

Methods for weight of evidence and expert elicitation

CAG Uncertainty meeting
5-6/3/18

What is the purpose of this exercise?

- Previous EFSA work has identified 73 pesticides as members of acute effects-CAG on functional effects on motor division
- The amount, reliability, relevance and consistency of the evidence varies
- Therefore there is uncertainty about which pesticides really belong to each CAG, with some being more likely to belong than others
- We aim to quantify this uncertainty so that it can be taken into account in risk assessment

The need for grouping

- Problem: assessing probability of membership for each pesticide *individually* is not feasible in the time available, and consistent weighing of the evidence would be difficult over so many pesticides
- Solution: assess membership probability for *groups* of pesticides that have *similar evidence*
- Groups formed by:
 - Identifying key criteria for summarising the evidence
 - Weight the criteria by relevance for membership
 - Calculate an overall score for each pesticide
 - Form groups of pesticides with similar scores
 - Check the groups are reasonable and not too heterogeneous

Definition of question for EKE

- WOE and EKE require a well-defined question
 - To ensure all experts are interpreting it the same way
 - ...and all probabilities refer to the same thing
 - ...and they can be used in probability calculations
- Well defined means:
 - The outcome should be *determinable in principle*
 - if suitable observations could be made, the outcome would be determined *unambiguously*
 - all observers would agree which outcome had occurred
 - But it's not essential that such observations are possible in practice

WG definition for motor division

As agreed at WG meeting in February:

- If the required set of studies (including neurotoxicity studies if relevant) was performed and reported perfectly, and the results were analysed and interpreted according to the standard procedure, would this chemical be assessed as positive for the motor division?

Weighting and scoring approach

- Evidence criteria were defined at the February meeting
- Weights were defined for each criterion
 - The weights are an approximate relative measure of the contribution that positive evidence for each criterion would make to increasing the probability of the chemical causing motor division effects
- The weights were multiplied together to obtain a total score for each pesticide
- The pesticides were ranked by total score and divided into groups (see Olaf's presentation)
- WG should now review the groupings and consider whether the weightings are okay or need revising

Note about weighting

- The purpose of the weighting and scoring is not to provide a measure of membership *per se* (this is done later, when we elicit probabilities) but to group pesticides for which the evidence of membership is of similar weight
 - Heterogeneity and/or misplacements in the resulting groups are taken into account later, when eliciting the probabilities

Expert elicitation methods

- Expert judgement is used to assess how many of the pesticides in each group really cause the effect, taking account of all the evidence
- The answer is uncertain so a distribution was elicited to quantify the uncertainty
 - Need a discrete distribution as the answer must be a whole number between zero and the number of pesticides in each group
 - The '*roulette*' method for elicitation was chosen because it allows us to elicit the discrete distribution directly, and seems to fit well with how the WG members think about their judgements

Reviewing the evidence

- For each group of pesticides, the experts will be given a sheet summarising the basic information about each pesticide, and its criteria scores
- Before making judgements, we will briefly review the evidence summary and clarify any questions about it

Individual judgements

- The same process will be used for each group of pesticides
- First the experts will be asked to work individually
 - Review the evidence
 - Place counters ('probs') on a template to build a distribution representing their judgement about the number of this group that cause motor division effects
- Then the distributions will be displayed together

Consensus judgements

- The facilitator leads a discussion to share and compare the individual judgements and rationales
- Then the facilitator leads a discussion aimed at agreeing a consensus distribution for the group
 - *What a rational independent observer might think after hearing the discussion and seeing the evidence*
- The WG will also be asked to agree a text summary of the reasoning for the consensus distribution

Review of elicitation protocol and training

Quantifying uncertainty by expert judgement

Expert Knowledge Elicitation

‘A systematic, documented and reviewable process to retrieve expert judgements from a group of experts in the form of a probability distribution’

EFSA EKE Guidance 2014



Expert knowledge elicitation (EKE)

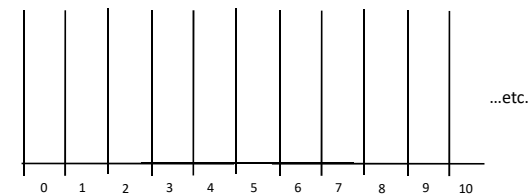
- Expert judgement is...
 - *not* guesswork!
 - *careful, reasoned, evidence-based, transparent*
 - a vital component of all EFSA assessments
 - essential and unavoidable for assessing uncertainty
 - subjective, and subject to psychological biases
- EKE methods are designed to...
 - mitigate the impact of the psychological biases
 - take account of the available evidence
 - help experts reach well-founded judgements
 - make the process as objective & transparent as possible

“Roulette” method for EKE

- Proposed for this assessment because
 - We are considering a specific set of substances, not a sample from a larger ‘population’
 - So question is how many of these are positive, rather than what proportion of a wider population
 - Asking for the number may fit better with how experts think than asking for the proportion
 - The true number of positives must be a whole number, fractions, so we need to elicit a discrete distribution, which the roulette method is well suited to

The Roulette grid

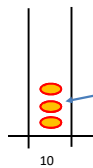
- You have a grid comprising a number of columns
- These represent the possible values of X, the true number of pesticides in the group for which the answer to the EKE question is ‘yes’
- Minimum = 0, maximum = no. of pesticides in group



Adapted from slides created by Prof A O'Hagan

Probs

- You also have 20 counters called *probs*
 - Because each one represents an amount of probability
 - As you have 20 probs, each represents a probability of 0.05 (5%)
- You are asked to place the probs in the bins on your grid, to specify your knowledge and beliefs about X



This expert has 20 probs, so each is worth 0.05
With these 3 probs she specifies that her probability that X is 10 is 0.12

Adapted from slides created by Prof A O'Hagan

SHELF v3.0

Roulette

17

The probability scale

- 0 to 1, or 0% to 100%

0% probability Impossible ← Less likely → More likely → 100% probability Certain



- Defined meaning, e.g.:**
 - 50% = *the expert judges* both outcomes to be *equally likely* (odds 1:1)
 - 66% = positive outcome *twice as likely* (odds 2:1 on)
 - 75% = positive outcome *3x as likely* (odds 3:1 on)
 - etc.

How to proceed (for CAG case)

- First decide what range of numbers you consider plausible and write these on the *results sheet* provided
 - The highest number you consider plausible
 - The lowest number you consider plausible
- Now use the *grid sheet* and consider each column between your lowest and highest plausible numbers
- Distribute your 'probs' among the columns to express your judgement about the probability of each number

Optional additional strategies

- Are there some pesticides you are virtually certain *do* cause motor division effects? If so, then your lower plausible bound must be this number or higher
- Are there some pesticides you are virtually certain *do not* cause motor division effects? If so, then your upper plausible bound must exclude at least this number
- If it helps, annotate the evidence sheet to identify which pesticides you think are more likely (+, ++, etc.) or less likely (-, --, etc.) to cause motor division effects
- If you think the group would be better divided into subsets, please raise this for discussion
- If you'd rather use more than 20 probs to specify your distribution more precisely, please ask

Now check that ...

- ... you have placed all your probs
- ... you have not placed any probs in bins that are outside your plausible range
 - If you have allocated one or more probs to the bins representing the lower and upper bounds of your plausible range, should your plausible range be wider?
- ... you are happy with the probability implied by the number of probs you have placed in each bin
 - Look at groups of bins and check that you are happy with their aggregate probabilities
 - Look at bins with the same numbers of probs and check that you think they are equally likely to be X
- And write your results (number of probs and corresponding probabilities) on your *results sheet*

Adapted from slides created by Prof A O'Hagan

Discussion

- Any questions about any aspect of this?
- Any worries or concerns?
- Ready to start?