7% of 134 air passengers arriving at Paris from sub-Saharan Africa carried bushmeat, and estimated that 273 tonnes per year are imported annually on these flights.
- A 2007 study estimated the total flow of illegal meat (not just bushmeat) from West Africa to UK as 1213 tonnes per year (95% CI 399 to 3082). This study included both air and sea routes.

Proportion contaminated

- Contamination with ZEBOV is more likely in bushmeat from areas experiencing active virus transmission in wildlife, especially for species susceptible to infection by ZEBOV and for animals found dead rather than hunted.
- A 2012 study detected EBOV in 17/33 apes and 1/22 non-primates found dead during a human Ebola outbreak. Of animals captured alive, 13/1418 bats were positive and 0/3891 other animals.
- The EFSA opinion lists species from which imported bushmeat has been reported as including apes, antelopes, pangolin, birds, porcupines, other rodents, crocodile and blue duiker
- Six bushmeat samples detected at Vienna airport originated from Nigeria (n=3), South Africa (2) and Ethiopia (1). Bushmeat detected in Switzerland was mainly from West Africa.

Processing effect

- Processing method is usually not reported, but may include salting, drying or smoking, which are expected to reduce ZEBOV infectivity.
  - Some studies suggest most carcasses sold in West African markets are processed to prolong shelf life
  - Of six bushmeat samples detected at Vienna airport, 1 was cooked, 2 smoked and 3 dried
  - Processing was known for only 2 of the bushmeat items detected in Switzerland: one was fresh, one dried
  - Bushmeat detected entering the USA included some fresh items, some raw transported in a cooler, some lightly smoked, some well dried
- There is almost no information on the effect of processing on ZEBOV. Heat and smoke constituents may lead to inactivation of ZEBOV. The virus has an envelope as outer membrane.

Transport effect

- Survival of ZEBOV in transport will depend on type of product, conditions (vacuum packing, temperature, etc.) and duration of transport
  - Some studies suggest most carcasses sold in West African markets are processed to prolong shelf life
  - Of six bushmeat samples detected at Vienna airport, 1 was cooked, 2 smoked and 3 dried
  - Processing was known for only 2 of the bushmeat items detected in Switzerland: one was fresh, one dried
  - Bushmeat detected entering the USA included some fresh items, some raw transported in a cooler, some lightly smoked, some well dried
- ZEBOV will survive better in fresh or frozen meat, less well in dried or smoked meat
- EFSA opinion notes that some items could contain moist inner tissues, which could favour survival of the virus
- There is no information on survival of ZEBOV in meat or animal products, but it is expected to be better at low temperature (4^{o}C) than at room temperature

Your estimates

Parameter name	Symbol	m	s	Units
Amount imported	A			tonnes
Proportion contaminated	p1			proportion of 1kg units
Processing effect	p2			proportion of 1kg units
Transport effect	p3			proportion of 1kg units

m: approximate central estimate
s: approximate standard deviation
These are for use in sensitivity analysis and will be explained at the course (Lecture 4)

TRAINING COURSE ON STEERING AN EXPERT KNOWLEDGE ELICITATION
Contract: OC/EFSA/AMU/2014/03-CT2

COURSE TIMETABLE: Final version

SESSION I. 1330 – 1800, DAY 1

PART 1. Problem definition: role of the Working Group

13:30 INTRODUCTION: Course objectives and agenda
13:40 LECTURE 1. Introduction – reasons and roles for the use of EKE in EFSA risk assessments
14:05 PRACTICAL 1. Examples of expert judgement in EFSA's work
14:35 LECTURE 2. Key principles for EKE
14:55 PRACTICAL 2 - plenary. Discussion of key principles
15:10 LECTURE 3. Probabilistic expert judgements
15:35 PRACTICAL 3. Probabilistic expert judgements - work individually
15:55 *Break*
16:25 LECTURE 4. Identifying priority parameters for EKE
17:00 PRACTICAL 4 - breakout groups. Identifying priority parameters for EKE: sensitivity analysis
17:40 PLENARY DISCUSSION Feedback from practical
18:00 SESSION ENDS

HOMEWORK – consider how what you've learned on day 1 would apply to an example assessment from your own area of work

SESSION II. 0900 – 1300, DAY 2

PART 2. The pre-elicitation phase: role of the Steering Group

09:00 LECTURE 5. Specifying questions for EKE
09:30 PRACTICAL 5 - breakout groups. Specifying questions for EKE
09:55 PLENARY DISCUSSION - report back from breakout groups
10:15 LECTURE 6. Identifying, selecting, motivating and training experts for an elicitation
10:45 *Break*
11:15 PRACTICAL 6 - breakout groups. Identifying, selecting, motivating and training experts for an elicitation
11:40 PLENARY DISCUSSION - report back from breakout groups
12:00 LECTURE 7. The evidence dossier
12:15 LECTURE 8. Sheffield Method
13:00 *Lunch*

SESSION III. 1400 – 1800, DAY 2

PART 3. The elicitation phase: role of the Elicitation Group

14:00 PRACTICAL 7 - breakout groups. Key aspects of steering the Sheffield method
14:30 LECTURE 9. Delphi Method
15:05 PRACTICAL 8 - breakout groups. Key aspects of steering the Delphi method
15:35 *Break*
16:05 LECTURE 10. Cooke Method
16:50 PRACTICAL 9 - breakout groups. Key aspects of steering the Cooke method
17:20 PLENARY DISCUSSION - report back from breakout groups
18:00 SESSION ENDS

HOMEWORK – consider how what you've learned on day 2 would apply to an example assessment from your own area of work

SESSION IV. 0900 – 1300, DAY 3

09:00 LECTURE 11. Selecting the appropriate elicitation method
09:25 PRACTICAL 10 - breakout groups. Selecting the appropriate elicitation method
09:55 PLENARY DISCUSSION - report back from breakout groups
10:25 *Break*

PART 4. The post-elicitation phase

10:55 LECTURE 12. Documentation: repeatability, transparency and confidentiality
11:10 LECTURE 13. Advanced topics in EKE
11:35 LECTURE 14. Steering the elicitation process: review of main points
11:55 PRACTICAL 11 - work individually. Planning EKE for examples from each participant's own area of work
12:35 PLENARY DISCUSSION - opportunities and challenges for uptake in participants' work areas
12:55 COURSE EVALUATION QUESTIONNAIRE.
13:00 COURSE ENDS

Training course on steering an expert knowledge elicitation

Introduction: Objectives and Agenda

Andy Hart

WELCOME TO THE COURSE !

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Uncertainty and Variability L0/1

Training in how to steer an expert knowledge elicitation (EKE)

- **Participants**: EFSA staff and experts who may be involved in steering EKE exercises in the future
- **Tutors** for September 2015:
  - Andy Hart
  - Tony O'Hagan
  - John Quigley
- Other contributors to course material:
  - Simon French
  - Fergus Bolger

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L0/2

EFSA Guidance on EKE

- Published June 2014
- Applicable to all areas of EFSA's work
- Part of EFSA's wider set of guidance on risk assessment methodology

EFSA (2014)

http://www.efsa.europa.eu/en/efsajournal/pub/3734.htm


efsa

GUIDANCE OF EFSA

Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment

European Food Safety Authority

ABSTRACT

KEY WORDS

Related EFSA Guidance and Courses

- E-learning module on Making Probability Judgements (in preparation)
- Draft Guidance on Uncertainty in EFSA Scientific Assessment (EKE plays a key role in this)
- Training Courses on Uncertainty and Variability in Risk Assessment
- Guidance on Weight of Evidence Assessment (in preparation)
- Guidance and Course on Systematic Review
- EFSA Prometheus Project (Promoting MeTHods for Evidence Use in Scientific assessments)

Learning objectives

On completing the course, participants shall be able to:

1. Recall the characteristics of Expert Knowledge Elicitation (EKE)
2. Explain the role and purpose of EKE in risk assessment
3. Explain probabilistic expert judgements
4. Identify and prioritise tasks in risk assessment suitable for EKE
5. Frame a problem for EKE
6. Identify, select, and motivate experts for an elicitation
7. Decide on training needs for the experts
8. Produce background information for an elicitation
9. Recall typical protocols using the Cooke, Delphi and Sheffield methods
10. Discuss and select the appropriate elicitation method
11. Define the elicitation protocol, incl. adaptations, resources and selection of elicitors
12. Document and interpret results; discuss and handle risks of elicitations
13. Produce a complete documentation of an EKE
14. Discuss handling of confidentiality during an EKE
15. Discuss issues of repeatability of an EKE.

Course agenda

- Part 1: Problem definition phase
  - Introduction, key principles, probability judgements, identifying where EKE is needed
- Part 2: Pre-elicitation phase
  - Framing EKE questions, selecting experts, collating evidence
- Part 3: Elicitation phase
  - Three basic methods; choosing which to use
- Part 4: Post-elicitation
  - Documentation, advanced topics, overview

Slides include references to more detail in the Guidance Document, e.g. *GD 1.2*

Examples for practical sessions

- Chemical risk – dermal exposure to bisphenol A
- Nutrition – Vitamin B12 requirement
- Environmental risk – GM pollen transport
- Plant health – Citrus Black Spot entry pathway
- Animal Health – Rift Valley Fever
- Biohazard – Ebola in bushmeat

Note: Examples have been simplified for purpose of training course

Your contribution

- Before the course:
  - Choose one of the course examples (previous slide)
  - *Also*, choose an example assessment from your area of EFSA and bring relevant documentation to the course
- During the course:
  - Consider how the course content applies to your examples
  - Ask if anything isn't clear
  - Please return promptly after breaks!
- After the course:
  - Give feedback to help us improve the course
  - Work with your colleagues to apply what you learned
  - Refer to the Guidance Document for more detail

Lecture 1: Introduction – reasons and roles for the use of EKE in EFSA risk assessments

Andy Hart

Outline

- Role of expert judgement in risk assessment
- Purpose of Expert Knowledge Elicitation ***(EKE)***
- Major challenges and choices in EKE
- EKE process and responsibilities in EFSA
- Example from EFSA work

Types of information used in risk assessment

- **Data** – from the open literature, applicant submissions, national and international databases, etc.
- **Expert knowledge**
  – Information – facts, sources, default values, etc.
    - Expert either knows it, or knows where it is
  – Judgements – correct answer is unknown
    - Qualitative – yes/no questions, categories, etc.
    - ***Quantitative judgements – estimates, assumptions, etc***

EKE Guidance focusses on quantitative judgements

Qualitative judgements are mentioned briefly in GD Appendix A.1.2 *GD 1.2*

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Uncertainty and Variability L1/3

Examples of quantitative expert judgement in EFSA

EFSA Scientific Committee Opinion on Default Values

- Reviews and harmonises use of *default* values
- Identifies places where no defaults can be given and *case-specific judgements* are required


efsa

SCIENTIFIC OPINION

ABSTRACT

KEY WORDS

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L1/4

Examples of quantitative expert judgement in EFSA

Case-specific judgements required for:

- Uncertainty factors for deficiencies in available data
- Extrapolation from subacute to chronic exposure
- Extrapolation from LOAEL to NOAEL
- Factor to account for the severity of an effect


SCIENTIFIC OPINION

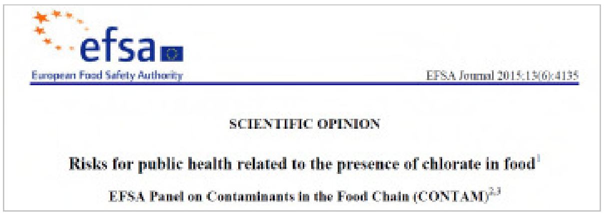
Guidance on selected default values to be used by the EFSA Scientific Committee, Scientific Panels and Units in the absence of actual measured data

EFSA Scientific Committee

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L1/5

Examples of quantitative expert judgement in EFSA

- *Read-across*: data from one context is extrapolated to another context, where data is lacking
- E.g. Assessment of *chlorate*:
  - Tolerable daily intake (TDI) set to 10x TDI for *perchlorate*, accounting for difference in potency


efsa European Food Safety Authority

EFSA Journal 2015;13(6):4135

SCIENTIFIC OPINION

Risks for public health related to the presence of chlorate in food[1]

EFSA Panel on Contaminants in the Food Chain (CONTAM)[2,3]

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L1/6

Examples of quantitative expert judgement in EFSA

- Values defining *exposure scenarios* are commonly a combination of data and expert judgement
- E.g. EFSA statement on melamine, 2008:

Table 2: Scenarios for potential melamine dietary intake from consumption of contaminated confectionary imported from China.

Melamine concentration		Dietary exposure mg/kg b.w. per day			
		60 kg adult		20 kg child	
		Mean[a]	95th percentile[b]	Mean[a]	95th percentile[b]
Milk powder	**Milk toffee (10%)**				
29 mg/kg	2.9 mg/kg	0.0007	0.0020	0.0020	0.0061
2563 mg/kg	256.3 mg/kg	0.0598	0.1794	0.1794	0.5382
	Chocolate (25%)				
29 mg/kg	7.3 mg/kg	0.0017	0.0051	0.0051	0.0152
2563 mg/kg	640.8 mg/kg	0.1495	0.4485	0.4485	1.3456

a) Mean daily consumption of 0.014 kg

b) 95th percentile daily consumption of 0.042 kg

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L1/7

Examples of quantitative expert judgement in EFSA

- The case studies used in this course:
  - Chemical risk – dermal exposure to bisphenol A
  - Nutrition – Vitamin B12 requirement
  - Environmental risk – GM pollen transport
  - Plant health – Citrus Black Spot entry pathway
  - Animal Health – Rift Valley Fever
  - Biohazard – Ebola in bushmeat

Quantitative expert judgement is used in all areas of EFSA's work

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L1/8

Quantitative expert judgements

- Estimates, assumptions, etc.
- Correct answer is unknown

- Personal, subjective and uncertain
- Differ between experts

→*Need to select appropriate experts*
→*Need to take account of their uncertainty*

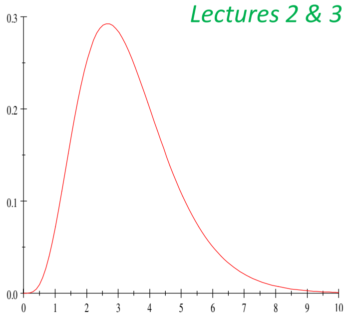
GD 1.2

EFSA EKE Guidance

*…uses **probability distributions** to express uncertainty*

- Expresses the *range* of possible values and their *relative likelihoods*
- Avoids the ambiguity of verbal expressions
- Enables us to calculate the impact of the uncertainty on the risk assessment outcome

→ more in Lectures 2 & 3



GD 2.1.2

Expert Knowledge Elicitation (EKE)

Defined by EFSA as:

'A systematic, documented and reviewable process

…to retrieve expert judgements from groups of experts

…in the form of a probability distribution'

Glossary, EFSA EKE Guidance

GD Glossary

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L1/11

Challenges to EKE

- Human thinking processes can cause *biases* in judgement (anchoring, availability, etc.) → *more in Lecture 3*

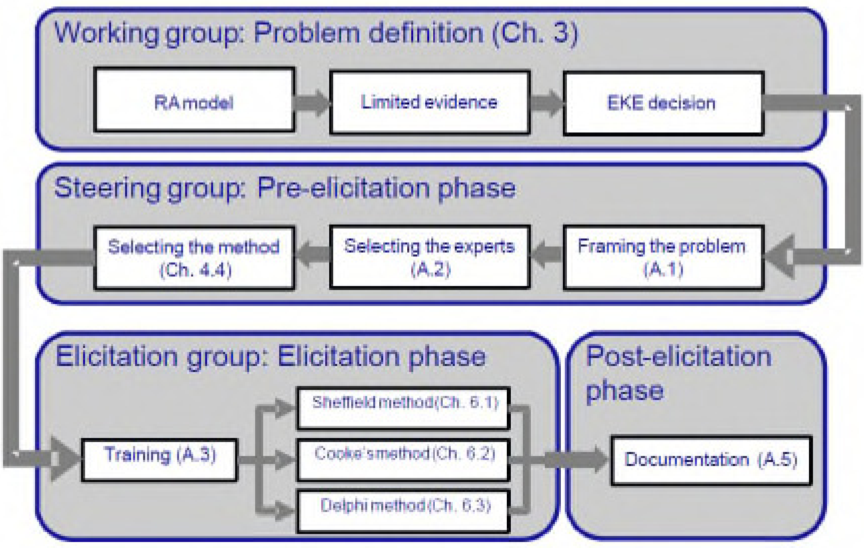
EKE methodology seeks to reduce these biases

- Time and resource limitations
- Practical choices (what & how to elicit, etc.)

EFSA Guidance provides a structured process

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L1/12

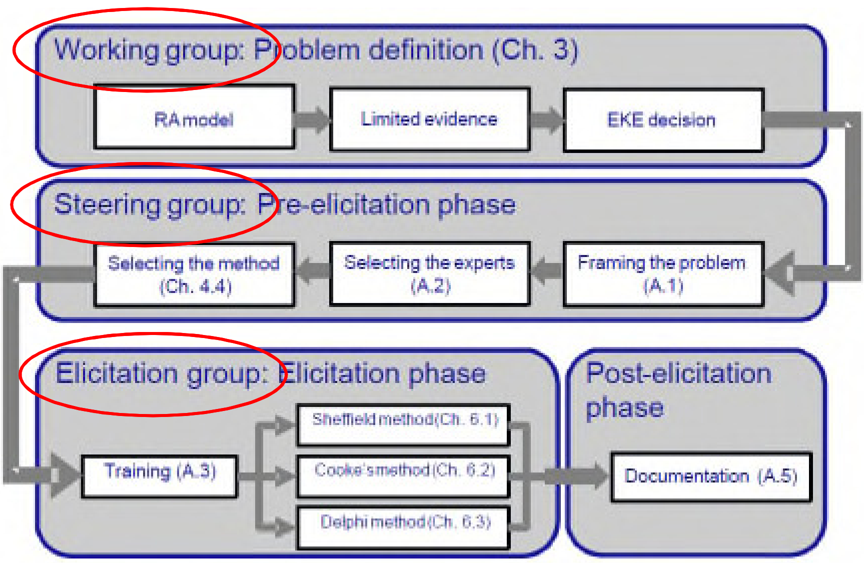
EFSA *process* for EKE



GD Figure 1, p. 8

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L1/13

EFSA *roles and responsibilities* for EKE

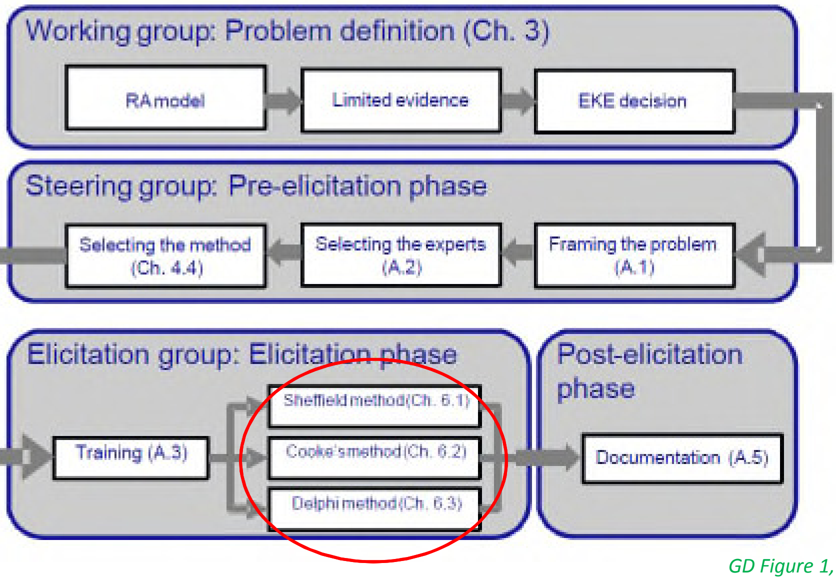


GD Figure 1, p. 8

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L1/14

3 basic methods for EKE

→ details in Lectures 8-11

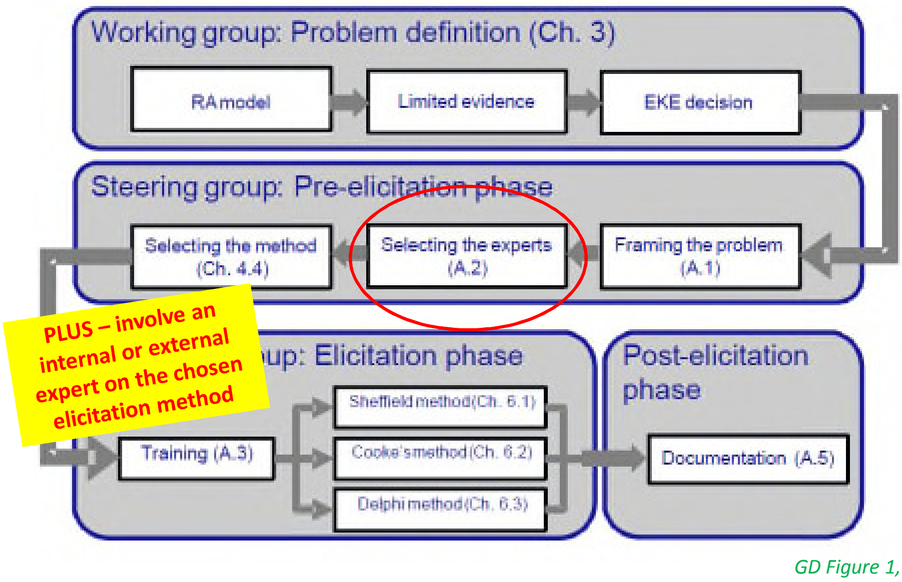


GD Figure 1, p. 8

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L1/15

Wider field of *experts*

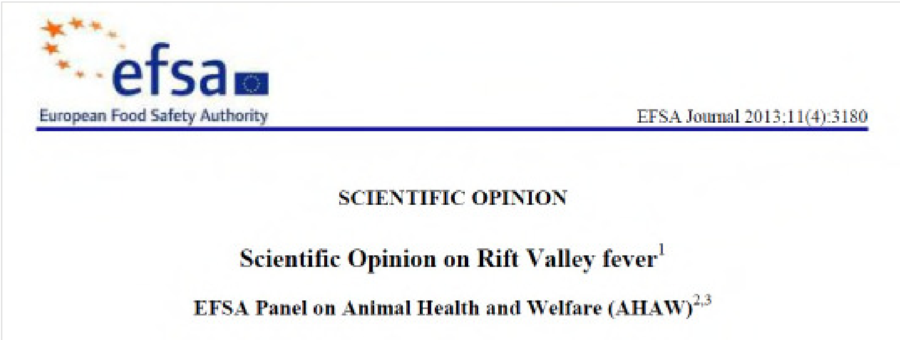
→ more in Lecture 6



GD Figure 1, p. 8

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L1/16

An early example of EKE from EFSA work


European Food Safety Authority

EFSA Journal 2013;11(4):3180

SCIENTIFIC OPINION

Scientific Opinion on Rift Valley fever[1]

EFSA Panel on Animal Health and Welfare (AHAW)[2,3]

Rift Valley Fever

- Affects cattle, sheep, goats and camels
- Virus transmitted by mosquitoes
- Endemic in East and West Africa
- Region of concern is North Africa

Map by Strebe, via Wikimedia Commons

Rift Valley Fever

- AHAW Panel used a simple model to estimate the number of animals introduced in an outbreak year
- Limited information for key model parameters:
  - Number of animals exported
  - Prevalence in exported animals
  - Proportion remaining infectious
- Used EKE to elicit distributions

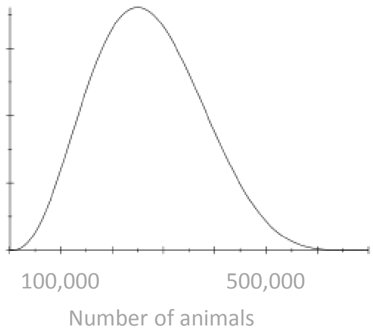
EKE process for Rift Valley Fever

- Invited relevant experts from Africa and the EU:
  - Egypt (2), Israel (2), Morocco (2), Mauritania, Senegal, Palestinian Territories, Tunisia (4), Kenya, Saudi Arabia
  - Spain (3), Italy, France
- Two 2-day workshops
- EKE using the Sheffield method
  - Judgements combined by discussion and consensus

Number of animals exported from West region

Factors considered by the experts:

- Undocumented animal movements can be expected, especially around the Feast of the Sacrifice
- Nomadic lifestyle facilitates movement by herding
- Trucks can take larger numbers but travel on controlled roads
- The border between Mali and Algeria was completely closed during 2013
- Morocco has stringent controls, but import of camels could still occur by desert roads
- Libya was previously estimated to import 130,000 ruminants in 2012
- Political unrest could have an enormous influence

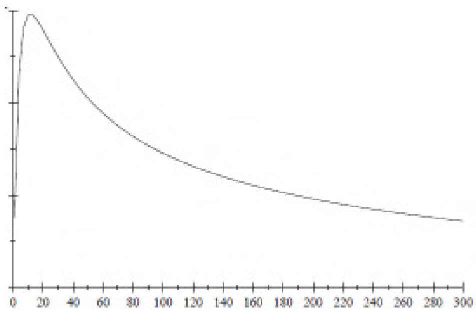


'The experts judged that it would be very unlikely that import from the west source into the RC would be below 25,000 and above 500,000 ruminants in 2013. The median was set at 260,000, with a high uncertainty.'

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L1/21

Rift Valley Fever

- Elicited distributions fed into Panel's model
- Output distribution peaks around 20
- Very uncertain – some probability of exceeding 1000



Number of infected animals entering Region of Concern in a year with an outbreak in both the East and West regions

AHAW conclusion: 'some hundreds of RVFV-infected animals will be moved into the RC when an epidemic in the source areas occurs'

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L1/22

Summary – what is EKE?

- EKE is a process of
  - representing the judgements of experts
  - concerning an uncertain quantity
  - as a probability distribution
- EKE methods are formal, rigorous probabilistic judgement techniques
  - designed to encourage careful, thoughtful judgements
  - and reduce psychological biases
- EFSA EKE Guidance implements EKE in an efficient, rigorous and transparent manner
  - targetted on most important uncertainties
  - subject to critical review at key decision points
  - fully documented
- EKE plays a key role in EFSA's Draft Guidance on Uncertainty

References

- EFSA, 2008. Statement of EFSA on risks for public health due to the presences of melamine in infant milk and other milk products in China. EFSA Journal (2008) 807, 1-10.
- EFSA, 2012. Guidance on selected default values to be used by the EFSA Scientific Committee, Scientific Panels and Units in the absence of actual measured data. The EFSA Journal, 2579, 1-32.
- EFSA, 2013. Scientific Opinion on Rift Valley fever. EFSA Journal 2013;11(4):3180 .
- EFSA, 2014. Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment. EFSA Journal 2014;12(6):3734.
- EFSA, 2015. Risks for public health related to the presence of chlorate in food. EFSA Journal 2015; 13(6):4135
- EFSA, 2015. Guidance on uncertainty in EFSA scientific assessment. Draft for public consultation.

Practical 1: Examples of expert judgement in EFSA's work

Andy Hart

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation P1/1

Objectives

- To identify examples where expert judgement is used in EFSA work

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation P1/2

Task

- Identify examples of numbers in risk assessment being wholly or partly based on expert judgement
  - From your work for EFSA
  - Or other EFSA work you are aware of

- Work on your own or with your neighbour
- 5 minutes to identify examples
- 25 minutes round table:
  - *Introduce yourself: name and Panel/Unit*
  - *Short description of your example*

Lecture 2: Key principles for EKE

Tony O'Hagan,
Fergus Bolger,
John Quigley

Outline

- Parameters, uncertainty and variability
- Probabilities and judgements
- Science and subjectivity

This lecture extends and updates GD 2.1.3

Quantifying uncertainty

Lecture 1 stated that:

> Knowledge or uncertainty about a parameter in a risk assessment will be formally described by a probability distribution

GD 2.2.2

Uncertainty

- Before proceeding further we need to consider the nature of parameters in RA and how probability represents uncertainty about them
  - In particular, we need to distinguish between:
    - Quantities whose values are uncertain because they vary randomly
    - Quantities that are fixed and unique, but which are uncertain because we are unsure of their true values
- This is important…
  - … because the parameters that we require expert judgements about are generally of the second kind
  - Even where there are quantities in the RA that vary randomly, the parameters we ask experts to judge are fixed and unique aspects of that variability

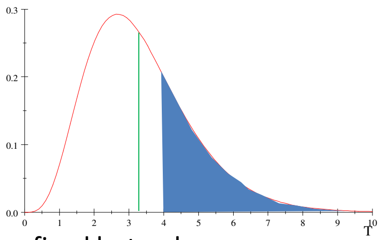
Consider the following example →

Example – Pathogen in meat

- Meat is displayed in a butcher's window until purchased
  - In a chiller cabinet at 10ºC
- If a certain pathogen is present in the meat when put in the cabinet, how much will be present when the customer buys it?
- Three uncertain parameters
  - Pathogen load when placed in the cabinet
    - This varies randomly
  - Time on display until purchased
    - This also varies randomly
  - Rate of reproduction of the pathogen
    - This is a fixed quantity

Example – Time on display

- What is the uncertainty about time on display?
- It varies randomly between customers and between butchers
  - We can describe that randomness with a probability distribution
  - Probability density function is a familiar graphical representation
- This distribution is fixed but unknown
  - For instance we don't know the average time on display
  - Or the proportion of customers who buy when the meat has been on display for more the 4 hours
  - The average and the proportion are fixed but unknown
  - We will elicit expert judgements about these parameters



Uncertainty and variability

The parameters that we require expert judgements about have fixed and unique, but uncertain, values

Even where there are quantities in the RA that vary randomly, the parameters we ask experts to judge are fixed and unique, but uncertain, aspects of that variability

GD 2.1.3

The meaning of probability (1)

- It is also important to know what probabilities mean
- One way to define a probability is as the proportion of times that something happens (over very many occasions)
  - This is called the frequency definition
  - For instance, in the frequency definition an event has probability 0.6 if it happens on 60% of occasions

Probability 0.6 → 60% of occasions

GD 2.1.3

The meaning of probability (2)

- Consider the average time on display – what would it mean to say that the probability that this parameter lies between two values *a* and *b* is 0.6?
  - That probability of 0.6 *cannot* mean that it lies between *a* and *b on* 60% of occasions

~~Probability 0.6 → 60% of occasions~~

  - Because the average time on display has a unique true value. It doesn't vary
  - The frequency definition cannot apply to a parameter like this whose value is fixed (but unknown)

GD 2.1.3

The meaning of probability (3)

- Frequency probability is

  - Because most of the parameters for which we seek expert judgements in RA are not variable
  - They have unique, fixed values

We need another definition – an alternative way to think about probability

Probabilities are judgements

- The probability distributions in RA will be *expert judgements*
  - So the statement that a parameter has a 60% probability of lying between *a* and *b* is a judgement
  - Representing the expert's *degree of belief* that the parameter's true value will be between *a* and *b*
  - EFSA's risk assessment will use expert judgement-based probability distributions for parameters
    - In order to assess the degree of uncertainty in the RA outcomes or conclusions

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L2/11

Subjective probability

- These kind of probabilities are called subjective probabilities
  - They are subjective because they are the personal judgements of individual experts
- "Surely this is totally unscientific?"
  - A common reaction when first introduced to subjective probability
  - But please read on …

GD 2.1.3

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L2/12

Subjective, but scientific (1)

You want to use subjective probability judgements? Isn't that totally unscientific? Science is supposed to be objective.

Yes, objectivity is the goal of science, but scientists still have to make judgements. These judgements include theories, insights, interpretations of data. Science progresses by other scientists debating and testing those judgements. Making good judgements of this kind is what distinguishes a top scientist.

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L2/13

Subjective, but scientific (2)

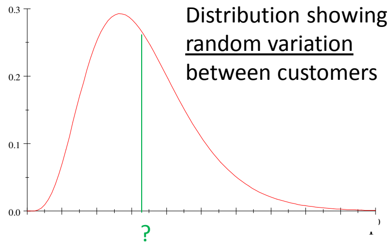
But subjective judgements are open to bias, prejudice, sloppy thinking …

Subjective probabilities are judgements but they should be careful, honest, informed judgements. In science we must always be as objective as possible. Probability judgements are like all the other judgements that a scientist necessarily makes, and should be argued for in the same careful, honest, informed way.

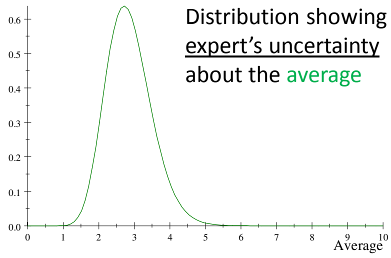
FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L2/14

Example – Average time on display

- We could ask for a public health inspector's beliefs about the average time on display
  - Which is a fixed but unknown feature of the variability in times between customers and shops
- The expert's knowledge will be represented by a subjective probability distribution
  - Based on the expert's judgements
  - Should be as objective and scientific as possible



Distribution showing <u>random variation</u> between customers



Distribution showing <u>expert's uncertainty</u> about the average

Best practice

- EKE methods are formal, rigorous probabilistic judgement techniques
  - Designed to encourage careful, thoughtful judgements
  - Explicit and documented
  - And structured to eliminate prejudice, bias, guessing, superstition, wishful or sloppy thinking, manipulation …

EKE is <u>best practice</u> for quantifying uncertainty in RA

Subjective probability judgements

- Let's actually make some probability judgements
- Remember, your probability for a proposition E is a measure of your degree of belief in the truth of E
  - If you are certain that E is true then P(E) = 1
  - If you are certain it is false then P(E) = 0
  - Otherwise P(E) lies between these two extremes
- Exercise 1 – How many Muslims in Britain?
  - Refer to the two questions on your sheet
  - The first asks for a probability
    - Make your own personal judgement
  - The second asks for another probability

Don't use the internet!!!

References

- EFSA, 2014. Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment. EFSA Journal 2014;12(6):3734.

Commentary on the Exercises in EFSA Training Course on "Steering an Expert Knowledge Elicitation"

The Muslims Exercise

This exercise was run in all three of the training courses in Parma in 2015. The purpose is to confirm the effect of anchoring in an experimental context that is more relevant to actual EKE than has been studied previously.

The principal feature of this exercise that is not present in other demonstrations of anchoring is that the first and second question both ask for probability judgements. These are judgements of the form P(X > x) that might realistically be used in practical EKE. The idea is to show that the initial choice of a value for x will serve as an anchor and bias the experts' judgements.

It should be said that this experiment (like most experiments in the psychology of judgement) does not involve real experts and is based on a simple 'almanac'-type question. Also, the participants had received only the briefest introduction to making probability judgements. We may not necessarily expect to see anchoring effects of a comparable magnitude in real EKE.

The parameter being judged here, M, is the number of people (in millions) in England and Wales who reported their religion as "Muslim" in the 2011 census. The participants randomly received one of two different versions of the exercise. In one, they were asked first to give their probability P(M > 2) and then P(M > 8). In the second, the order of questions was reversed. In each case, they could not see the second question until they had answered the first.

The aggregated results for the three Parma courses are shown in Table 1.

	2 First	**8 First**
P(M > 2)	0.692	0.810
P(M > 8)	0.370	0.422

Table 1. Average responses, Muslim exercise, all three Parma courses

The figures in each cell are averages from all probability judgements obtained on the three courses, in each case from about 30 respondents. For instance, the average value given for P(M > 2) among all the respondents who received the version of the questionnaire which asked first for P(M > 2).

On the basis of the psychological findings of anchoring, we would expect the respondents who received the P(M > 8) question first would give higher probability judgements on average than those who received the P(M > 2) question first, because they had been anchored on the figure 8 rather than 2. This is indeed what we see in Table 1, with average values in the second column higher than the corresponding values in the first.

The evidence so far supports the anchoring theory. The sample size is not large enough for the findings to be statistically "significant", but we would certainly expect data from future deliveries of the course to continue to strengthen the evidence in favour of the anchoring effect.

It is worth noting that this exercise has also been run in a variety of other training courses with a variety of audiences. The aggregate averages from all the courses (more than 70 respondents in each case) are given in Table 2. Although the differences are now statistically significant, such an analysis is questionable because of the heterogeneity of the audiences. The value of the Parma data is that the three audiences were all made up of people drawn from the same pool (EFSA staff and experts), and this is a strong reason for continuing to run the exercise in future EFSA deliveries of the course.

	2 First	**8 First**
P(M > 2)	0.692	0.794
P(M > 8)	0.318	0.397

Table 2. Average responses, Muslim exercise, all courses

It should also be noted that in all courses, responses from some participants have been excluded from the above figures because they gave inconsistent judgements (with a higher probability for M > 8 than for M > 2, or some other clear evidence of their having misunderstood the task). This is perhaps an inevitable consequence of the decision to place this exercise at a point in the course where the participants have not had any real training in probability judgement.

The Time to Linate Exercise

This exercise was also run in all three Parma courses. It was designed first to test whether respondents would produce appreciably different intervals when asked for either a 90% interval or a credible interval (meaning one with almost 100% probability). In general, if an individual's uncertainty is represented by a unimodal distribution then the credible interval should in most cases be much wider than the 90% interval. But the exercise sought to see whether in fact respondents might make essentially no distinction between the two.

The exercise was designed rather like the Muslims exercise, with two different versions asking for both 90% and credible intervals, but in different orders. Again, the second question was not visible until they had answered the first. At the point in the course where the exercise was given to the participants, they had not had any discussion of these intervals or training in how to judge them.

The parameter in question in this exercise was the average time (averaged over all working days in the year) for an EFSA shuttle to travel to Milan Linate airport if it left the EFSA main building in Parma at 16:00.

Table 3 shows the average widths of the intervals, in each case based on about 30 respondents over the three courses.

Considering the original purpose of the study, we see from the upper right and lower left cells (just looking at responses to the first question they were asked) that respondents did give appreciably wider credible intervals than 90% intervals on average. So they were *not* treating them as effectively equivalent, i.e. as if simply asking for an interval that the travel time was very likely to lie in.

	Credible	**90% First**

	First	
90% width	37.3	49.8
Credible width	71.3	107.1

Table 3. Average interval widths (minutes), Linate exercise, all courses

The more interesting finding in Table 3 is that the average widths in the right hand column are larger than the corresponding values in the left hand column. This is like the anchoring effect in the Muslims exercise, but in this case stems from the fact that, even though respondents gave wider credible intervals than 90% intervals in their first answers the difference was nevertheless not wide enough. When they started with a 90% interval and then widened it for their second answers they gave wider credible intervals than if they had been asked for them first. Similarly, when they started with a credible interval and narrowed it for their second answer they produced a 90% interval narrower than if they'd been asked for it first.

Although the sample sizes are not large enough for these differences to be formally significant, the same ordering of widths was observed in each of the three courses separately. So it is to be expected that the effects will be confirmed by repeating this exercise in future deliveries of this course.

It may be noted that again there were a number of rejected responses in each course (for instance where respondents gave 90% intervals that were wider than their credible intervals). Furthermore, it was clear that some did not appreciate the difference between a judgement about an individual travel time and about the requested average travel time.

The Italian Speakers Exercise

This exercise was added for the third course in Parma, and so we only have one set of responses. The intention was to explore the effect that the choice of bins has on respondents' probability distributions elicited using the roulette method (which is mentioned in the Appendix of the EKE Guidance document but does not figure in the three recommended protocols). The specific hypothesis is that respondents tend to use the full range of bins provided, and so their distributions should have larger standard deviations if the range of bins is wider. We might also find that the means of their distributions are higher if the middle of the range of bins is higher.

The parameter in this case was the proportion of EFSA employees in Parma who speak Italian fluently. The definition of "fluently" was B2 or higher in the Common European Framework of Reference for language skills, where B2 means "Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party."

Three different versions of this exercise were distributed randomly to the course participants, defined as follows.

- *Five narrow*. Five bins of width 10%, starting with 30% - 40% and ending with 70% - 80%.
- *Seven narrow*. Seven bins of width 10%, starting with 10% - 20% and ending with 70% - 80%.
- *Five wide*. Five bins of width 15%, starting with 15% - 30% and ending with 75% to 90%.

Respondents were given 20 counters (representing 0.05 probability each) to distribute among the bins, and were told that they could put them in the space to the left of the first bin or to the right of the last bin if they thought the range provided for the proportion did not cover their distribution.

Table 4 shows the results, as averages for the means and standard deviations of the respondents' distributions. For these calculations, their distributions were treated as discrete, with the bin probabilities concentrated at the centres of the bins.

	Mean	Standard deviation
Five narrow	0.625	0.096
Seven narrow	0.709	0.105
Five wide	0.686	0.159

Table 4. Average means and standard deviations, Italian exercise, final course

The table shows some interesting results which are not entirely as expected. The principal hypothesis is supported to the extent that the figures are increasing as we read down the standard deviations column, because the widths of the ranges of bins are also increasing as we read down (0.5 for *Five narrow*, 0.7 for *Seven narrow* and 0.75 for *Five wide*). However, the difference between the first two is really smaller than this hypothesis would suggest. What actually appears to be the case is that the standard deviation is driven by the width of an individual bin. The versions with narrow bins of width 10% have average standard deviations close to 0.1, while the version with wide bins of width 15% gave an average standard deviation close to 0.15.

The findings do not support the secondary hypothesis at all, because the *Seven narrow* version has the lowest central bin and yet has the highest average mean.

The sample sizes are so small (7 or 8) that any or all of these findings could easily be due to chance, so it will be interesting to see if they are supported by future deliveries of this course.

We can note that there were no rejected responses for this exercise. People generally find the roulette method easy to understand and to use – deceptively so because this exercise does suggest some unwanted influence from the choice of bins.

Practical 2: Key principles for EKE

Tony O'Hagan

Objectives

- To discuss the points made in Lecture 2: Key Principles for EKE

Task 2.1

- Some of the ideas and arguments in Lecture 2 will be challenging for some participants, maybe even disturbing
- It's important therefore to understand why these principles are indeed essential for Expert Knowledge Elicitation
- So feel free to *ask questions, challenge the lecturer and present your own opinions*

Lecture 3: Probabilistic expert judgements

Tony O'Hagan,
Fergus Bolger,
John Quigley

Outline

- Psychology
  - Probability judgements
  - Heuristics, anchoring, availability, overconfidence
- The basics
  - Two step process for specifying a subjective probability distribution
  - Multiple experts, aggregation

Just for background/refresher. Not too serious!
Much more in the e-learning course.

PSYCHOLOGY

Key messages from psychology

- Probabilities are not sitting preformed in people's heads
  - Just waiting for us to elicit them
- Judgements are formed only when needed
  - In response to questions
  - So the way we ask questions, and the order in which we ask them, influences the expert's judgements
- EKE methods are designed to avoid distorting experts' judgements

Heuristics

- How do we make judgements?
  - And what does it tell us about how our brains work?
- Our brains evolved to make quick decisions
  - Heuristics are short-cut reasoning techniques
  - Allow us to make good judgements quickly in familiar situations
- Judgement of probability is not something that we evolved to do well
- The old heuristics now produce biases

Anchoring and adjustment

Availability

Overconfidence

And many others !

GD 2.3.2

Anchoring and adjustment

- Exercise 1 was designed to exhibit this heuristic
  - The probabilities should on average be different in the two groups
- Any number that we have in our heads influences our next judgement
  - The number in our heads is an unconscious starting value for the new judgement
  - It acts like an anchor
  - Judgement is made by adjusting away from it
  - Adjustment is typically inadequate
- Remedy
  - Careful phrasing and sequencing of questions
    - Avoid creating anchors

GD 2.3.3

Availability

- The probability of an event is judged more likely if we can quickly bring to mind instances of it
  - Things that are more memorable are deemed more probable
    - High profile train accidents lead people to imagine rail travel is more risky than it really is

- Remedy
  - Review all relevant evidence at the start of the elicitation

Extends GD 2.3.2

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L3/7

Overconfidence

- It is generally said that experts are overconfident
  - When asked to give (e.g.) 95% intervals, far fewer than 95% contain the true value
- May be overstated but several possible explanations
  - Wish to demonstrate expertise
  - Anchoring to a central estimate
  - Difficulty of judging extreme events
  - Not thinking ‘outside the box’
    - Experts have their own specialist heuristics

- Remedy
  - Warn experts of the issues
    - and make it clear that we want honest expressions of uncertainty

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L3/8

THE BASICS

Eliciting a distribution

- We wish to elicit a probability distribution for a parameter in a risk assessment
- Sometimes this reduces to eliciting a single probability
  - For instance the probability that a bird flu virus mutates so that human transmission becomes possible
  - The parameter can only take two values, 0 or 1
    - Like a switch that is off or on – it mutates or it doesn't mutate
  - And a single probability determines the distribution
    - Because the probability of the switch being off (virus doesn't mutate) is one minus the probability of it being on (virus mutates)
- Sometimes the parameter has only a few possible values
  - Then we can elicit probabilities for each value

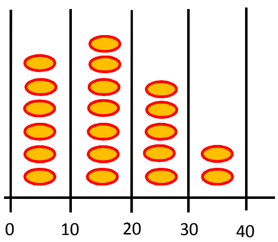
Too many probabilities!

- But parameters can usually take very many possible values
  - For instance the mean incubation time for a disease can take any positive value
- One way to define a distribution for a parameter X that can take any value in some range is as a set of probabilities
  - $P(X < x)$ for all possible x values
  - That's a lot of probabilities to elicit!
- If we sat down to elicit them one by one, the interrogation would never finish!
  - And we'd have serious anchoring problems!
- We need a pragmatic approach

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L3/11

A practical approach

- Ask the expert to construct his/her probability distribution visually
  - The range of possible values is marked out into a number of 'bins'
  - Expert places counters in the bins to represent probability
  - With 20 counters, for instance, each is worth probability 0.05 (5%)
- Exercise 2
  - What proportion of EFSA employees in Parma speak Italian?



GD Appendix A.3.5

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L3/12

A better pragmatic approach!

- In practice we (step 1):
  - elicit just a small number of carefully chosen judgements
    - recognising that it is impractical to ask for too many
- Then (step 2) we:
  - fit a reasonable probability distribution to those judgements
- We now consider these two steps
  - What judgements should we ask for?
  - What are the options for fitting a distribution?

Asking about the right things (1)

- Probabilities
  - Not expectations, standard deviations or other statisticians' favourites
  - Despite all the psychological warnings, probabilities are generally judged best
    - Such as the probability P(X < x) that the quantity X is less than some value x
    - Or the probability P(a < X < b) that X lies between values a and b
  - But as soon as we give a value to x or a or b we create an anchor
    - We should avoid introducing numbers

GD 2.4.4

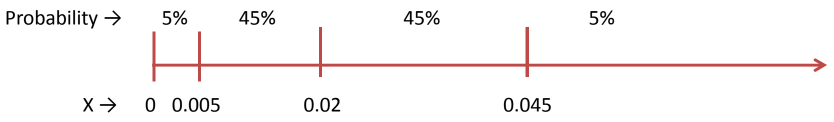
Asking about the right things (2)

- Quantiles
  - To avoid anchoring, ask for quantiles
    - So don't ask for P(X < x) because your choice of x influences the judgements
    - E.g. ask for the median value M such that P(X < M) = 0.5
    - Other quantiles by changing the 0.5
- Typically elicit 3 to 6 such judgements
  - Central value such as median, plus quantiles either side to quantify uncertainty
  - More of this in Practical 3

GD 2.4.4

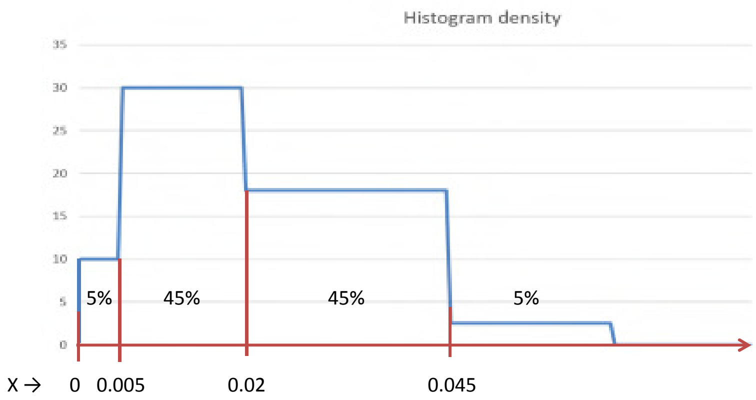
Example of fitting (1)

- X is the proportion, X, of batches of aquarium plants for import to the EU from a particular country that contain floating pennywort (Hydrocotyle ranunculoides)
  - She provides the following quantile judgements
  - 5th percentile 0.005 (0.5% of batches), median 0.02 (2%), 95th percentile 0.045 (4.5%)
- Her judgements only partially determine the distribution
  - We know only the probabilities shown



Example of fitting (2)

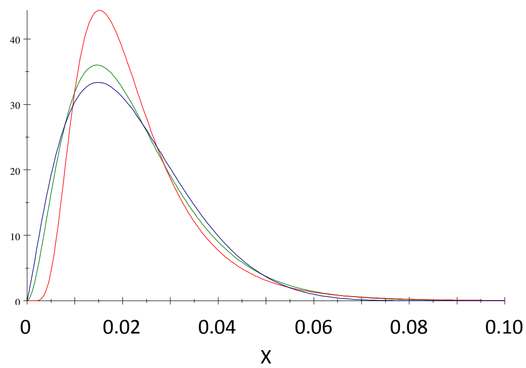
One option is simply to spread the probability evenly over each range of values



FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L3/17

Example of fitting (3)

- The other option is to fit a smooth density function
  - Of a standard form
- The graph shows three different distributions fitted to the expert's judgements
  - The blue and green are almost identical
  - The red line is actually not such a good fit to the elicited judgements



Beta, gamma and lognormal fitted distributions

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L3/18

Multiple experts

- The case of multiple experts is important
  - We generally want to use the skill of as many experts as possible
- Two approaches to get a single distribution to represent their combined knowledge
  - Mathematical aggregation (pooling)
    - Elicit a distribution from each expert separately
    - Combine them using a suitable formula
    - Experts can be weighted
  - Behavioural aggregation
    - Get the experts together
    - Elicit a single 'consensus' distribution

See Lectures 8-11

GD 2.4.5

Summary

- Probability judgements may be affected by biases
  - Arising from psychological heuristics
    - Anchoring and adjustment, availability, overconfidence etc.
  - EKE methods are designed to minimise these effects
- Specifying a distribution involves two steps
  - Elicit a few judgements
    - Usually of quantiles
  - Fit a distribution to those judgements
    - Histogram or smooth density
- We generally have multiple experts
  - Need to aggregate to obtain a single distribution
    - Mathematical or behavioural
  - Main source of differences between EKE methods

Exercise 3

- This is a taster for the coming Practical 3
- Our parameter is the average journey time for a shuttle to Milan Linate airport, leaving EFSA offices in Parma at 16:00

References

- EFSA, 2014. Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment. EFSA Journal 2014;12(6):3734.
- O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E. and Rakow, T. (2006). Uncertain Judgements: Eliciting Expert Probabilities. John Wiley and Sons, Chichester. 328pp. ISBN 0-470-02999-4.

Practical 3: Probabilistic expert judgements

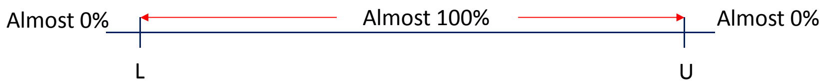
Tony O'Hagan,
John Quigley

Objectives

- To understand the judgements that experts are asked to make
- To practice making those judgements in a realistic setting
  - 5th and 95th percentiles and median (Cooke approach)
  - Credible range, median and quartiles (Sheffield and EFSA Delphi approaches)

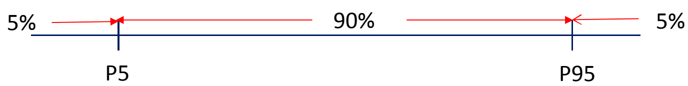
Credible range L to U

- Expert is asked for lower and upper credible bounds
  - Lower bound L, p ‘almost 0%’
    Upper bound U, p ‘almost 100%’
  - Expert would be very surprised if X was found to be below L or above U
    - It’s not impossible, just highly unlikely
  - Practical interpretation might be p = 1% and p = 99%



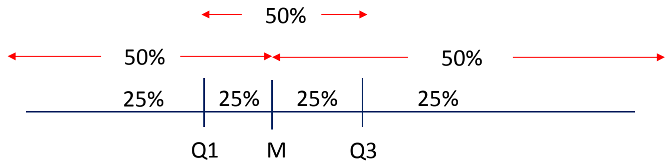
90% range P5 to P95

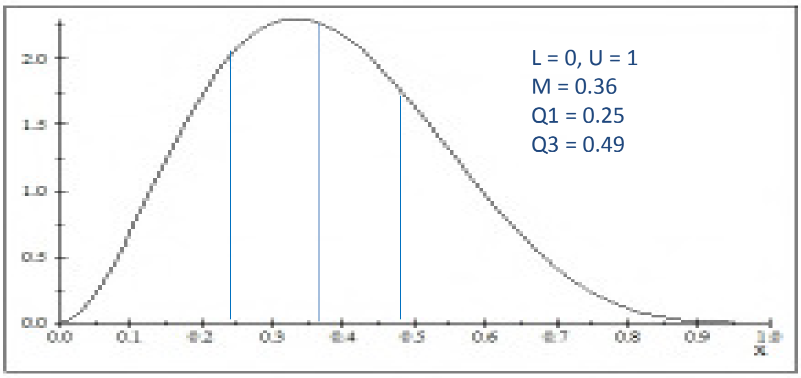
- A similar judgement, the 5th and 95th percentiles
  - Expert should judge that there is only a 5% chance (probability 0.05, or one chance in 20) that X lies below P5
  - And also a 5% chance that X lies *above* P95
  - So a 90% chance that X lies within this 90% range
    - Probability 0.9, or highly likely
    - But should not be as wide as the credible range



Median M and quartiles Q1 and Q3

- M is the 50th percentile
  - The value of x for which the expert judges X to be equally likely to be above or below x
- Q1 is the 25th percentile and Q3 is the 75th percentile
  - X is equally likely to be in any of the four sections:
    - below Q1, Q1 to M, M to Q3, above Q3





- Note:
  - M is not generally in the middle of the range
  - Q1 and Q3 are generally closer to M than to U or L

TASKS

Task 3.1

- Parameter X1 is the average journey time (minutes) for a shuttle to get to Milan Malpensa airport if it leaves EFSA headquarters in Parma at 16:00
- *Write down your judgements of the 90% range and median for X1*
  - The question sheet has a summary of definitions and advice for these judgements

Task 3.2

- Parameter X2 is the average number of people sitting in the cafeteria in EFSA headquarters at 13:00 (averaged over all working days in 2014)
- *Write down your judgements of the credible range, the median and quartiles for X2*
  - The question sheet has a summary of definitions and advice for these judgements

Practical 3 – Task 3.1

The parameter X1 is the **average** time (minutes) for a shuttle to get to Milan Malpensa airport if it leaves EFSA headquarters in Parma at 16:00.

For this task, you will specify your median and your 90% range for X1. Please read the notes carefully before making your judgements.

Notes:

1. Remember that X1 is the **average** journey time, **averaged** over all journeys to Malpensa leaving at 16:00 on any working day in the year.
2. The median value M is such that you think it equally likely that X1 will be above M or below M. It is a kind of estimate of X1, but an estimate with this specific meaning that you judge there to be a 50% chance that the average journey time is shorter than M and a 50% chance that it is longer than M.
3. The 90% range has a lower limit P5 and an upper limit P95. You should feel 90% certain that X1 will be between P5 and P95. (Again, **remember** that you are expressing uncertainty about the **average** journey time, **not** a single journey.) You should feel that there is a 5% chance (one in twenty) that X1 is below P5 and a 5% chance that it is above P95.

P5 = ………………………….. (minutes)

M = ……………………………. (minutes)

P95 = ……………………….. (minutes)

Practical 3 – Task 3.2

The parameter X2 is the **average** number of people sitting in the cafeteria in EFSA headquarters at 13:00 (averaged over all working days in 2014)

For this task, you will specify your credible range, median and quartiles for X2. Please read the notes carefully before making your judgements.

Notes:

1. Remember that X2 is the **average** number of people, **averaged** over all working days in the year.
2. The credible range has a lower limit L and an upper limit U. You should feel that it is extremely unlikely (but not impossible) that X2 wold be less than L or more than U. If someone were to tell you that X2 really was below L, or above U, you would think that they had made a mistake.

L = …………………………….

U = ……………………………

Notes:

3. The median value M is such that you think it equally likely that X2 will be above M or below M. It is a kind of estimate of X1, but an estimate with this specific meaning that you judge there to be a 50% chance that the average number of people is less than M and a 50% chance that it is more than M.

M = …………………………….

Notes:

4. The lower quartile Q1 is a value between L and M such that you believe that X2 is equally likely to be in the range [L to Q1] or in the range [Q1 to M]. Similarly, Q3 is a value between M and U such that you believe that X2 is equally likely to be in the range [M to Q3] or in the range [Q3 to U].
5. Also, you should feel that X2 has a 50% chance of being between Q1 and Q3 (and a 50% chance of being outside this range)

Q1 = …………………………..

Q3 = ……………………………

Lecture 4: Identifying priority parameters for EKE

Andy Hart & Tony O'Hagan

Outline

- The need for prioritisation
- How to identify parameters for which EKE is not necessary
- Minimal assessment of lower-priority parameters
- Sensitivity analysis to prioritise important parameters

The need for prioritisation

- Formal, rigorous EKE demands non-trivial resources
- It is neither feasible nor necessary conduct full EKE for every parameter in a risk model.
- Therefore, *prioritisation is needed*
  - a task for the **Working Group**

Often, a small number of parameters are responsible for most of the uncertainty

GD 3.3.1

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L4/3

Identifying parameters for which elicitation is not necessary

- Value known, or with negligible measurement error

Uncommon?

- Parameters where uncertainty is well-quantified
- Parameters where uncertainty requires small inflation

GD 3.3.1

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L4/4

Parameters for which uncertainty is well-quantified

- For example:
  - A measurement with known accuracy
  - Estimate from statistical analysis or meta-analysis of data, with standard error
- Assign appropriate distribution (often Normal) with:
  - mean = estimate from data/analysis
  - standard deviation = standard error of estimate

GD 3.3.1

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L4/5

Uncertainty requires small inflation

- Value or uncertainty is known for a related parameter (e.g. for a related chemical, pathogen, or scenario)
- Extrapolation between the parameters can be covered by adding a degree of extra uncertainty, s
- This can be estimated by informal expert judgement *provided* s is much smaller than the uncertainties about other parameters in the model

GD 3.3.1

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L4/6

Uncertainties for which EKE is desirable

- Often, most parameters won't meet the requirements described above
- Full EKE is *desirable* for the remaining parameters but usually *not feasible for all of them*
- …so we need a *strategy to prioritise them*

GD 3.3.1

Strategy recommended in Guidance Document

1. Carry out minimal assessment for all parameters requiring EKE *(GD 3.3.2)*
2. Evaluate their relative importance by sensitivity analysis *(GD 3.3.3)*
3. Submit the most important to formal EKE
4. Use the minimal assessment of uncertainty for the others

Minimal assessment

- WG assigns a probability distribution by a simplified expert judgement process:
  - consider the evidence for the parameter and select a best estimate, m
  - select a margin of error, s, for that estimate
    - such that the true value is *at least twice as likely* to lie in the range from (m – s) to (m + s), as outside that range
    - if appropriate, make s assymmetric (s_{up}, s_{down})
  - choose an appropriate distribution, e.g. Normal or Gamma, with mean m and standard deviation s
- These should be *careful*, *consensus*, *documented* judgements of the WG

GD 3.3.2

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L4/9

Minimal assessment is a crude procedure

- Relies on WG expertise and ignores the generally accepted principles of good practice for EKE
- *Adequate in practice provided the more important parameters have been submitted to full formal EKE*
  - so that minimal assessment parameters make only a small contribution to overall uncertainty

GD 3.3.2

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L4/10

Role of minimal assessment

- *Minimal assessment will be an intrinsic part of most risk assessments*
- *BUT 'it should never be acceptable to use minimal assessment for all parameters in the risk assessment'*
  - 'if full EKE is not done for at least some parameters, then the risk assessment should be qualified by a statement that the conclusions may not be robust to unquantified uncertainty in model parameters'

GD 3.3.2

Sensitivity analysis

- Use *sensitivity analysis* to prioritise parameters for full EKE
  - identifies those which contribute most to the uncertainty of the risk assessment outcomes
- Doing this by expert judgement is unreliable because the importance of a parameter depends on two factors:
  - how strongly it features in the model
  - the amount of uncertainty regarding the parameter

GD 3.3.3

One-way sensitivity analysis

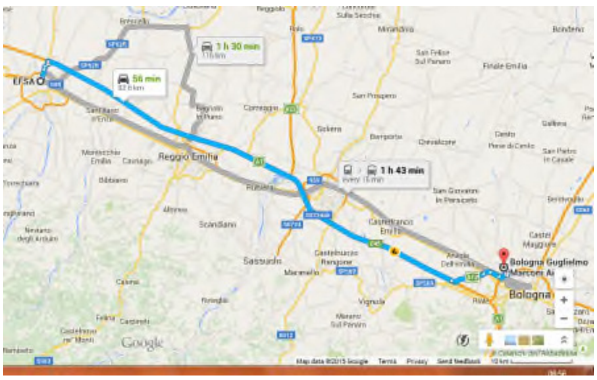
- Use the m and s values from minimal assessment
- Compute the output of the risk model twice, with:
  - parameter X set to (m – s) and all others set to their m
  - parameter X set to (m + s) and all others set to their m
  - the *measure of importance* for X is the difference between the two output values
- Repeat for every parameter
- *Rank the parameters in order of importance*
  - Elicit the most important by full EKE
  - Use minimal assessment distributions for the rest

GD 3.3.3

Example – driving time to Bologna *if leave at 6pm*

- A simple model: $Time_{total} = T_{stop} + \frac{D_{slow}}{S_{slow}/60} + \frac{(D_{total} - D_{slow})}{S_{fast}/60}$
  - Total time (minutes) = time stopped, time at slow speed and time at high speed

Which parameters are most uncertain?



Other methods for sensitivity analysis

- Methods of sensitivity analysis are reviewed in Frey and Patil (2002), Oakley and O'Hagan (2004)
- In more complex models it is preferable to use probabilistic sensitivity analysis (Saltelli et al., 2000)

GD 3.3.3

Summary

- Not practical to conduct full EKE for all parameters
- Identify parameters for which EKE is unnecessary
- Carry out minimal assessment for the rest
- Evaluate their importance by sensitivity analysis
- Submit *at least* the most important to formal EKE
- Use the minimal assessment for the others
  - Should be a part of most risk assessments

References

- EFSA, 2014. Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment. EFSA Journal 2014;12(6):3734.
- Frey HC and Patil SR, 2002. Identification and review of sensitivity analysis methods. Risk Analysis 22, 553-578.
- Oakley JE and O'Hagan A, 2004. Probabilistic sensitivity analysis of complex models: a Bayesian approach. Journal of the Royal Statistical Society, B 66, 751–769.
- Saltelli A, Chan K and Scott EM, 2000. Sensitivity Analysis. John Wiley and Sons.

Practical 4: Identifying priority parameters for EKE: sensitivity analysis

Tony O'Hagan,
Andy Hart

Objective

- Practice prioritisation of parameters for EKE, including:
  - minimal assessment
  - one-way sensitivity analysis

Practical task 4.1

- Form a group with other participants who chose the same practical example
- Review the risk assessment scenario and consider the listed parameters
  - Discuss the summary of evidence provided for the parameters and their role in the risk assessment
  - *Consider which parameters are likely to be prioritised for EKE*
    - This is an informal judgement
    - Involves both how influential a parameter is expected to be in the RA and how uncertain it is

Practical task 4.2

- Apply the technique of minimal assessment
  - Apply to each parameter in turn
    - Don't take too long, this is supposed to be a rough assessment
- Apply one-way sensitivity analysis using the minimal assessments of uncertainty
  - *Which would be your top priority for EKE?*
  - *How does it differ from expectations in Task 4.1?*
- Reflect on the value of formal sensitivity analysis
  - Does it seem to be an effective prioritiser?
  - Even when based on crude minimal assessment and one-way SA?

Examples for practical sessions

- Chemical risk – dermal exposure to bisphenol A
- Nutrition – Vitamin B12 requirement
- Environmental risk – GM pollen transport
- Plant health – Citrus Black Spot entry pathway
- Animal Health – Rift Valley Fever
- Biohazard – Ebola in bushmeat

Note: Examples have been simplified for purpose of training course

Lecture 5: Specifying questions for EKE

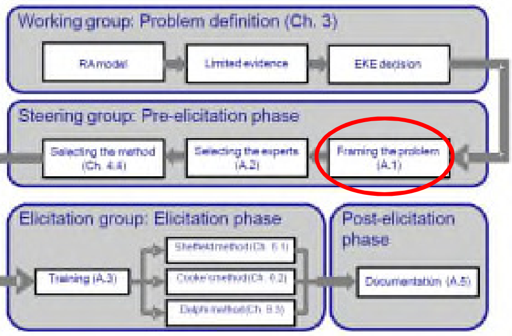
Andy Hart

Outline

- Roles
- Need for precise specification
- Challenges
- Choice of scale
- Uncertain variables
- Examples

Roles

- Working Group prioritises parameters and decides which to submit for EKE
- *Steering Group defines the precise questions to be asked about these parameters*



From Lecture 1

Precise specification of questions

- Need to *specify in precise terms*:
  - Parameter to be elicited
  - Metric, scale and units
  - Familiar to experts
  - Usable in risk assessment
  - Spatial and temporal context/scope (when and where)

Is this a well-specified question?

- What will be the exchange rate for euros and dollars next year?

GD 4.1

Precise specification of questions

- The quantity should be *in principle observable*
  - if suitable observations could be made, the outcome would be determined *unambiguously*
  - i.e. if betting on the outcome, you would know who had won the bet
- Advantageous if observable *in practice*
  - → opportunity to update and calibrate later

Is this a well-specified question?

- What will be the exchange rate for euros and dollars next year?

GD 4.1

Precise specification of questions

- If the question is *NOT well specified* then:
  - different experts interpret it differently
  - their answers may be inappropriate for use in the assessment
  - the relevance of future observations will be unclear

This question is NOT well-specified:

- What will be the exchange rate for euros and dollars next year?

GD 4.1

Specifying elicitation questions is a major task

- Highly interactive within the Steering Group
- Needs input from:
  - the substantive scientist on the problem,
  - a person knowledgeable about elicitation to find possible question formats,
  - administrative staff to decide on resources (e.g. timeline, possible number of experts, possible number of questions, etc.)

GD 4.1

Defining elicitation questions is a major task

- May require revising the structure of risk assessment model (requires consultation with Working Group)
- Test the draft questions on selected experts (e.g. WG)
- Finalise question after protocol selected, Elicitation Group appointed and experts recruited

GD 4.1

Scales for elicitation questions

- EFSA EKE Guidance uses quantitative scales
- Experts sometimes baulk at using numbers to express uncertainty
  - difficulty/unfamiliarity with quantitative expression
  - concern that numbers imply unjustified precision
  - concern that numerical estimates may be misinterpreted by decision-makers, stakeholders and the public
- …and may just prefer to be qualitative

GD A1.3

Questions about *variable* quantities

- Time taken to travel from EFSA to Linate by shuttle VARIES from trip to trip
- How does this affect framing for EKE?

From Lecture 3

Questions about variable quantities

- Many quantities in risk assessment are variable
  - Over time, over space, between individuals, etc.
- Need to specify what is relevant for the risk assessment:
  1. A particular instance of the quantity?
  2. The whole population of possible instances?
  3. A particular subset of the population?
- If (1), specify which particular instance is required
- If (2) or (3) then we need to:
  - Specify the population or subpopulation of interest
  - Elicit judgements about parameters that quantify the variability, e.g. median and ratio of P90/median

See also Lectures 2 and 13 - Not discussed in GD

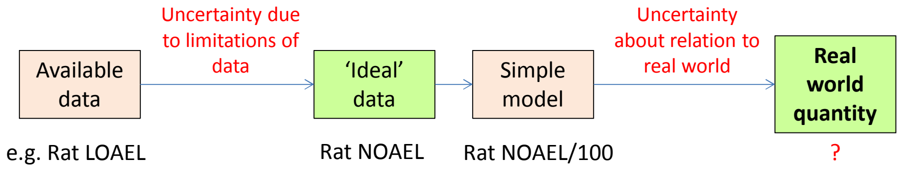
Consider at what level to pose the question

- Many questions can be broken into sub-questions
  - E.g. distance to Linate & driving speed
- Some questions require more than one type of expertise, from different experts
- Breaking into sub-questions may help – but involves more work *– more in Lecture 13*

GD A1.2

Question-framing in simple risk models

- Many EFSA assessments use simple deterministic calculations with conservative scenarios & assumptions

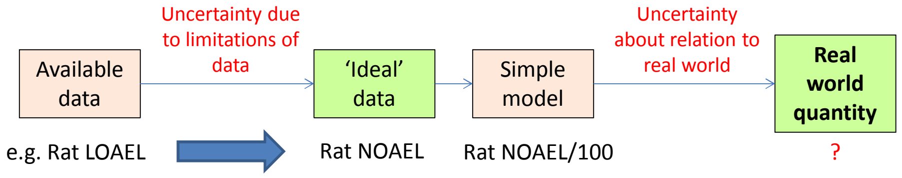


- WG may want to address *uncertainties affecting the data*
- *Relation of parameters to real world* often not defined
  - so it's *difficult to frame precise questions* for EKE in terms of observable outcomes in real world

(Not in EKE guidance – see Draft GD on Uncertainty)

Question-framing in simple risk models

- *Short-term solution*: focus on uncertainty of inputs
  - frame questions in terms of *what ideal data would be*, e.g. the NOAEL from a new, good quality study



- *Longer-term*: define relation of parameters to real world
  - may require revision of assessment procedure

(Not in EKE guidance – see Draft GD on Uncertainty)

Supporting information for questions

- Accompanying information should include:
  - the purpose of the elicitation task
  - any assumptions that are being made concerning the problem/model (with justifications)
  - how the output from the exercise will be used
  - the Evidence Dossier - *see Lecture 7*
- Ensure problem is framed consistently for all experts
- Elicitor must take care to avoid introducing inappropriate or inconsistent information

GD A1.2

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L5/15

Summary

- Defining questions for EKE is a major task
- They should:
  - have clearly specified metric, scale and units
  - be quantitative where possible
  - be adapted to the experts' language
  - be in principle observable
  - handle variability appropriately
  - be broken into subquestions where helpful
  - be accompanied by supporting information

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L5/16

References

- EFSA, 2014. Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment. EFSA Journal 2014;12(6):3734.
- EFSA, 2015. Draft Guidance on Uncertainty in EFSA Scientific Assessment. Draft published for Public Consultation.

Practical 5:
Specifying questions for EKE

Andy Hart

Objective

- Practice specifying questions for EKE

Practical tasks

- Work with the same group as for Practical 4
- Specify a suitable question for eliciting judgements about the parameters
  - *Start with the first parameter in the list*
  - If time permits, continue with the parameters you prioritised for EKE
- *Bring your finished questions back to plenary on flip charts*

List of parameters

Case study	Parameter name	Description
PLH	Proportion infected	Proportion of fruits which are infected
AHAW	Volume	Number of animals to be transported from endemic countries to the RC.
BIOHAZ	Amount imported	Amount of bushmeat illegally imported into Europe
CHEM	Number of fingers	Average number of fingers that touch receipts during handling
ENV	Wind and rain	Effect of wind and rain on pollen concentrations
NDA	Growth factor	Additional cobalamin requirement for growth

Lecture 6: Identifying, selecting, motivating and training experts for an elicitation

John Quigley,
Fergus Bolger and Simon French

Outline

- The nature of expertise
- Identifying experts
- Tasks of the Steering Group:
  - Expert roles and profiles
  - How many experts?
  - Long-listing
- Tasks of the Elicitation Group:
  - Screening and short-listing
  - Creating heterogeneity
  - Inviting the experts
  - Expert motivation and retention

Note: identifying expertise needed and recruiting the experts requires *your* judgement and expertise

Who is a suitable expert to assess risk of salmonella poisoning at fast food outlets?

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L6/3

What are the pros & cons of each?

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L6/4

Types of expertise

Think of Questions …

What …?
Who …?
When …?
Why …?
How …?
…..

GD A.2.2.6

Types of expertise

→ Different types of knowledge

Know - What
Know - Who
Know - When
Know - Why
Know - How
….. …..

Use these as prompts to think about the sorts of knowledge and expertise that you will need

The Guidance Document gives formal terms and definitions for different types of knowledge

GD A.2.2.6

SG: Recruitment stage I

- The role of the Steering Group (SG):
  - refines parameters to be elicited
  - identifies required expert knowledge and relevant roles

    → *profile matrix*
    - Helps identify expertise needed

Example: risk of salmonella poisoning at fast food outlets

GD 4.2. & A.2.2.6

Example: expertise profile matrix

Knowledge requirements			Country	Expert Roles			
				Industry		Govt (Inspector)	Academia (Scientist)
Substantive Expertise	**Importance**	**Specificity**		Supply chain	Production		
Immunity to levels of salmonella	Essential	Specific	AA				
			BB				
		Comparable	AA				
			BB				
Quality of Food in supply chain	Essential	Specific	AA				
			BB				
		Comparable	AA				
			BB				
Conditions of fast food kitchen	Essential	Specific	AA				
			BB				
		Comparable	AA				
			BB				
Standard contamination metrics	Desirable	Specific	AA				
			BB				
		Comparable	AA				
			BB				
Expressing risk and uncertainty as probability	Desirable	Specific	AA				
			BB				
		Comparable	AA				
			BB				

Note: this is an example, *not* a template! Design the matrix for your context

SG: Recruitment stage I

- *Profile matrix*
  - Helps identify expertise needed
  - Can help determine how many experts required
  - Can help determine elicitation method
    → creates Elicitation Group
- Be aware that experts can have specific knowledge relevant to your study
  - E.g. know about precise issue in region of concern

  or comparable knowledge
  - E.g. know about similar issue in similar region

GD 4.2. & A.2.2.6

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L6/9

Identifying experts

- Often useful to have all types of expertise in a group
  - Creates challenge: explicit or implicit
- EFSA has expertise database to give you a good start
  - Ask colleagues
  - Ask experts who you have identified for other experts
  - Etc. – see guide for suggestions.
- Can include experts who would not normally be eligible for EFSA Working Groups, e.g. industry
- Expert availability/diaries and politics may be major constraints

GD A.2.2.6

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L6/10

SG: How many experts?

- Profile matrix gives general indication of the number required *but*:
  - not all cells may need filling
  - may be quotas (e.g. member-state representation)
  - You can sample down columns and across rows to ensure 'even' coverage
- The more experts the better?
  - Increases reliability

  but
  - practical problems → choice of elicitation method
  - trade-off quantity vs. quality
  - broad vs. deep approaches
  - diminishing returns
- May wish to limit to between 8 and 15 (min. 5)
- Important to *over-recruit* initially
- Adding experts little use if expertise is *homogeneous*

GD 2.3.5. & A.2.4.2.

EG: Screening

- May wish/need to cut down long list → short list
  - EFSA guidelines: develop a questionnaire to assist this
- Relevant considerations include:
  - Job description – title, expertise area, years experience, practical vs. theoretical etc.
  - Experience of making judgements:
    - amount of judgment vs. data and models
    - feedback on accuracy
    - data availability and quality
    - nature and experience of judging probability and risk
  - Training received → training needs
    - Training in expressing uncertainties as probabilities
    - e-learning material

GD A.2.2.7.

EG: Inviting the experts

- Inviting the short list – points for invitation letter:
  - What? Nature of problem & motivation for EKE
  - When? Timing of EKE, key dates
  - Why? Reasons expert selected
  - Where? Venue for elicitation(s)
  - How? EKE procedure and meeting agendas
- Also information about:
  - Constitution of expert group
  - Confidentiality and anonymity
- Additional information to and from experts?
  - e.g. their concerns, conflicts of interest
  - EFSA-run EKE: DOI* required but not evaluated
  - Out-sourced EKE: DOI not required

** DOI: Declaration of Interests*

GD 5.1. & A.2.3.

EG: Motivating and retaining experts

- Often the entire EKE exercise will take many months and substantial expert commitment
- Important to keep experts 'on board' through provision of e.g.:
  - regular feedback re. use of their expertise
  - positive reinforcement & incentives
  - information regarding progress of EKE exercise
- May also want to use experts again in future:
  - 'exit' questionnaire
  - retention of CV's, questionnaires?
  - keeping information about non-responders

GD A.2.3.3. & A.2.3.4.

Summary

- The expert recruitment *process*
- Tasks of the Steering Group:
  - Expert roles and profiles
  - How many experts?
  - Long-listing
- Tasks of the Elicitation Group:
  - Screening and short-listing
  - Creating heterogeneity
  - Inviting the experts
  - Expert motivation and retention
- But wherever you sit in the recruitment/management process, you will have to use your judgement

Reference

- EFSA, 2014. Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment. EFSA Journal 2014;12(6):3734.

Practical 6: Identifying, selecting, motivating and training experts for an elicitation

Fergus Bolger & Simon French

Work in groups with the same example parameters you chose in Practical 5

Objective

- Practice developing an expertise profile

Task: develop an expertise profile

- Work in the same groups as for Practical 5
- Focus on the first parameter in the list for your example
- Consider what types of expertise would be essential and desirable for an EKE of this parameter
- *Draw up an expertise profile table on a flip chart, in suitable format*
- *Decide how many experts you need*

List of parameters

Case study	Parameter name	Description
PLH	Proportion infected	Proportion of fruits which are infected
AHAW	Volume	Number of animals to be transported from endemic countries to the RC.
BIOHAZ	Amount imported	Amount of bushmeat illegally imported into Europe
CHEM	Number of fingers	Average number of fingers that touch receipts during handling
ENV	Wind and rain	Effect of wind and rain on pollen concentrations
NDA	Growth factor	Additional cobalamin requirement for growth

Lecture 7: The Evidence Dossier

Tony O'Hagan

What is the dossier?

- A document summarising
  - the evidence regarding each parameter to be elicited
  - based on the researches of the Working Group
    - possibly supplemented by work of the Steering Group
    - and also possibly with additional evidence from experts
  - for use by the experts when making judgements
- The Guidance document does not present clear guidance about the dossier
  - Material in this lecture therefore extends the Guidance
  - Based on the experience of the course presenters

This lecture extends and updates GD 6.1.2, 6.2.1

Importance of the dossier

- RA should, as far as practicable, be evidence based
- Experts' judgements should differ only because of their expertise and interpretation of the evidence
  - Not from having different data
  - Aggregation is otherwise much less reliable/effective
- Availability heuristic makes it important to review all the evidence together

We need to assemble the evidence

Evidence should be shared

It should be available to experts during elicitation

Assembling the dossier

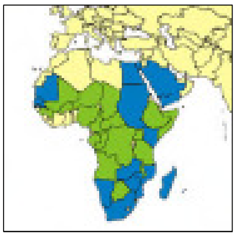
- Steering Group prepares a first draft dossier
  - The Working Group will already have done some review of literature and other evidence
  - This should be revisited for parameters chosen for the RA
    - Don't include minimal assessments!
  - Possibly with additional research undertaken
- Elicitation Group sends this out to experts before the elicitation, with request for them to identify omissions
  - As part of their preliminary briefing
  - Deadline for responses at least several days before elicitation
  - If experts bring up new evidence at the time of the elicitation it needs careful scrutiny!
- Final dossier incorporates new evidence from experts
  - Made available to experts for use in the elicitation

Writing the dossier

- The dossier should *summarise* the principal relevant evidence
  - Not too long
    - Otherwise it's hard for experts to assimilate all the evidence when making their judgements
  - Point out weaknesses
    - Sample size, sampling/experimental technique
    - Parameter relates to different region/species/duration/age/etc.
  - Tabular form recommended
- Reference list itemises *all* relevant sources

Examples

- Rift Valley Fever



- *Pomacea*

Summary

- The evidence dossier is an important mechanism to ensure that all relevant information is assembled and is available to all the experts during elicitation
- It presents a summary of the most important evidence
  - With references
- It is assembled by Steering Group
  - Based on initial evidence review by Working Group
  - Supplemented where appropriate by additional research
  - And including any new evidence submitted by experts prior to the elicitation

Lecture 8: The Sheffield method

Tony O'Hagan

Outline

- Key features of the Sheffield method
  - Behavioural aggregation
  - Meaningful result
  - SHELF structured process
- Pre-elicitation
  - Selection of experts and elicitor
  - Preparation for workshop, timescale
- Elicitation
  - Individual judgements and the SHELF system
  - Consensus judgements and managing experts
  - Roles of elicitor and recorder
- Post-elicitation
  - Documentation
  - Review

GD 6.1

KEY FEATURES

Behavioural aggregation

- The Sheffield protocol uses behavioural aggregation
  - Experts meet together in an *elicitation workshop*
- Both individual and group judgements are made
  - Experts make initial probability judgements individually
  - Discuss differences of opinion, reasons, etc.
  - The group makes consensus judgements
  - Aggregate distribution is fitted to those judgements
- Information and interpretations of evidence are shared
  - Making best possible use of both evidence and expertise

GD 6.1.1, 6.1.4

Meaningful result

- Nature of group 'consensus' judgements
  - Group asked to make judgements that would be reasonable for a *rational, impartial observer (RIO)*
    - Having seen the experts' individual judgements
    - And having heard their opinions and the discussion
- Nature of the resulting aggregate distribution
  - Has a *genuine interpretation as a subjective probability distribution*
    - Unlike the results of mathematical aggregation
    - The viewpoint of a rational, impartial observer is what EFSA needs

GD 6.1.4

SHELF structured process

- The elicitation follows the SHELF system
  - Experts are asked to make judgements following a strict sequence and framing
  - *Designed explicitly to avoid/minimise biases from common heuristics*
  - Guidance notes for elicitor
  - Software for fitting distributions
- SHELF is a widely used and stable system

GD 6.1.1, 6.1.3, g.1.4

PRE-ELICITATION

Selection of experts

- Aim for 4 to 8 experts in a workshop
  - Too many will result in unnecessary discussion
    - Avoid duplication of opinions
  - With too few experts we may not cover the range of opinion
- Avoid hierarchies
  - Junior people will tend to defer to seniors if present
  - Aim for a discussion among equals
- Avoid including people for 'political' reasons
  - Experts should not be included simply as representatives
    - Of some nation, stakeholder, pressure group, etc.
    - They are often ignorant of, or unwilling to give serious consideration to, the opinions of others
  - Their input can be sought as evidence
    - To be included in the evidence dossier
    - But not as members of the elicitation workshop

GD 6.1.2

The elicitor / Elicitation Group

- Steering Group's appointment of the Elicitation Group is a critical task
  - Conducting an elicitation workshop is a skilled job
    - The elicitor has to manage the group of experts
      - See discussion later
    - Many of whom will be strong-willed or opinionated
  - EG will need substantial experience in facilitating behavioural aggregation
    - Preferably using the Sheffield method

GD 6.1.2, 6.1.6

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L8/9

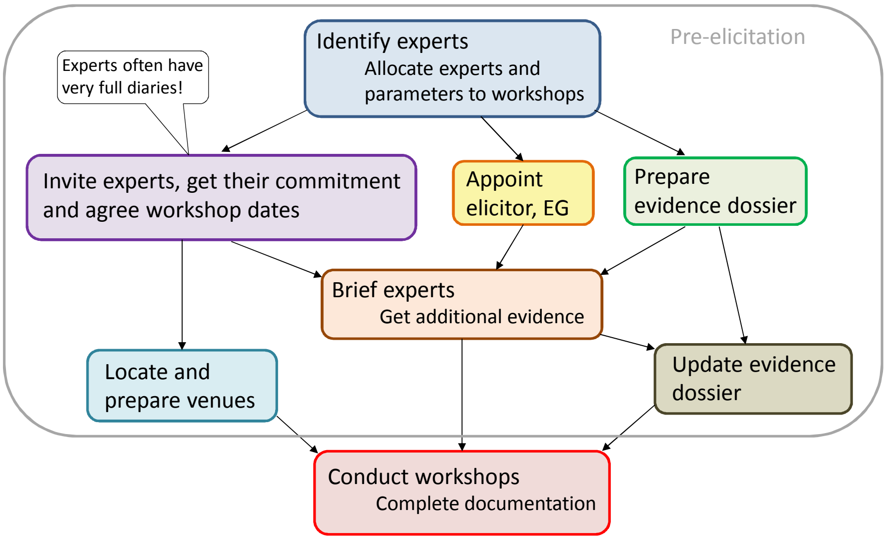
Prepare workshops

- In one workshop of 1 or 2 days, we can expect to elicit distributions for 2 to 5 parameters
  - Each parameter should be such that all the experts in the workshop can contribute usefully to the elicitation for that parameter
  - More than one workshop may be required if elicitation is needed for many parameters, or sufficiently diverse parameters
- Workshop venue
  - One suitably large room with good facilities
    - Boardroom style layout
    - Data projector and screen, flipchart, tea/coffee etc.

GD 6.1.2

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L8/10

Timescale flowchart



Extends GD 6.1.2,

ELICITATION

Workshop flowchart

Introductions
Roles, review purpose of workshop, etc.
- Completion of SHELF1 form – see later

Training
Review principal ideas from online e-learning course
Ideally, run a practice elicitation
- Partly for the experts to practice the skills learnt online
- Partly so that they see how the Sheffield process works
- Particularly group discussion, consensus judgements

Parameter 1

Review evidence

Individual judgements
Credible range, median and quartiles
- Privately without discussion

Discussion
Individual judgements are revealed and discussed
- Particularly high, low, wide or narrow distributions challenged and justified
- General discussion of reasons for individuals' judgements

Group judgements
Group 'consensus' judgements are made
- The rational, impartial observer

A probability distribution is fitted
- Feedback and opportunity to revisit judgements
- Confirmation of final elicited distribution
- Completion of SHELF2 form – see later

Parameter 2

Review evidence

....

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L8/13

The SHELF1 form

- At the beginning of a workshop this form is completed
- It records basic information
- Note
  - This record
  - Orientation and training
  - Declarations of interests
  - Evidence
  - Structuring
  - Definitions

ELICITATION RECORD – Part 1 – Context

Elicitation title	
Session	
Date	
Part 1 start time	

Attendance and roles	
Purpose of elicitation	
This record	Participants are aware that this elicitation will be conducted using the Sheffield Elicitation Framework, and that this document, including attachments, will form a record of the session.
Orientation and training	
Participants' expertise	
Declarations of interests	
Strengths and weaknesses	
Evidence	
Structuring	
Definitions	

GD 6.1.3

Part 1 end time	
Attachments	

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L8/14

The SHELF2 form

- This form is completed for each elicited parameter
- It provides a record of the elicitation
- There are two judgement phases
  - Individual judgements
  - Group consensus judgements

GD 6.1.4

ELICITATION RECORD – Part 2 – Distribution

Quartile Method

Elicitation title	
Session	
Date	
Quantity	
Start time	

Definition	
Evidence	
Plausible range	
Median	
Upper and lower quartiles	
Fitting	
Group elicitation	
Fitting and feedback	
Chosen distribution	
Discussion	

End time	
Attachments	

Example

- This example is based on a real case
  - A practice elicitation for training purposes
    - Eliciting beliefs about the population of Portugal
    - Write down your credible range now!
  - With three experts
  - And elicitor
  - [SHELF 2 Sample.pdf]

Individual judgements

Definition	⟸	Precise definition (Framing)
Evidence	⟸	Review the evidence (Availability)
Plausible range	⟸	Individual's plausible range (Anchoring, Overconfidence)
Median	⟸	Individual's median (Quantile judgement)
Upper and lower quartiles	⟸	Individual's quartiles (Quantile judgements, Anchoring)
Fitting	⟸	Sheffield method fits a full distribution

Each step guided by findings in psychology

GD 6.1.4

Fitting using MATCH

- The MATCH software is used in the e-learning course to fit a distribution to a set of judgements
  - http://optics.eee.nottingham.ac.uk/match/uncertainty.php
  - SHELF also has software to fit distributions
- Let's take a quick look
  - Using the judgements of a volunteer
  - About the height in metres of the EFSA building

Group discussion

- Individual elicitation is followed by group discussion
- Elicitor should:
  - Show the collection of fitted distributions and invite comments from outliers
  - Allow discussion all the while it seems to be developing ideas
    - But not let them keep going over old ground
    - Not aiming for complete consensus and 'agree to differ' is OK
  - Make sure all opinions are heard and properly considered
    - Keep bringing in the quieter members if necessary
    - Bear in mind each expert's expertise– what they bring to the group
    - Not allow ranting or lecturing
  - Listen carefully
    - Try to get a sense of the strengths of competing arguments

GD 6.1.4

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L8/19

Consensus judgements

- Elicitor now seeks group 'consensus' judgements
  - Median and quartiles
    - But could instead ask for specific probabilities e.g. P(X < 50)
    - No anchoring issues now
  - Not aiming for literal consensus
    - 'Agree to differ' is OK
    - And in fact inevitable
  - Asking experts to make reasonable judgements
    - Such as would be the opinion of a *rational, impartial observer*
    - They are the best people to do this
      - Assessing how much weight should be given to each argument

GD 6.1.4

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L8/20

Elicitor is not passive

- Elicitor should be prepared to challenge the experts' judgements
  - If those judgements don't look like those of RIO
  - For instance:
    - If the experts don't appear to have given proper consideration to an argument that sounded valid
    - If any expert's initial beliefs are not represented in the group judgements
      - With no obvious reasons for that expert to have changed
    - We don't want valid opinions to be overlooked or lost
      - Through force of personality
      - Or because an expert can't be bothered to keep arguing

GD 6.1.4, 6.1.6

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L8/21

Fitting and feedback

- Finally, a distribution is fitted to the consensus judgements
  - SHELF provides some software
  - Feedback is given
  - Experts confirm the final elicited distribution
    - Or may wish to revise and refit
  - Any final discussion is recorded

Group elicitation	
Fitting and feedback	
Chosen distribution	
Discussion	

GD 6.1.4

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L8/22

The roles of elicitor and recorder

- The preceding slides demonstrate the importance of the elicitor
  - It is a *skilled job*
  - Must manage experts and deal with psychological issues arising in a group
  - Must have respect of the experts
- To free the elicitor to concentrate on interaction with and between experts, it is important to have a *recorder*
  - Takes notes for completion of SHELF forms
    - Also a skilled job!
  - Runs software to fit distributions and provide feedback

GD 6.1.1

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L8/23

POST-ELICITATION

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L8/24

Documentation and review

- SHELF forms document the elicitation process
  - A firm basis for the formal EFSA reporting
    - See Lecture 12
- Parts may be completed 'live' in the workshop
  - But inevitably most is completed later from the recorder's notes
    - And the elicitor's memory!
- Essential for forms to be reviewed by experts to identify any significant errors/omissions
  - Constitutes formal post-EKE feedback to experts
    - See Lecture 12

GD 6.1.5

Summary

- Key features
  - Behavioural aggregation
  - Meaningful result
  - SHELF structured process
- Two rounds of judgement
  - Individual and group consensus (RIO)
- Importance of skilled elicitor
  - Managing the group discussion
  - Recorder is another important and skilled role
- SHELF documentation

GD 6.1

References

- EFSA, 2014. Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment. EFSA Journal 2014;12(6):3734.
- SHELF: http://tonyohagan.co.uk/shelf
- MATCH: http://optics.eee.nottingham.ac.uk/match/uncertainty.php

Practical 7: Key aspects of the Sheffield method

Tony O'Hagan

Objective

- To gain some practical understanding of issues around steering an expert knowledge elicitation using the Sheffield method

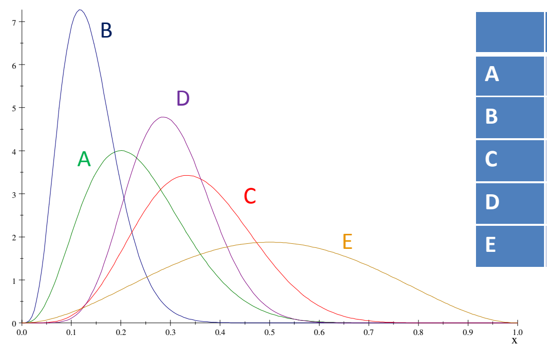
Practical task 7.1

- Work in the same group as for Practical 6
- You have just been appointed as members of the Steering Group for an EKE exercise
- Your constraints/resources:
  - WG has agreed that EKE is required for the parameter you considered in Practical 6
  - It is proposed to use the Sheffield method
  - You would like to have at least 5 or 6 experts to cover the full range of expertise and opinion needed
  - A quick brainstorming of possible experts has identified 10 good possibilities, all but one of whom is based in Europe
  - WG requires a report in 5 weeks

- *Develop a suitable timeline plan on a flip chart*
  - *With milestones*

Practical task 7.2

- Parameter x relates to a hypothetical assessment
- Five experts (A, B, C, D and E) have given the median and quartile judgements shown in the table
- The graph shows the five fitted distributions
- What are the principal differences between the experts' judgements?
- *As the elicitor, how would you lead a discussion towards reaching consensus judgements?*



	Q1	M	Q3
A	0.15	0.22	0.3
B	0.1	0.14	0.18
C	0.27	0.35	0.45
D	0.25	0.3	0.35
E	0.35	0.5	0.65

Lecture 9: Delphi Method

Fergus Bolger and Tony O'Hagan

Outline

- History and rationale
- The Delphi method
  - Procedures and tools
- Pre-elicitation concerns
  - Expert selection
  - Resource implications
- Elicitation concerns
  - Liaising with the elicitor
  - Managing interactions between elicitor and experts

History and rationale

- Designed by RAND Corporation in the 1950s to improve quantitative forecasting
- Uses groups to pool expertise but tries to minimize adverse group effects by:
  - restricting interpersonal interaction
  - controlling information flow
- Accentuates positive attributes of groups and downplays negative

GD 6.3. & A.4.4.

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L9/3

Principles of Delphi method

- Anonymity
  - Reduces social and political pressures that often emerge within interacting groups
    - Experts never meet
- Feedback
  - Of opinions of other experts from the “group”
  - Structured and controlled by the elicitor
  - Enables a small degree of expert “interaction”
- Iteration
  - Experts review their own opinions
    - Using information from their peers
  - Stop when no further revision
    - Or due to resource limitations or excessive drop-out
- Equal weighting of experts’ judgments

GD 6.3. & A.4.4.

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L9/4

How Delphi works

Improvement in accuracy over Delphi rounds comes about because more-expert panellists (the "hold-outs") maintain their judgment over rounds, whilst the less-expert panellists (the "swingers") alter their judgments towards the group average (Dalkey, 1975)

Assumes swingers are less expert than hold-outs

Studies support the advantage over traditional groups (in terms of increased accuracy) by 5 to 1

(Rowe & Wright, 1999)

Further improvement may be gained by exchanging rationales → *virtuous opinion change*

GD 6.3., A.4.4. & Bolger & Wright (2011)

Delphi and EFSA Delphi

- "Classic" Delphi
  - Paper survey, fixed rounds
    - Real-time web-based alternatives also used
  - Quantitative estimates plus "confidence"
    - And rationales
  - Feedback averages, individual estimates, rationales
- EFSA Delphi
  - E-mail or web-based survey, fixed rounds
  - Probabilistic judgements (median, quartiles, …)
    - And rationales
  - Feedback judgements, probability distributions, rationales

GD 6.3. & A.4.4.

Delphi exercise – round 1

- In what year did Karl Benz build his *Motorwagen*, the first modern (internal combustion engine) automobile?
  - Upper credible limit
  - Lower credible limit
  - Median
  - (Rationale for median)

- Make a copy to be collected
- NB this is simplified
  - just 3 estimates (usually 5 in EFSA Delphi)
  - rationale for median only

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L9/7

Step 1: Choose survey medium

- An advantage of Delphi is that experts do not need to be brought together in a single place or time
- Delphi can use web-based software or e-mail
  - Web-based
    - No software available presently for EFSA Delphi
  - E-mail
    - Excel template available for EFSA Delphi judgements
    - Experts must make judgements in defined sequence

GD 6.3.1.

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L9/8

Steps 2 & 3: Develop and pilot survey

- Develop the survey
  - Write an introduction to the survey
  - List all questions that need to be answered
    - Restrict to what can be achieved in about 30 minutes
  - Write a closure to the survey
- Pilot the survey
  - As experts do this remotely and individually it is important that they
    - Understand questions and how to respond
    - Do not find it too difficult and/or time consuming
  - Pilot subjects must be trained in probabilistic judgements

GD 6.3.2. & 6.3.3.

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L9/9

Step 4: Estimate timeline

- How long for each round? Typical timings:
  - Day 0: Distribute questionnaire by e-mail to experts
  - Day 7: First reminder e-mail sent out
  - Day 11: Second e-mail reminder sent out
  - Day 14: Telephone call to essential experts who have not returned their questionnaires
  - Day 16: Close of polling
- How many rounds?
  - Minimum 2
  - Maximum unknown
    - i.e. when stability occurs, but usually 3 is sufficient

GD 6.3.4.

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L9/10

Estimate timeline for whole survey

Step	**Estimated time needed**
Delphi round 1	**About 5–10 weeks**
1. Choose survey medium	1 day
2. Survey development	1–3 weeks
3. Pilot of survey	1-2 weeks
4. Estimate timeline	1 day
5. Survey out with expert participants	2–3 weeks
6. Data collation & analysis	1–2 weeks
7. Subsequent rounds each	**4-9 weeks**
8. Reporting	1-2 weeks
Total for a 3-round Delphi	**14-30 weeks**

Steps 5 & 6: Execute and analyse survey

- Delphi round 1
  - Send out questionnaire
    - With well-framed definitions and evidence dossier
  - Reminders of how to make judgements
    - Even though experts should have taken e-learning course!
  - Email and telephone reminders to get good response rate
- Analyse results and prepare feedback for next round
  - With few experts (e.g. fewer than 10), feed back individual distributions and rationales
  - With many experts, need to summarize e.g.:
    - Aggregate distribution
    - Means and ranges of judgments
    - Categorization of rationales
  - But summaries lose the link between rationale and answer

GD 6.3.5.

Delphi exercise – round 2

- Date of Benz's *Motorwagen*
  - Feedback is given from Round 1 regarding
    - Mean and range of the medians
    - Mean and range of the upper and lower credible limits
    - (Categorized rationales)
  - Please make new judgements of:
    - Upper credible limit
    - Lower credible limit
    - Median
  - The true value is revealed!

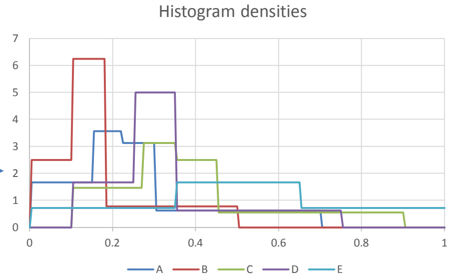
Step 7: Iterate

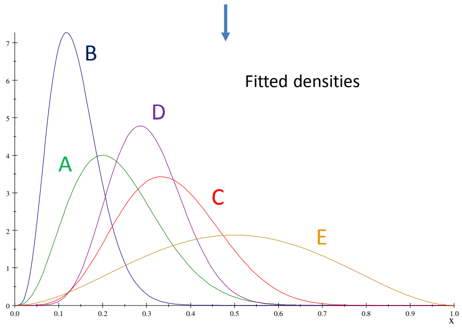
- Delphi Round 2
  - Feedback from Round 1
  - Repeat questions from Round 1
    - May be revised if significant issues identified at Round 1
  - Experts invited to consider feedback and revise their opinions if they wish
- Round 3 …
  - Repeat until experts stop revising
  - If resources available
- Final mathematical aggregation of experts' judgements
  - Equally weighted

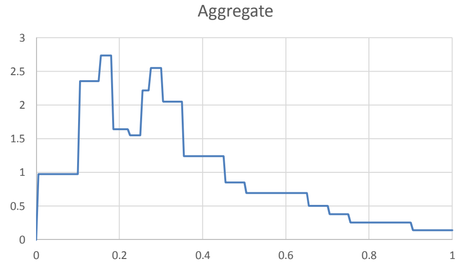
GD 6.3.7.

Aggregation example

	L	Q1	M	Q3	U
A	0	0.15	0.22	0.3	0.7
B	0	0.1	0.14	0.18	0.5
C	0.1	0.27	0.35	0.45	0.9
D	0.1	0.25	0.3	0.35	0.75
E	0	0.35	0.5	0.65	1







Step 8: Reporting

- Technical support document
  - expert training given
  - participant list (incl. pilot)
  - invitation and other letters
  - for all Delphi rounds
    - Delphi questionnaires (incl. pilot)
    - responses (‘raw data’) + analysis
  - final aggregate distributions
  - expert exit form and responses
    - includes appraisal of the EKE exercise

GD 6.3.8.

Pre-elicitation concerns

- Experts
  - The Guidance recommends a minimum of 5
    - Suggest maximum 10
      - So that feedback can show all judgements and rationales
  - Avoid duplication of opinions
    - Opinion change is greater if different perspectives
- Resource implications
  - Time-consuming (min. 1 month per round)
  - Need frequent reminders to/contact with experts to
    - Keep to timeline
    - Minimize dropout and missing data

GD 6.3..

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L9/17

Elicitation concerns

- The elicitor
  - Difficult to find anyone with experience of EFSA Delphi!
  - May need to appoint someone with experience of classic Delphi
    - Bring them up to speed with requirements of EFSA Delphi
    - This kind of expertise may be available in-house?
- Managing interactions between elicitor and experts
  - Who is responsible for recruitment and follow-ups to the experts?
    - EFSA staff or WG members may have more weight
  - Elicitor should be chiefly responsible for survey design and preparing feedback

GD 6.3.

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L9/18

Summary

- EFSA Delphi modifies classic Delphi to elicit probabilistic judgements
  - In the same sequence as Sheffield method
  - Very limited experience
- Two or more survey rounds
  - Experts get feedback, revise judgements iteratively
  - Allows limited passing of information between experts
- Planning, piloting and running survey is time-consuming
  - Even two rounds can take several months

References

- EFSA (2014). Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment. *EFSA Journal*, 12(6), 3734
- Bolger, F, & Wright (2011) Improving the Delphi process: Lessons from social psychological research. *Technological Forecasting and Social Change*, 78, 1500–1513
- Dalkey, N.C. (1975). Toward a theory of group estimation, in: H.A. Linstone, M. Turoff (Eds.), *The Delphi Method: Techniques and Applications*. Addison-Wesley: Reading, MA. 236–261
- Rowe, G & Wright G. (1999). The Delphi technique as a forecasting tool: Issues and analysis. *International Journal of Forecasting*, 15, 353–375

Practical 8: Key aspects of steering the Delphi method

Tony O'Hagan
Fergus Bolger

Objective

- To gain some practical understanding of issues around steering an expert knowledge elicitation using the EFSA Delphi method

Delphi practical preparation

- Here a parameter that you made judgements for earlier (in your course handbook, in Practical 3)

  Parameter X1 is the average time (minutes) for a shuttle to get to Milan Malpensa airport if it leaves EFSA headquarters in Parma at 16:00

- Working individually, retrieve your judgement of the median from Practical 3 and write a rationale for your judgement
- Now carry out the following tasks in your groups

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation P8/3

Practical task 8.1

- Using your group's experiences of writing a rationale, think of how experts will handle this task
  - What kind of rationales would be most useful?
  - What are the obstacles to experts responding well?
  - How could they be helped?
- *What guidance would you give to experts to help them in writing their rationales?*

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation P8/4

Practical task 8.2

- Refer to the handout sheet
  - This shows hypothetical round 2 feedback for the parameter X1
  - Plus some questions that the experts asked after receiving the round 2 survey
- Discuss
  - The quality of the experts' round 1 responses
    - Their judgements and rationales
  - How to handle experts who give poor responses or clearly have misunderstandings
  - The implications of their questions
- *Write on a flip chart the lessons learned from this example for running a Delphi EKE*

Delphi Practical (Task 8.2)

Example Round 2 feedback for the parameter X1 (elicited in Practical 1).

Expert	Lower	Median	Upper	Rationale
A	65	95	100	That is the airport to the south of Milan, isn't it (the nearest)? Usually it takes an hour or so, maybe a bit longer at that time of day.
B	90	100	180	I have only done that trip a couple of times. Once, I think was pretty quick, an hour and a half maybe. The other time there was an accident so it took nearly 3 hours, but that would be unusual.
C	100	120	150	I think that the fastest you could do it is about 100 minutes, and the longest it is likely to take is 150, so I am guessing the average is midway between these.
D	70	150	180	I only did this journey once and it took about 2 and a half hours, but there were road works and a lot of traffic, so normally I expect it is much, much quicker! (But possibly this is the norm – and it will be rush hour!).
E	110	125	140	I have done this trip many times and it usually takes just under 2 hours, but at this time of day it may take a little longer. I think it is equally likely the average is above or below 125 mins. and 90% sure it will be 15 mins. either way.
Mean	**87**	**118**	**150**	

Some of the experts e-mailed with queries about the Round 2 questionnaire:

"Do I have to explain why I changed my median judgement? I feel it was too low the first time but I do not know why (there are other estimates with convincing rationales both above and below mine)."
"Could you remind me how to make the quartile judgements?"
"I have been discussing this with a colleague and realize I was thinking of the wrong airport. What should do?"

Note that these are fictitious data and questions for purposes of illustration and discussion only!

Lecture 10: Cooke Method

John Quigley &
Simon French

Outline

- Key Features of the method
  - Performance Based Aggregation
  - Seed Questions
  - Results & Excalibur
  - Documentation
- Pre-elicitation
  - Expert and elicitor selection
  - Preparation & planning

Note: we will skim over the technical details in many places. The elicitor/analyst will deal with these.

Performance Based Aggregation

- Experts provide their judgements in isolation from other experts
  - maintains a degree of independence
- Individual judgements elicited on:
  - seed questions
  - target questions or variables of interest
- Answers for seed questions are known by elicitor but not by experts; these are used to assess performance
- Aggregate judgements – weighted average where weights are based on performance on seed questions

Note: Cooke's method appears very quantitative, but the elicitor should gather the qualitative reasoning and thinking behind each expert's quantitative assessment.

GD 6.2

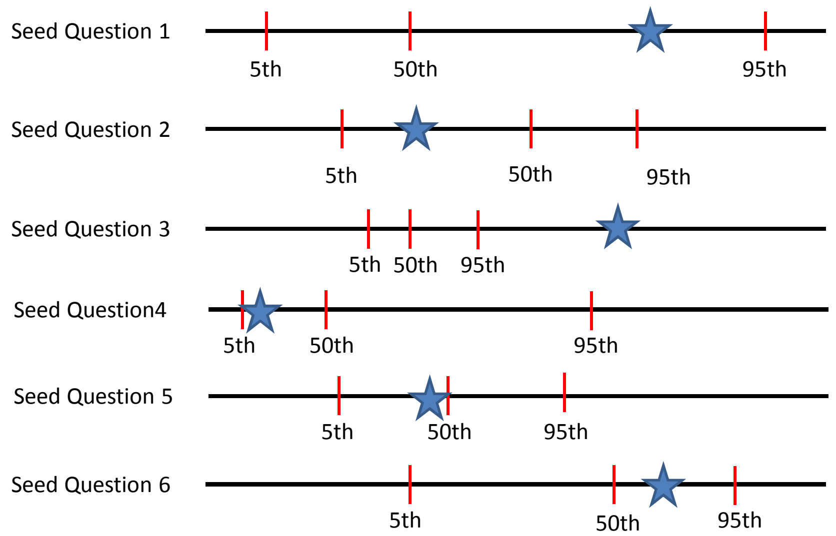
Practical Issues with Seed Questions

- Typically use 10 seed questions
- Need to involve domain experts in framing seed questions
- The seed variable should sufficiently cover the case structures for elicitation
  - *Domain variables*: same dimensions as target question, from previous studies or similar conditions
  - *Adjacent variables*: different dimensions, but about which experts should be able to give an educated guess
- Seed variables must be provided for all sub-fields
- Independence amongst seed variables
- Seed variables may be identified as such during elicitation, but not necessarily

Example Seed Questions

- *Domain variable*: Subjects held for 60 seconds a 8×12 cm portion of thermal paper containing 27.2 mg BPA/g paper. Deposited BPA was swiped and measured by HPLC.
  - What was the average amount of BPA swiped from clean dry hands?
- *Adjacent variable*: Subjects held BPA coated receipts for 4 minutes then held a French fry for 4 minutes and finally consumed the French fry.
  - What was the average BPA concentration in subjects' urine 90 minutes later?

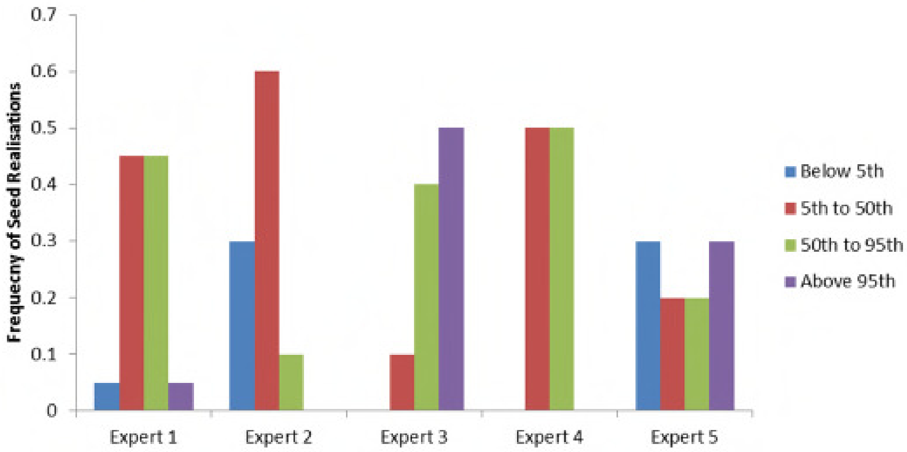
Observing Degree of Calibration



Calibration

- Calibration measures the degree to which the frequency of true values appearing in intervals corresponds to the probabilities assigned by the expert
  - i.e. in a statistical sense, the calibration score measures how close the actual seed values are to an expert's predictions.
  - Calibration scores actually correspond to *p*-values in a particular hypothesis test
- Higher scores indicate greater agreement

Observing Degree of Calibration



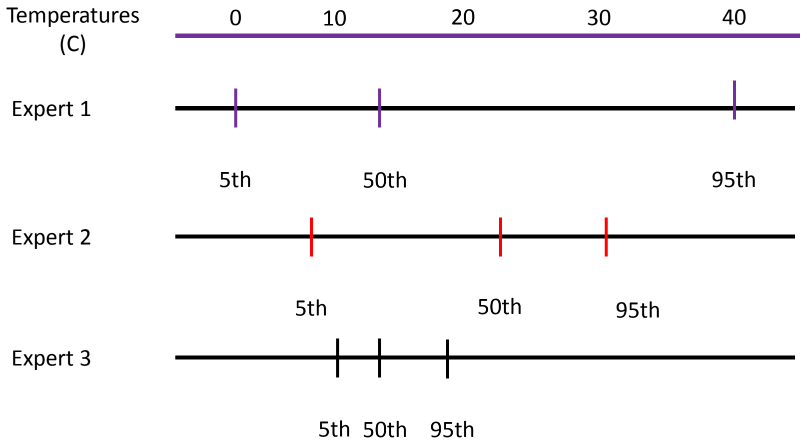
Perfectly calibrated | Over estimates | Under estimates | Too wide | Too narrow

Observing informativeness

Consider three experts, each asked for the recorded temperature (C) in Toronto on September 30 2015.

- Expert 1:
  0 for 5^{th} percentile, **13** for 50^{th}, and **40** for 95^{th}
- Expert 2:
  9 for 5^{th} percentile, **22** for 50^{th}, and **30** for 95^{th}
- Expert 3:
  10 for 5^{th} percentile, **13** for 50^{th}, and **20** for 95^{th}

Informativeness



Expert 3 is more informative because his 90% bounds are tighter about the median:

- More precise (informative) but could, of course, be less accurate (poorly calibrated)

Informativeness

- Measures the informativeness of an expert relative to the other experts
- Rewarding those with probabilities assigned to shorter intervals
- Does not use true values so can evaluate information on both target and seed questions
- Does not vary to same extent as calibration score
  - Poorly calibrated experts not compensated with good information
  - Differentiates between equally calibrated experts

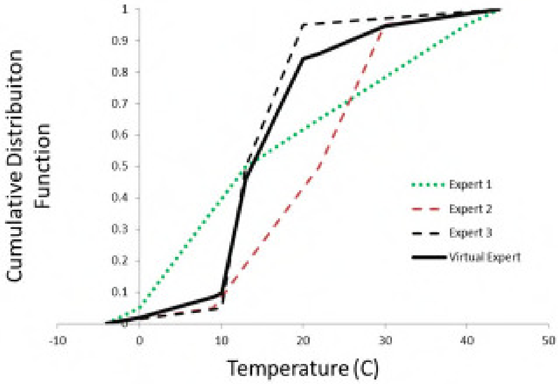
Aggregation – Example

	L	5%	50%	95%	U
Expert 1	-4	0	13	40	44
Expert 2	-4	9	22	30	44
Expert 3	-4	10	13	20	44
Virtual Expert	-4	4	14	31	44

Cooke's method forms a *virtual expert*, called *decision maker* (DM) in the Excalibur software, by forming a weighted average of the experts judgements.

Note:

- Higher calibration gives higher weight
- To a lesser extent, higher information gives higher weight
- Some weights may be zero



Weights

- Raw weights are obtained by multiplying calibration score with information score
- Experts who do not achieve a minimum calibration score are excluded
  - i.e. given zero weight
- Analyst chooses minimum score
- Raw weights are then adjusted to ensure they sum to 1

Illustrative Example

- Four experts each assessed 11 quantities
- Each predicted the temperature recorded on BBC website for 11 cities
- Each provided a 5th, 50th and 95th percentile for each city
- We will look at output from Cooke's Excalibur software
  - Somewhat dated
  - But well tested

Excalibur Table of Results

Best performing expert on information

Worst performing expert on calibration

Best performing expert on calibration

Worst performing expert on information

Expert scores: Example with temperatures

Results of scoring experts
Bayesian Updates: no Weights: global DM Optimisation: no
Significance Level: 0.01 Calibration Power: 1

Nr.	Id	Calibr.	Mean relative total	Mean relative realization	Numb real	UnNormalized weight	Normaliz.weigt without DM	Normaliz.weig with DM
1	e1	0.04576	0.6405	0.6405	11	0.02931	0.1154	0.1002
2	e2	0.001832	0.5291	0.5291	11	0	0	0
3	e3	0.09406	0.345	0.345	11	0.03245	0.1278	0.111
4	e4	0.6378	0.3014	0.3014	11	0.1922	0.7568	0.6572
5	GW	0.3697	0.1041	0.1041	11	0.03849		0.1316
6	EW	0.01247	0.09314	0.09314	11	0.001162		0.0045[illegible]7

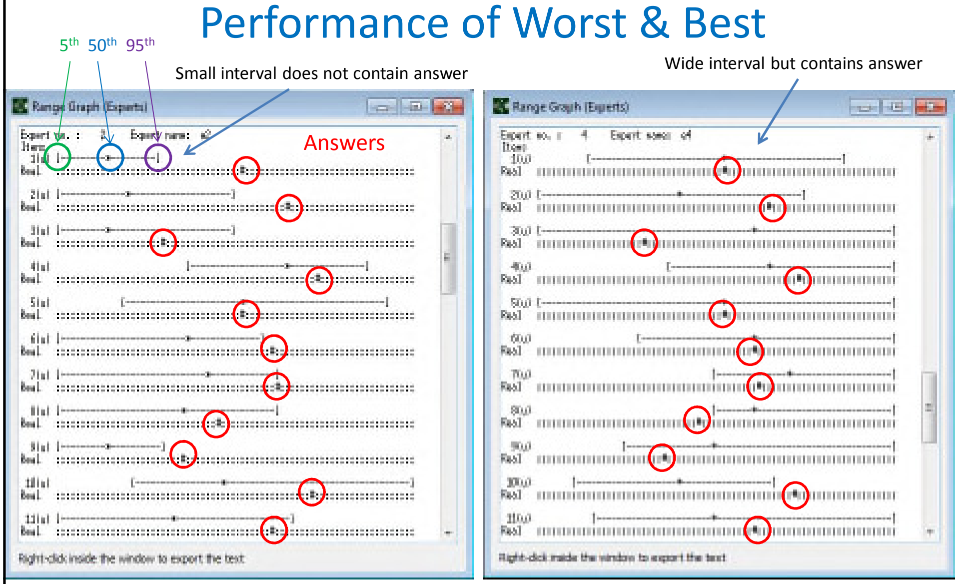
EW is equal weights performs poorest for calibration & information

GW is Global weights with significance level cutoff at 1%

Weights assigned to experts using Global Weights

GD 6.2.7

Performance of Worst & Best



Feedback and Documentation

- Experts must have access to:
  - their assessments
  - their calibration and information scores
  - their weighting factors
  - any conclusions about over- or underconfidence
  - conclusions about their tendency to over- or underestimate.
- All relevant information such as seed questions, performance analysis and experts rationale are recorded

GD 6.2.9

Expert Selection

- Aim for 6 to 10 experts
- Number of questions answered by expert will vary depending on:
  - Subject
  - cognitive ability of experts
  - number of mental models required
- If experts can answer many questions with few mental models then as many as 100 questions can be answered

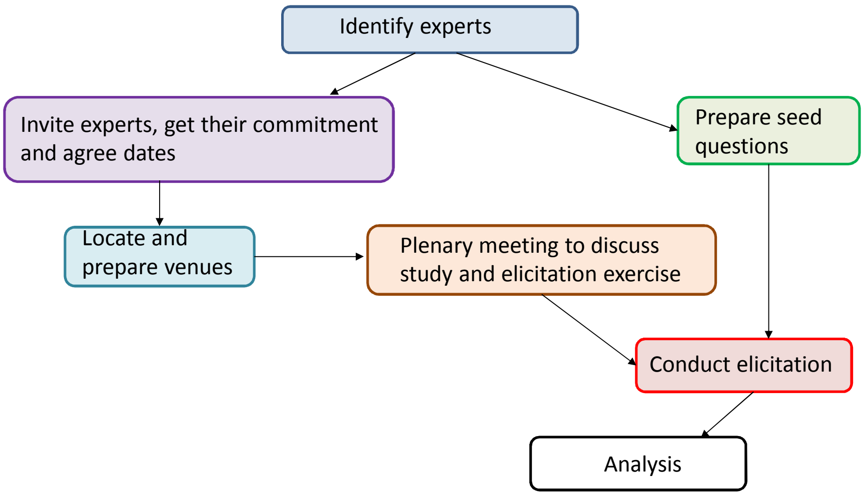
GD 6.2.2

Elicitor

- Normative Elicitor is knowledgeable about elicitation method
- Substantive Elicitor is knowledgeable about the topic being investigated
  - May not need a substantive elicitor
  - But certainly need a normative one!
- Elicitations are one on one
- Dry run
  - Check protocol is clear
  - Anticipate questions from experts
  - Elicitor remains neutral never coaches

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation GD 6.2.8 L10/19

Timescale flowchart



FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L10/20

Workshop

- Half a day to three days
- Start with plenary meeting with all experts (day 1)
  - Study design explained
  - Short elicitation conducted
  - Experts shown how process works
  - Experts can share understanding prior to individual elicitation
- Conduct individual elicitation (day 2)
- Feedback results and discuss (day 3)
- Recommend eliciting both seed and variables of interest at the same time following plenary
- Not always possible to bring all experts together in plenary and is not necessary

Summary

- Weights based on performance
  - Expertise is assessed
- Measure combines calibration and information
  - More emphasis on calibration
- Need for suitable seed questions
  - Need to ensure all experts share same understanding of seed (and target) questions
- Qualitative and quantitative analysis of expert performance on seed questions

References

- EFSA, 2014. Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment. EFSA Journal 2014;12(6):3734.

Practical 9: Developing Seed Questions

John Quigley & Simon French

Objective

- Explore the development of seed questions for the Cooke method

Practical 9: Developing Seed Questions

- Work in the same group as for Practicals 7 and 8
- Start with the parameter listed first for your example
- Try to identify relevant seed variables: i.e.
  - Variables 'cognitively or scientifically' close to the target parameter
  - Variables for which you can obtain the actual values – and can explain how you would obtain those values.
- If time permits, move on to other parameters in your example
- *Write on a flip chart:*
  - *Examples of your seed questions*
  - *Any lessons learned about the process of developing seed questions*

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation P9/3

Plenary: Feedback from Practicals 7-9 Sheffield, Delphi and Cooke methods

Andy Hart

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation P7-9/1

Objective

- Review lessons learned regarding the three basic methods for EKE: Sheffield, Delphi and Cooke

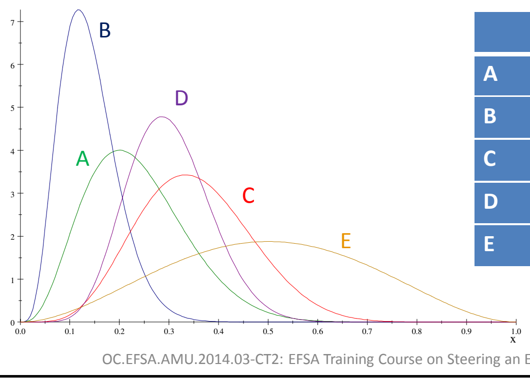
FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation P7-9/2

Topics for discussion

- Sheffield method:
  - Timetable for a Sheffield EKE
  - How to lead experts towards a consensus judgement
- Delphi method:
  - Guidance to experts on writing their rationales
  - Lessons learned on running a Delphi EKE
- Cooke method:
  - Example seed questions
  - Lessons learned on developing seed questions

Practical task 7.2

- Five experts (A, B, C, D and E) have given the median and quartile judgements shown in the table
- The graph shows the five fitted distributions
- What are the principal differences between the experts' judgements?
- *As the elicitor, how would you lead a discussion towards reaching consensus judgements?*



	Q1	M	Q3
A	0.15	0.22	0.3
B	0.1	0.14	0.18
C	0.27	0.35	0.45
D	0.25	0.3	0.35
E	0.35	0.5	0.65

Lecture 11: Selecting the appropriate elicitation method

Tony O'Hagan,
Fergus Bolger,
John Quigley

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L11/1

Outline

- Three methods – review of principal differences
- Strengths and weaknesses
  - Generic considerations
  - Context-specific considerations
- Making the choice

GD 4.4 with some extensions

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L11/2

Three methods

- The Sheffield, EFSA Delphi and Cooke methods are all presented in the Guidance as suitable for use in EFSA risk assessments
- It is possible to vary these standard methods in many ways
  - Using information in Appendix A of the Guidance
  - But this is advised only after gaining experience with the standard versions
- Steering Group must decide which method to use
  - For each parameter that will be elicited

Principal differences

Method	Sheffield	EFSA Delphi	Cooke
Aggregation	Behavioural Individual judgements followed by 'consensus' judgements	Mixed Limited behavioural followed by unweighted pool	Weighted pool Weights derived from performance in judging seed variables
Managing experts	Workshop Experts meet together and interact fully	Remote Conducted by email with limited interaction	Mixed Maybe a single location but usually no interaction
Quantiles elicited	5 Credible bounds, median and quartiles	5 Credible bounds, median and quartiles	3 5^{th} percentile, median and 95^{th} percentile
Distribution fitted	Smooth With feedback	Histogram	Histogram

GD 4.4

STRENGTHS & WEAKNESSES – GENERIC

Aggregation

Methodl	Pros and cons
Sheffield Behavioural	Advantages: Experts share and discuss opinions Aggregate distribution with explicit interpretation Disadvantages: Difficulty of managing experts Possible additional biases from group interaction
Cooke Weighted linear pool	Advantages: Objective weighting through seed variables Avoids problems of group interaction Disadvantages: Difficulty of constructing seed variables No discussion between experts
EFSA Delphi Mixed	Advantages: Controlled sharing of reasons for judgements Easy to use. Avoids problems of group interaction Disadvantages: Dropout. Arbitrary aggregation rule Communication difficulties due to remote working

GD 4.4.1-

Accumulated experience

Method	Pros and cons
Cooke	Method has been used unchanged over many years Substantial accumulated experience and database Some accumulated evidence of good performance
Sheffield	Builds on long established use of behavioural aggregation Sheffield method itself used in same basic form for 7 years Now used widely
EFSA Delphi	Simple Delphi has a very long history, very widespread use EFSA Delphi is a substantial modification Only a few applications

GD 4.4.1

Informed by psychology

Method	Pros and cons
Sheffield	Explicitly based on psychological research Elicitor uses templates to enforce good framing
EFSA Delphi	Traditional Delphi informed by psychology, EFSA Delphi has same framing as Sheffield But experts may not comply
Cooke	Not explicitly informed by psychology But has features (e.g. weighting) to control biases

GD 4.4.1

Calibration

Method	Pros and cons
Cooke	Badly calibrated experts will be removed through seed variables
Sheffield	Experts who make unrealistic judgements should be recalibrated through group discussion
EFSA Delphi	Experts who make bad judgements will sometimes be persuaded to change by seeing other experts' judgements and- rationales

GD 4.4.1

STRENGTHS & WEAKNESSES – CONTEXT-SPECIFIC

Geographical

Method	Pros and cons
Sheffield	Requires all experts to come together in elicitation workshop
Cooke	Does not require all experts to attend together, although this is preferable
EFSA Delphi	Managed remotely, so experts can be widely spread

GD 4.4.2

Language

Method	Pros and cons
Sheffield	Depends on interaction Experts and elicitor should be reasonably fluent in a common language, speaking and listening
Cooke	Also depends on interaction but to a lesser extent Expert and elicitor should nevertheless be reasonably fluent in a common language, speaking and listening
EFSA Delphi	Also depends on interaction but only in written form Experts and elicitor should be reasonably fluent in a common language, reading and writing

- Instructions, questions and evidence can be translated into expert's own language
- Real-time interpretation can moderate language problems, but requires exceptional interpreters

GD 4.4.2

Diversity of background

Method	Pros and cons
Cooke	Requires all experts to be able to make judgements without discussion, about both the seed variables and the parameters, so all must have enough of a common background
EFSA Delphi	Requires all experts to be able to make judgements without discussion, so all must have enough of a common background
Sheffield	Through discussion each expert can benefit from the expertise of others having different backgrounds/disciplines

GD 4.4.2

Skill requirement

Method	Pros and cons
EFSA Delphi	Skill needed to write questionnaires and to summarise responses – relatively low requirement
Cooke	Skill needed to develop seed variables and to work with experts individually – relatively high requirement
Sheffield	Skill needed to work with and manage a group of experts – relatively high requirement

Extends GD 4.4.2

Time requirement

Method	Pros and cons
Sheffield	Lead time for recruiting several experts to attend a workshop can be substantial Weeks
Cooke	Developing good seed questions is a substantial commitment Also generally aims to have experts meeting together Weeks to months
EFSA Delphi	Time must be allowed for experts to respond to each questionnaire round, and for responses to be summarised between rounds Months

GD 4.4.2

Sensitive parameters & conflicts of interest

Method	Pros and cons
Cooke	Biased experts should be down-weighted But this requires that seed variables not be identifiable
Sheffield	Expert bias hopefully moderated by the group But this requires that conflicts of interest are declared
EFSA Delphi	Delphi has no real mechanism to control potential bias from, e.g., industry experts

Extends GD 4.4.2

Other factors

- Methods may differ on cost
  - But no experience yet
  - E.g. Delphi probably cheaper in expert time but more expensive in staff time
- Although experts will take the e-learning course reminders/revision will still be advisable
  - Easier in Sheffield and Cooke

MAKING THE CHOICE

Choosing for generic reasons

- Some people will have a preference for one method
  - Based on generic balance of strengths and weaknesses on aggregation, accumulated experience, etc.
    - E.g. a preference for or against behavioural aggregation
  - This was the case for some members of the Working Group that prepared the Guidance
  - Or a generic preference may be acquired through experience
- But context-specific reasons might over-rule a generic preference
  - E.g. Sheffield might be preferred but experts cannot be brought together for a face-to-face workshop

Choosing for context-specific reasons

- Other people will be happy to place context-specific considerations at the forefront
  - Choice based on balance of strengths and weaknesses for geography, language, diversity, resources, sensitivity
  - No generic preference for one of the three methods
- What do you think?
  - Next practical …

Summary

- The three methods differ in many ways
  - So they have different strengths and weaknesses
- Generic considerations
  - Aggregation method, accumulated experience, informed by psychology, calibration
- Context-specific considerations
  - Geography, language, diversity of background, time requirement, skill requirement, sensitivity/conflicts
- Choice can be based on any of these
  - Generic preference may be over-ruled by context

References

- EFSA, 2014. Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment. EFSA Journal 2014;12(6):3734.

Practical 10: Selecting the appropriate elicitation method

Tony O'Hagan, Fergus Bolger,
John Quigley, Andy Hart

Objective

- Assess the relative applicability of the Sheffield, Delphi and Cooke methods to the parameter considered in earlier Practicals
- Decide and justify which method to recommend for this parameter

Practical task

- Discuss the generic considerations for choosing between methods
  - Aggregation; accumulated experience; informed by psychology
- Discuss the context-specific considerations in applying each method to elicit expert knowledge about the parameter defined in Practical 5, from experts such as identified in Practical 6
  - Geographical; language; diversity; resources, etc.
- Consider which of the three methods you would recommend for this case, with reasons
  - First, do this individually
  - Then discuss any differences in preferences and reasons between group members
- *Summarise the group's preferences & reasons for the plenary*

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation P10/3

Lecture 12: Documentation: Repeatability, Transparency and Confidentiality

Fergus Bolger, Simon French,
Tony O'Hagan & John Quigley

Outline

- Transparency and repeatability of EKE
- Confidentiality of experts' judgments
- 3 milestones of technical documentation:
    1. Decision to carry out EKE
    2. Choice of elicitation protocol and selection of experts
    3. Execution of the elicitation process
- Builds up a Knowledge base of EKE practice
    - Evidence to develop *good* practice
- Also useful for general feedback to the experts
    - For detailed feedback, see guidance

Transparency and repeatability

- Full public documentation is a fundamental characteristic of EFSA's work
  - ensures that risk assessment procedures are done in a transparent manner
  - the topics and stages in risk assessment procedures must be predefined and clearly stated
- As already noted EKE itself is a full process rather than a single method
  - The documentation must therefore summarise all steps and decisions taken from the initiation until the final result
  - Each specific EKE process must be fully repeatable from the documentation

GD A.5.

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L12/3

Confidentiality

- EKE has its particular confidentiality requirements due to the involvement of external experts
- Disclosing personal data that might identify individual experts with their judgments
  - is neither an objective of the EKE process nor necessary to fulfil transparency requirements
  - may discourage experts from taking part in the process or influence their responses
- Participating experts should therefore be assured on the confidential treatment of their individual answers
  - 'Chatham House rules' reports
    - Who took part
    - What was said but not who said it.
  - EFSA will outsource EKE if needed for complete confidentiality

Acera, 2010; Chatham House Rules; GD A.5.

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L12/4

Problem definition of the initiation phase

- Documentation of Milestone 1
  - decision to carry out EKE
- *Authored by Working Group*
  - Evidence dossier (including RA model)
  - Existing information on parameter of interest
  - Justification and necessary conditions for EKE evaluation by the corresponding panel and EFSA administration

GD 7.1. & A.5.2.

Elicitation protocol of the pre-elicitation phase

- Documentation of Milestone 2
  - definition of the elicitation protocol and selection of experts
- *Authored by Steering Group*
  - final elicitation questions
  - description of expert selection procedure
  - decision on the elicitation method
  - final project plan for elicitation
  - external review, if applicable

GD 7.1. & A.5.2.

Result report of the elicitation phase

- Documentation of Milestone 3
  - execution and documentation of the elicitation process
- *Authored by Elicitation Group*
  - expert panel constitution
  - evidence dossier and training sessions provided
  - elicitation methods, process, time line and questions
  - data analysis methods
  - anonymised expert rationales for judgements
  - results for use in risk assessment
  - discussion of assumptions, qualitative uncertainties and constraints of result
  - complaints regarding result (if any)
  - evaluation of the process and result by SG and WG

GD 7.2., A.5.1. & A.5.2.

Summary

Type of report	**Content/audience**	**Author**
Result report	Summarises the results and will be used and published in the risk assessment procedure	Elicitation Group
Technical support document	Includes a full description of the process and enables the public to review the study	Working Group
	Decision for expert knowledge elicitation	Working Group
	Definition of the elicitation protocol and selection of experts	Steering Group
	Execution and documentation of the elicitation process	Elicitation Group
Expert feedback	Confidential documentation for the individual expert summarising the input from each expert	Elicitation Group

Reference

- EFSA, 2014. Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment. EFSA Journal 2014;12(6):3734.
- ACERA (Australian Centre of Excellence for Risk Analysis), 2010. Process manual – Elicitation Tool. ACERA, Melbourne, Australia, 40 pp. Available online:
- http://www.acera.unimelb.edu.au/materials/endorsed/0611-process-manual.pdf
- Chatham House Rule: http://www.chathamhouse.org/about-us/chathamhouserule

Lecture 13: Advanced topics

Tony O'Hagan,

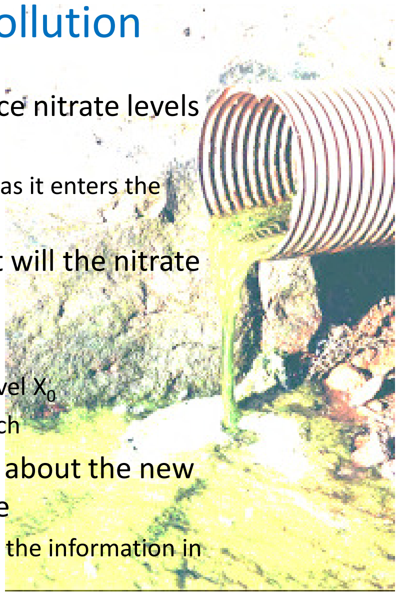
Outline

- Elaboration
  - When thinking is hard
  - The general technique
  - Heterogeneity and modelling
- More applications
  - You get to choose
- Imprecision

This lecture goes beyond the material in the GD

ELABORATION

Nitrate pollution



- New technique proposed to reduce nitrate levels in river water
  - By treating groundwater in a trench as it enters the river
- If this scheme is carried out, what will the nitrate level X be afterwards?
- Relevant evidence relates to
  - Sample measurements of current level X_0
  - Efficacy of the method in a test trench
- Direct elicitation of expert beliefs about the new concentration level X is not simple
  - Requires the expert to synthesise all the information in her head

Nitrate pollution elaboration

- Instead of eliciting judgements directly about X, express it in terms of
  - Current level X_0
  - Proportion P due to groundwater sources
  - Reduction factor R under test conditions
  - Fraction F of that reduction achievable in the field

$$X = X_0 (1 - PRF)$$

- Elicit judgements about X_0, P, R and F
  - Distribution of X derived via the equation
  - Evidence relating to the different components is now separated
  - Expert might judge them independent (see later)

When thinking is hard

- We often need to elicit judgements in this kind of situation
  - The parameter is complex
  - Data are complex
  - It’s hard to think about all the factors
- Then it makes sense to break the parameter down

Elaboration

- Elaboration is the process of expressing things that are difficult to quantify in terms of other, simpler things
- Express parameter X as a function of quantities
  $Y = \{Y_1, Y_2, Y_3, \ldots\}$
  - Where the distribution of Y is easier to elicit
- How to achieve this?
  - Find out how the experts think about X
    - Expertise often involves developing heuristics to break problems down into components that can be thought about individually
  - Or break down according to data sources
    - As in the nitrate pollution example

The downside

- More work!
  - We now have several parameters to elicit distributions for, instead of one
- Is it worth it?
  - In principle, we will get a more accurate assessment about X using elaboration
    - But it may make negligible difference in the risk assessment
  - Informally recognising the elaboration can still be useful when thinking is hard
    - Less effective but less work

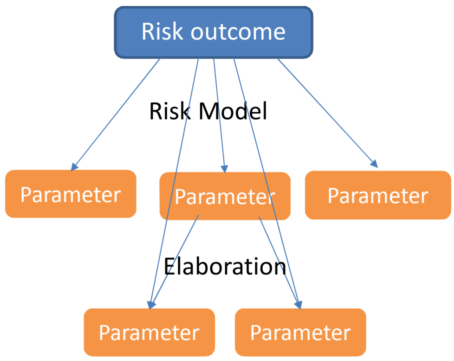
Heterogeneity

- We've talked about the problem of heterogeneity in the expert group
  - Difficult for experts to make judgements when some aspects are outside their expertise
  - And when the EKE method doesn't let them share knowledge
- Elaboration can often help
  - Break down the parameter into parts that lie in the different areas of expertise
- But it can also greatly increase the workload
  - Not just eliciting distributions for several parameters instead of one
  - We also need multiple expert groups

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L13/9

Elaboration is modelling

- The Working Group is responsible for the risk model
  - Elaborates the risk outcomes in terms of the model parameters
- Elaborating parameters is a further refinement of the model
  - Can be done by experts and elicitor during elicitation
    - But that's often impractical
  - Better if done by Steering Group or Elicitation Group in preparing for elicitation
- Could combine the two layers of modelling
  - Replacing X in the risk model with its elaboration



FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L13/10

MORE APPLICATIONS

Il menù del giorno

- Bridging data gaps
  - We have data but it refers to a slightly different parameter
- Correlated parameters
  - Judgements about parameters are not independent
- Extremes
  - How to elicit knowledge about rare events
- Parameters that are functions
  - Dose-response
  - Variability

You get to choose!

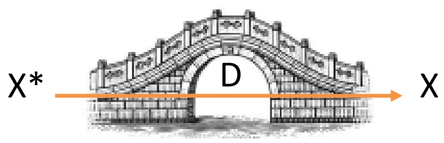
Bridging data gaps (1)

- Do we need expert elicitation when we have data?
  - We can just do statistical analysis of the data
  - Uncertainty about the parameter calculated statistically
    - E.g. estimate and confidence limits
- That's fine, IF …
  - there is no additional expert knowledge, and
  - the data concern precisely our parameter
- If there is additional expert knowledge, we can use Bayesian statistics
  - Elicit prior distributions and combine with data
- But what if the data refer to a slightly different parameter?

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L13/13

Bridging data gaps (2)

- Very often the data relate to a similar but different parameter, say X*
  - E.g. a different chemical, a different age group, a different environment
  - "Read across" is common, using data on X* as if it were on X, but we can do better
- Elaboration
  - X = X* + D or
  - X = X* x F
    - Elicit expert judgements about difference D or factor F
  - Combine with statistically-derived distribution for X*



FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L13/14

Correlated parameters (1)

- We often want experts to provide judgements about more than one uncertain quantity
  - Toxicities of two chemical agents
  - Hours of sunshine and growth rate
- With multiple quantities, need to think about dependence
- Two or more uncertain quantities are independent if:
  - When you learn something about one of them it doesn't change your beliefs about the others
  - It's a personal judgement, like everything else in elicitation!
    - They may be independent for one expert but not for another
- Independence is nice
  - Independent quantities can just be elicited separately

Correlated parameters (2)

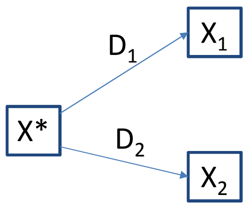
- When parameters are not independent, elicitation is much more complex
- Elaboration may allow you to transform a set of correlated parameters into an independent set
  - Express quantities of interest $X = \{X_1, X_2, X_3, \ldots\}$ as a function of quantities $Y = \{Y_1, Y_2, Y_3, \ldots\}$
    - Where the Y_is are independent
    - As judged by the experts
- Even when elaborating a single parameter X it is important to find an elaboration where the Y_is are independent!
  - If possible

Correlated parameters (3)

- The elaboration in bridging data gaps is a simple example
  - To choose between X = X* + D and X = X* x F, ask whether D or F would be independent of X*
- The bridge is also one way of eliciting knowledge about toxicities of two chemicals
  - If we have data for one toxicity, X*, but not for the other, we can construct a distribution for X by a suitable bridge
  - X and X* will be correlated

Correlated parameters (4)

- In the last example, there was asymmetry between the two toxicities – data on one but not the other
- More often we will have similar quality of knowledge about both
  - To elaborate two (or more) toxicities X_1 and X_2, introduce a reference chemical with toxicity X*
    - A well studied chemical in the same class
    - Or just the average of the class
  - Build bridges from X* to X_1 and X_2
  - With independence between X*, D_1 and D_2
    - Or X*, F_1 and F_2

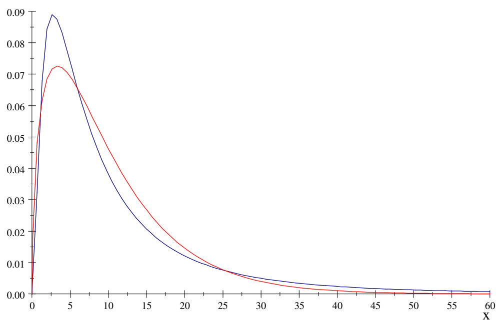


Extremes (1)

- In risk assessment, the risk is very often associated with extreme values of one or more parameters
  - Either very high or very low values
  - So we are interested in the tails of their probability distributions
- Eliciting tail probabilities is hard
  - Experts cannot reliably assess very small probabilities
- We fit a distribution to an expert’s assessment of quantiles like median and quartiles
  - Distributions with widely differing tail thicknesses can all fit the elicited quantiles equally well
  - But give very different probabilities for extremes

Extremes (2)

- The two distributions are both fitted to the same elicited quartiles
  - Q1 = 4
  - M = 7.5
  - Q3 = 15



- They give very different probabilities in the tails
  - Red curve (gamma): P(X > 40) = 0.007
  - Blue curve (lognormal): P(X > 40) = 0.046

Extremes (3)

- In my opinion the best way to tackle tail probabilities is by elaboration
  - The parameter of interest is elaborated so that extreme values arise from less extreme values of the components
- The RVF risk model is an example
  - The number of infected animals entering the RoC can be very large if all the components v, p, d, t, e are moderately large
  - Its tail thickness is not determined by the tails of the components

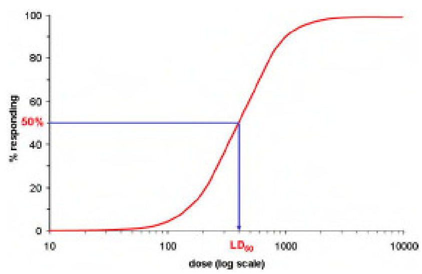
FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L13/21

Functions (1)

- Risk models often involve uncertain functions
- For instance a dose response function

  R = f(D)

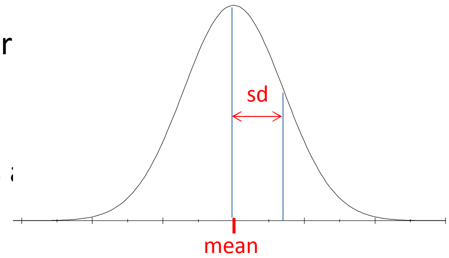
  - where D is (log) dose, R is response
  - The whole function f is uncertain
- Elaboration
  - Assume f has a standard form
    - E.g. logistic
  - Characterised by a small number of parameters
    - E.g. LD_{50} and gradient at the LD_{50}
    - Gradient could be expressed as, e.g., difference between LD_{90} and LD_{50}
  - So elicit expert knowledge about these
    - Reasonable to assume independent



FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L13/22

Functions (2)

- Another example is when we have random variability
  - The random variation is described by a probability distribution
  - But that distribution is uncer
  - As discussed in Lecture 2
- Elaboration
  - Assume the distribution has a standard form
    - E.g. Gaussian
  - Characterised by a small number of parameters
    - E.g. mean and standard deviation
  - Elicit expert knowledge about these
    - Often reasonable to assume independence



IMPRECISION

It's not an exact science!

- Elicitation can never be exact
  - The elicited judgements are only approximate
  - And they only partially specify the distribution
- RA must take account of imprecision as well as uncertainty
  - Sensitivity analysis can check whether the RA conclusions are robust
    - Varying the elicited quantiles and/or the fitted distribution
    - Within reasonable bounds of imprecision
  - Alternative theory of imprecise probabilities
- If conclusions change materially we can try to remove the sensitivity
  - Elicit more judgements from experts
  - Or involve more experts
- Although this may not be feasible!

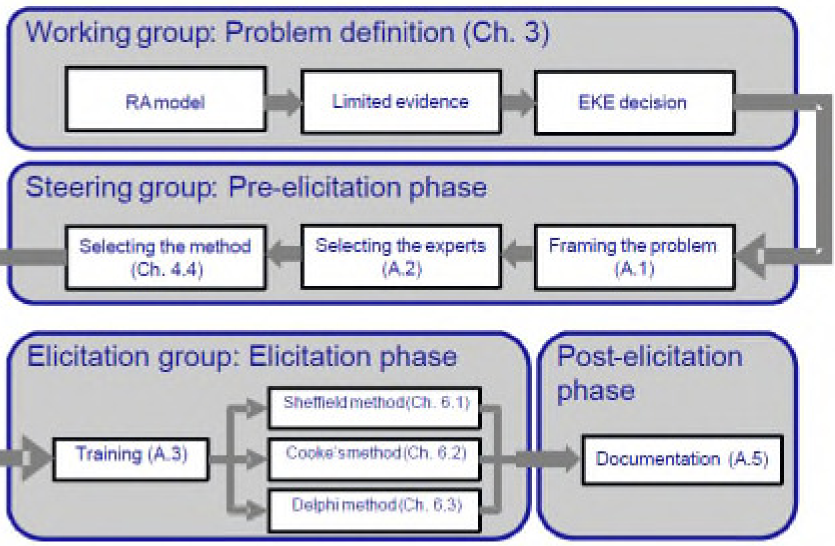
References

- EFSA, 2014. Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment. EFSA Journal 2014;12(6):3734.
- O'Hagan, A. (2012). Probabilistic uncertainty specification: Overview, elaboration techniques and their application to a mechanistic model of carbon flux. Environmental Modelling and Software 36, 35-48.

Lecture 14
Steering the elicitation process: review of key points

Andy Hart

Outline



GD Figure 1, p. 8

Lecture 1: Reasons and roles for EKE

Expert judgement is used in all areas of EFSA's work

Including:

- Chemical risk
- Nutrition
- Environmental risk
- GMOs
- Plant health
- Animal Health
- Biohazards

...and is a key part of EFSA's Guidance on uncertainty

EKE Guidance focusses on quantitative judgements

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L14/3

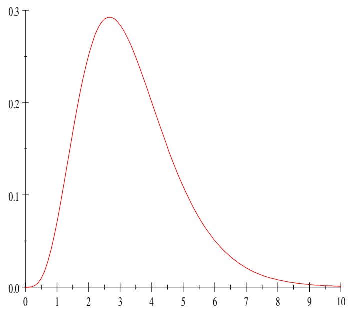
Expert Knowledge Elicitation

Defined by EFSA as:

'A systematic, documented and reviewable process

...to retrieve expert judgements from groups of experts

...in the form of a probability distribution'



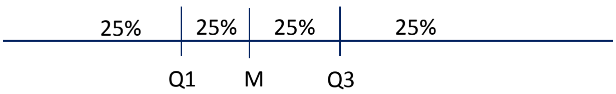
EFSA GD Glossary

GD Glossary

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L14/4

Lectures 2 & 3: Principles of EKE

- Subjective probability
- Careful, reasoned judgements
- EKE procedures are designed to counteract psychological biases



FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L14/5

Lecture 4: Prioritisation

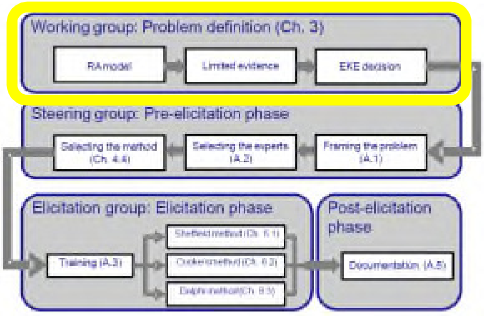
- Working Group prioritises parameters for EKE:
  - Minimal Assessment
  - Sensitivity Analysis

Often, a small number of parameters are responsible for most of the uncertainty

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L14/6

Deliverables of the Working Group:

- *Background report* including:
  - risk assessment model
  - available evidence
  - justification for EKE
  - preliminary timeline
  - SG membership
  - resource estimates
  - (contract specification)



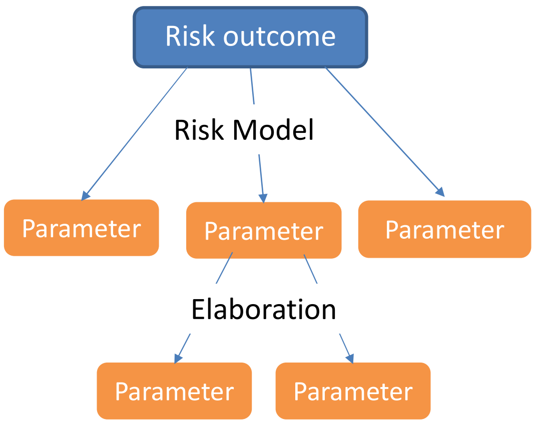
- *Reviewed by Panel Chair & Unit Head before proceeding*

GD 3.6, 3.7

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L14/7

Lecture 5: Question definition

- Precise specification
- Potentially observable
- Elaborate model if necessary (*Lecture 13*)

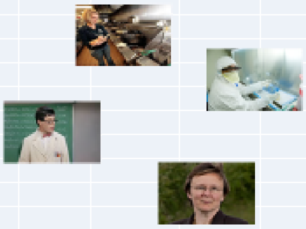


FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L14/8

Lecture 6: Experts

- Identification, Recruitment & Retention

Knowledge requirements			Country	Expert Roles			
				Industry		Govt	Academia
Substantive Expertise	Importance	Specificity		Supply chain	Production	(Inspector)	(Scientist)
Immunity to levels of salmonella	Essential	Specific	AA				
			BB				
		Comparable	AA				
			BB				
Quality of Food in supply chain	Essential	Specific	AA				
			BB				
		Comparable	AA				
			BB				
Conditions of fast food kitchen	Essential	Specific	AA				
			BB				
		Comparable	AA				



Lecture 7: Evidence dossier

- WG and SG summarise available evidence

Lecture 11: Choice of EKE method

- SG decision
- General and Specific considerations

Protocol	Sheffield	EFSA Delphi	Cooke
Aggregation	Behavioural Individual judgements followed … judgem…	Mixed Limited behavioural …	Weighted pool Weights derived from …
Managing experts	Works… Experts interact…		
Quantiles elicited	5 Credible … and qua…		
Distribution fitted	Smoot… With fe…		

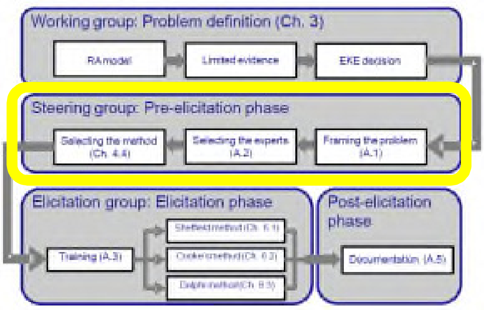
Method	Pros and cons
Cooke	Method has been used unchanged over many years
Sheffield	
EFSA Delphi	

Method	Pros and cons
Sheffield	Lead time for recruiting several experts to attend a workshop can be substantial Weeks
Cooke	Developing good seed questions is a substantial commitment Also generally aims to have experts meeting together Weeks to months
EFSA Delphi	Time must be allowed for experts to respond to each questionnaire round, and for responses to be summarised between rounds Months

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L14/11

Deliverables of the Steering Group

- *Elicitation protocol*
  - the elicitation question
  - long list of experts
  - proposed EKE method
  - revised timeline
  - project plan



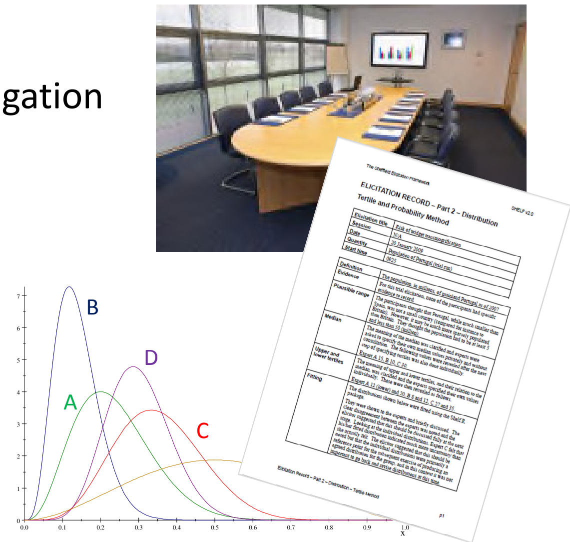
- *Reviewed by EFSA and WG*

GD 4.6, 4.7

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L14/12

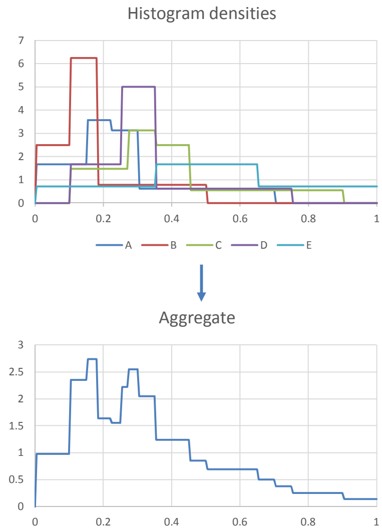
Lecture 8: Sheffield method

- Behavioural aggregation
- Skilful facilitator
- SHELF software



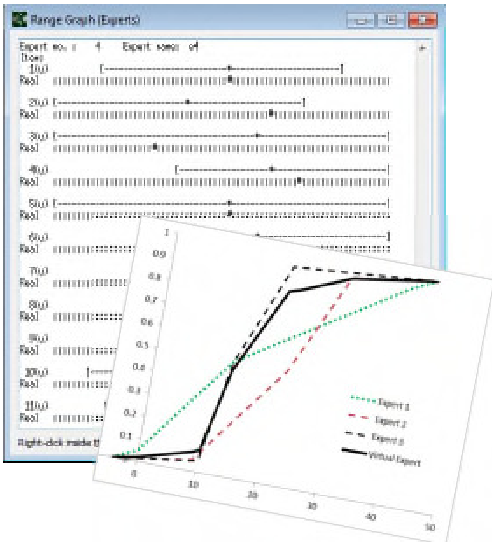
Lecture 9: Delphi method

- Remote interaction
- Equal weighting



Lecture 10: Cooke method

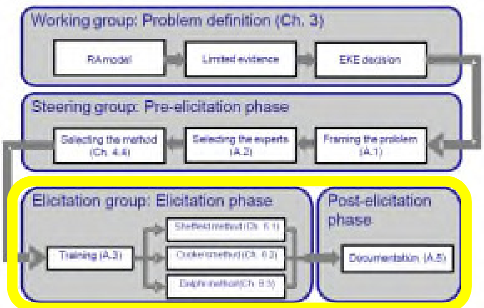
- Independent judgements
- Performance-based aggregation:
  - Calibration
  - Informativeness
- Excalibur software



FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L14/15

Deliverables of the Elicitation Group

- *Technical Documentation*
- *Result Report*
  - suitable for publication
- Feedback to experts
- *Reviewed by the Steering Group*

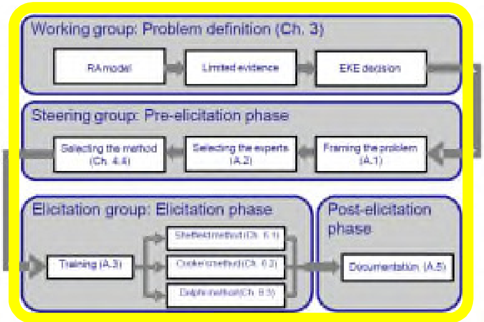


GD 6.4

FINAL OC.EFSA.AMU.2014.03-CT2: EFSA Training Course on Steering an Expert Knowledge Elicitation L14/16

Lecture 12: Documentation

- Repeatability
- Transparency
- Confidentiality

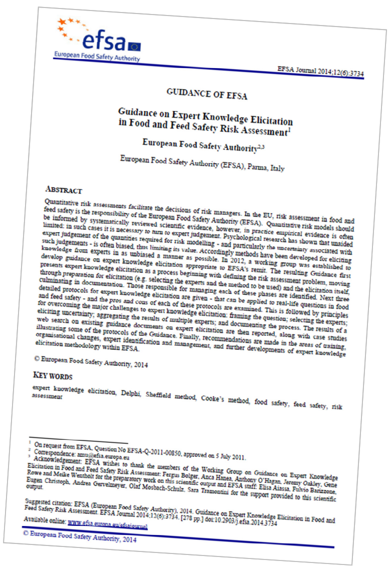


For more information…

- Guidance Document
- AMU Unit
- External specialists

EFSA (2014)

http://www.efsa.europa.eu/en/efsajournal/pub/3734.htm


efsa
European Food Safety Authority

EFSA Journal 2014;12(6):3734

GUIDANCE OF EFSA

Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment[1]

European Food Safety Authority[2,3]

European Food Safety Authority (EFSA), Parma, Italy

ABSTRACT

KEY WORDS

© European Food Safety Authority, 2014

Summary

- EKE is a process of
  - representing the judgements of experts
  - concerning an uncertain quantity
  - as a probability distribution
- EKE methods are formal, rigorous probabilistic judgement techniques
  - designed to encourage careful, thoughtful judgements
  - and reduce psychological biases
- EFSA EKE Guidance implements EKE in an efficient, rigorous and transparent manner
  - targetted on the most important uncertainties
  - subject to critical review at key decision points
  - fully documented
- EKE plays a key role in EFSA's Draft Guidance on Uncertainty

Major discussion points from this course

- *To be added during course…*

Practical 11: Planning EKE for examples from each participant's own area of work

Andy Hart, Tony O'Hagan,
John Quigley, Fergus Bolger

Objective

- Apply learning from the course to an assessment in your own area of work
- Identify potential challenges and discuss possible solutions
- Leave with a preliminary plan of how to proceed

Practical task 7.1

- Take an EFSA assessment from your area of work
- *Work individually, complete the template provided*

TEMPLATE FOR EKE COURSE PRACTICAL 11 Name........................

Example assessment:........................

Identify 3 or more uncertain parameters:

What:........................Why:........................

What:........................Why:........................

What:........................Why:........................

Which contributes most to the overall uncertainty of the assessment? Why?

Define a precise question for your selected parameter

How would you use results of EKE for this parameter in the assessment?

Practical 11

Types of expertise/experience | Names of potentially suitable experts

How many experts would be needed in total?........................

Which EKE method do you think is most suitable?........................

Why?........................

Challenges you expect to encounter | Possible solutions

Practical 11

- *And then…*

Practical task 7.2

- Discuss with your neighbours:
  - Lessons learned
  - Applicability of EKE to your areas of EFSA's work
  - Implications for current working practices
- *Identify key points for feedback to rest of group*
- Take your template away at the end of the course as a starting point for applying what you have learned

TEMPLATE FOR EKE COURSE PRACTICAL 11

Name…………………………………………………

Example assessment:……………………………………………………………………………

……………………

Identify 3 or more uncertain parameters:

What:………………………………………………..Why:………………………………

…………………………………

What:………………………………………………..Why:………………………………

…………………………………

What:………………………………………………..Why:………………………………

…………………………………

Which contributes most to the overall uncertainty of the assessment? Why?

……………………………………………………………………………………………

…………………………………..

……………………………………………………………………………………………

…………………………………..

Define a precise question for your selected parameter

……………………………………………………………………………………………

…………………………………..

……………………………………………………………………………………………

…………………………………..

FINAL

How would you use results of EKE for this parameter in the assessment?

……………………………………………………………………………………………………
…………………………………

Types of expertise/experience **Names of potentially suitable experts**

……………………………………………………………
……………………………………………………………

……………………………………………………………
………………………………………………………………

……………………………………………………………
………………………………………………………………

How many experts would be needed in total?…………………………………………………………

Which EKE method do you think is most suitable?……………………………………………………

Why?………………………………………………………………………………………
………………………………

……………………………………………………………………………………………………
…………………………………

Challenges you expect to encounter **Possible solutions**

……………………………………………………………
………………………………………………………………

……………………………………………………………………

……………………………………………………………………

……………………………………………………………………

……………………………………………………………………

……………………………………………………………………

……………………………………………………………………

Final plenary

Andy Hart, Tony O'Hagan,
John Quigley, Fergus Bolger

Topics for discussion

- What challenges do you expect in implementing EKE in your Panel/Unit?
- How will EKE fit into existing EFSA practices?
- Any general comments arising from the course?

Thank you for your participation!

Please complete the evaluation form
and leave it when you depart

Training Course on Expert Knowledge Elicitation
[Course Date/s] 2015

efsa European Food Safety Authority

EVALUATION FORM

Thank you for your participation in this training course. It would be very much appreciated if you could please complete the following questions with regards various aspects of the course.

Your comments and feedback are very important and valued. They ensure we are able to fully address any potential areas of concern promptly, and to help inform continuous improvement of the training.

Your responses are anonymous, unless you choose to indicate your name at the end of the form, and will be reviewed as each training course concludes to inform the refinement and development of future training in this topic area. Additionally, a summary of responses received across the complete programme of training will be included in the final evaluation report submitted to EFSA.

For each question, please circle the numerical rating or descriptive option that best fits your opinion. Specific comments, particularly to explain any low ranking ratings or to highlight aspects that you found of most value and which worked especially well, will help ensure we are able to apply learnings to future training.

In addition to completing this questionnaire, if you have a specific query and/or comments that you wish to discuss in person, please speak to a member of the training team at any point during the course.

Thank you for your time.

---

1 OVERALL EXPERIENCE

1.1 Did the course fully meet your expectations and requirements?	
No, not at all Yes, completely 1 2 3 4 5	Comments:
1.2 Have you reached the intended learning outcomes of the course?	
No, not at all Yes, completely 1 2 3 4 5	Comments:
1.3 Has the course facilitated your future work for EFSA?	
No, not at all Yes, completely 1 2 3 4 5	Comments:

2 CURRICULUM AND TEACHING

2.1 Are you satisfied with the content of the course?	
No, not at all Yes, completely 1 2 3 4 5	Comments:
2.2 Was the course material at the correct level for your needs?	
Too basic Just right Too advanced	Comments:
2.3 Are you satisfied with the balance of practical sessions versus lectures?	
No, not at all Yes, completely 1 2 3 4 5	Comments
2.4 Was sufficient time allocated for discussions with fellow participants and tutors?	
No, not at all Yes, completely 1 2 3 4 5	Comments
2.5 Are you satisfied with the teaching ability of the tutors?	
No, not at all Yes, completely 1 2 3 4 5	Comments
2.6 Are you satisfied with the professional and technical competence of the tutors?	
No, not at all Yes, completely 1 2 3 4 5	Comments
2.7 If you requested additional information, was this provided?	
Yes No	Comments
2.8 Which part/s of the course did you find most and/or least useful/instructive and why?	
Comments	

Training Course on Expert Knowledge Elicitation
[Course Date/s] 2015

European Food Safety Authority

3 COURSE ADMINISTRATION & VENUE

3.1 Did the overall organisation and administration associated with the course, prior to and during the training, meet your requirements?	
No, not at all Yes, completely 1 2 3 4 5	Comments:
3.2 Did the venue and training facility provided meet your requirements?	
No, not at all Yes, completely 1 2 3 4 5	Comments:
3.3 How relevant and user friendly were the training materials/hand outs?	
Very poor Very good 1 2 3 4 5	Comments
3.4 How suitable was the scheduling, including duration, of the training?	
Not at all Completely 1 2 3 4 5	Comments

ANY ADDITIONAL COMMENTS

Please add any other comments that you have or suggestions on how the course and/or administration/ organisation could be improved.

TESTIMONIAL

If you are willing to offer a short testimonial below regarding the training you have received, please write in the space below. This will help us illustrate the benefits of participation in similar training opportunities in the future.

____

____

____

____

____

Please tick appropriate statement to confirm permission as to use:

I agree to my name being included alongside the testimonial….
(Name ……………………………………...Job Title/Role……………………………………………….)

Please do not name me….

THANK YOU FOR COMPLETING THE QUESTIONNAIRE.

Please leave as indicated by your course tutor ahead of your departure.

Appendix C. Detailed results of participant evaluations

1.1 Did the course fully meet your expectations and requirements?

Score	June	August	September
1 (low)	0	0	0
2	1	1	0
3	2	1	0
4	5	15	7
5 (high)	4	12	12

1.2 Have you reached the intended learning outcomes of the course?

Score	June	August	September
1 (low)	0	0	0
2	0	0	0
3	1	7	0
4	8	14	10
5 (high)	3	6	9

1.3 Has the course facilitated your future work for EFSA?

Score	June	August	September
1 (low)	0	0	0
2	0	0	0
3	2	5	3
4	5	16	8
5 (high)	5	6	7

2.1 Are you satisfied with the content of the course?

Score	June	August	September
1 (low)	0	0	0
2	1	0	0
3	0	1	0
4	8	14	10
5 (high)	3	13	9

2.2 Was the course material at the correct level for your needs?

Score	June	August	September
1 (too basic)	0	1	0
2 (just right)	11	23	19
3 (too advanced)	0	1	0

2.3 Are you satisfied with the balance of practical sessions versus lectures?

Score	June	August	September
1 (low)	0	0	0
2	1	1	0
3	2	7	4.5
4	7	10	5.5
5 (high)	2	10	9

2.4 Was sufficient time allocated for discussions with fellow participants and tutors?			
Score	June	August	September
1 (low)	0	0	0
2	1	2	1.5
3	4	9.5	4.5
4	7	12.5	4
5 (high)	0	4	9

2.5 Are you satisfied with the teaching ability of the tutors?			
Score	June	August	September
1 (low)	0	0	0
2	0	0	0
3	1	0	0
4	2	8	4
5 (high)	9	20	15

2.6 Are you satisfied with the professional and technical competence of the tutors?			
Score	June	August	September
1 (low)	0	0	0
2	0	0	0
3	0	0	0
4	2	5	4
5 (high)	10	23	15

2.7 If you requested additional information, was this provided?			
Score	June	August	September
1 (yes)	8	14	18
2 (no)	0	1	0
3 (not applicable)	2	3	1

3.1 Did the overall organisation and administration associated with the course, prior to and during the training, meet your requirements?			
Score	June	August	September
1 (low)	0	0	0
2	0	0	0
3	1	1	1
4	4	11	4
5 (high)	7	16	14

3.2 Did the venue and training facility provided meet your requirements?			
Score	June	August	September
1 (low)	0	0	0
2	0	1	0
3	1	0	1.5
4	5	12	8.5
5 (high)	6	15	9

3.3 How relevant and user friendly were the training materials/hand outs?			
Score	June	August	September
1 (low)	0	0	0
2	0	1	0
3	1	0	0
4	4	16	8
5 (high)	7	12	11

3.4 How suitable was the scheduling, including duration, of the training?			
Score	June	August	September
1 (low)	0	0	0
2	0	3	0.5
3	4	10	4.5
4	5	9	7
5 (high)	3	4	7