Radiation Protection Dosimetry vol. 90, No.3 303-309

Procedures Guide for Structured Expert Judgement in Accident Consequence Modelling

R.M. Cooke and L.H.J. Goossens

Delft University of Technology, Mekelweg 4, NL-2628 CD Delft, the Netherlands

COS 02

ABSTRACT

This paper sketches a protocol for using structured expert judgement to generate uncertainty data for uncertainty analyses. The paper emphasises the use of performance based weighting as an instrument to enable optimisation of the aggregated experts' assessments. Examples are shown from the EC/USNRC joint study on Probabilistic Accident Consequence Uncertainty Analysis.

1 INTRODUCTION

Governmental bodies are confronted with the problem of achieving rational consensus in the face of substantial uncertainties. The area of accident consequence management for nuclear power plants affords a good example. Decisions with regard to evacuation, decontamination, and food bans must be taken on the basis of predictions of environmental transport of radioactive material, contamination through the food chain, cancer induction, and the like. These predictions use mathematical models containing scores of uncertain parameters. Decision makers want to take, and want to be perceived to take, these decisions in a rational manner. The question is, how can this be accomplished in the face of large uncertainties? This paper describes the European Guide for Expert Judgement in Uncertainty Analysis(1) and the results of the performance based analysis of experts' assessment in the joint EC/USNRC study on Uncertainty analysis of nuclear probabilistic accident consequence codes.

2 WHAT IS UNCERTAINTY?

Uncertainty is that which is removed by becoming certain. In practical scientific and engineering contexts, certainty is achieved through observation, and uncertainty is that which is removed by observation. Hence uncertainty is concerned with the results of possible observations. Uncertainty must therefore be distinguished from ambiguity. Ambiguity is removed by linguistic conventions regarding the meaning of words. To be studied quantitatively, uncertainty must be provided with a mathematical representation, for instance, as probability.

Within the subjective interpretation of probability, uncertainty is a degree of belief of one person, and can be measured by observing choice behaviour. Viewed from the theory of rational decision(2) one subjective probability is as good as another. There is no rational mechanism for persuading individuals to adopt the same degrees of belief. In practice, however, decision makers rely on experts' subjective assessments.

3 STRUCTURED EXPERT JUDGEMENT

Expert judgement has always played a large role in science and engineering. Increasingly, expert judgement is recognised as just another type of scientific data, and methods are developed for treating it as such. Summaries are given in references(3,4,5).

For applications in uncertainty analysis, we are mostly concerned with uncertain quantities taking values in some continuous range. Our uncertainty is therefore described by a subjective probability distribution for uncertain quantities with values in a continuous range.

When expert judgements are cast in the form of distributions of uncertain quantities, the issues of conditionalisation and dependence are important. When uncertainty is quantified in an uncertainty analysis, it is always conditional on something. It is essential to make clear the background information conditional on which the uncertainty is to be assessed. This is the role of the "case structure". The case structure document describes in greater detail which areas of interest are to be assessed by the experts. From the case structure document the questionnaire on the elicitation variables is derived. For that reason the case structure document should also include a section on which conditions the questions are based and which issues should be taken into account in the uncertainty assessments and which issues are not part of that.

The expert judgement protocol has the following steps:

Preparation for elicitation:

(1) Definition of case structures document describing the field of interest for which expert judgements will be required: in general, this will be risk assessment codes or accident consequence codes.

(2) Identification of target variables: these are the variables whose uncertainty must be quantified through formal expert judgement.

(3) Identification of the query variables: these are the variables to be assessed by the experts. These variables must be observable. If a target variable can in principle be measured by a procedure with which experts are familiar, then these are also query variables. Target variables for which no such measurement procedures exist cannot be quantified by direct elicitation. For these variables other derived elicitation variables must be found. The uncertainty distributions over these derived elicitation variables must then be pulled back via probabilistic inversion onto the target variables (see step (14)).

(4) Identification of performance variables (or seed variables) to be assessed by the experts (that is the major issue of this paper).

(5) Identification of experts.

(6) Selection of experts.

(7) Definition of elicitation format document describing the exact questions and format for the experts elicitations.

(8) Dry run exercise describing the try out of the elicitation format document to a few experts.

(9) Expert training session describing the ingredients of training experts in preparing probabilistic assessments.

Elicitation:

(10) Expert elicitation session, whereby the experts' individual judgements are discussed in the presence of a normative analyst (experienced in probability issues) and a substantive analyst (experienced in the expert's field of interest).

Post-elicitation:

(11) Combination of experts' assessments describing the methods with which the individual expert assessments will be aggregated to one combined assessment. Aggregation can be done by equally weighting all experts (as is done for the uncertainty analysis of COSYMA described in this special issue) or by using performing based weighting (as explained in this paper).

(12) Robustness and discrepancy analysis describing the procedures to show the robustness of the combined results. Robustness analysis is available on experts and seed variables. Experts/seed variables are removed from the data set one at the time and the decision maker is recalculated, to account for the relative information loss to the original decision maker. If that loss is large, then results may not be replicated if another study were to be done using different experts and seed variables. Discrepancy analysis identifies items on which the uncertainty assessments of the experts differ most. These items should be reviewed to ascertain any avoidable causes of discrepancy.

(13) Feed back communication with the experts.

(14) Post-processing analyses describing the methods for processing the uncertainties of the combined expert assessments (resulting from the query variables, defined in step 3) into uncertainties on the target variables from step 2.

(15) Documentation of the results.

4 PERFORMANCE MEASURES AND RATIONAL CONSENSUS

The goal of applying structured expert judgement techniques is to enhance rational consensus. Necessary conditions for achieving this goal are laid down as methodological principles[(5)]:
Scrutability/accountability: All data, including experts' names and assessments, and all processing tools are open to peer review and results must be reproducible by competent reviewers.
Empirical control: Quantitative expert assessments are subjected to empirical quality controls.

Neutrality: The method for combining/evaluating expert opinions should encourage experts to state their true opinions, and must not bias results.
Fairness: Experts are not pre-judged, prior to processing the results of their assessments.
We claim that these are *necessary* conditions for rational consensus, we do not claim that they are sufficient as well. Hence, a rational subject could accept these and yet reject a method which implements them. In such a case, however, (s)he incurs a burden of proof to formulate additional conditions for rational consensus which the method putatively violates.

The requirement of empirical control will strike some as peculiar in this context. How can there be empirical control with regard to expert subjective probabilities? To answer this we must reflect on the question 'when is a problem an expert judgement problem?' We would not have recourse to expert judgement to determine the speed of light in a vacuum. This is physically measurable and has been measured to everyone's satisfaction. Any experts we queried would give the same answer. Neither do we consult expert judgement to determine the existence of god. There are no experts in the operative sense of the word for this issue. A problem is susceptible for expert judgement, if there is relevant scientific expertise. This entails that there are theories *and* measurements relevant to the issues at hand, but the quantities of interest themselves cannot be measured in practice. For example, toxicity of a substance for humans is measur*able* in principle, but is not measured for obvious reasons. However, there are toxicity measurements for other species which might be relevant to the question of toxicity in humans. Or again, we may be interested in the dispersion of a toxic airborne release at 50 km from the source. Although it is practically impossible to measure the plume spread at 50 km, it is possible to measure this spread at 1 km. If a problem is an expert judgement problem, then necessarily there will be relevant experiments which can in principle be used to enable empirical control.

5 THE CLASSICAL MODEL

The above principles have been operationalised in the so called Classical Model, a performance based linear pooling or weighted averaging model. The weights are derived from experts calibration and information performance, as measured on calibration or seed variables. These are variables from the experts' field whose values become known to the experts post hoc. Seed variables serve a threefold purpose: (i) to quantify experts' performance as subjective probability assessors, (ii) to enable performance-optimised combinations of expert distributions, and (iii) to evaluate and hopefully validate the combination of expert judgements. The name "classical model" derives from an analogy between calibration measurement and classical statistical hypothesis testing. It contrasts with various Bayesian models.

The performance based weights use two quantitative measures of performance, calibration and information. Calibration measures the statistical likelihood that a set of experimental results correspond, in a statistical sense, with the experts assessments[1]. Loosely, the calibration score is the probability that the divergence between the expert's probabilities and the observed values of the seed variables might have arisen by chance. A low score (near zero) means that it is likely, in a statistical sense, that the expert's probabilities are 'wrong'. Similarly a high score (near one, but bigger than, say, 0.05) means that the expert's probabilities are statistically supported by the set of seed variables. Information represents the degree to which an expert's distribution is concentrated, relative to some user-selected background measure[2]. "Good expertise" corresponds to good calibration (high statistical likelihood) and high information. The weights in the classical model are proportional to the product of statistical likelihood and information. When a combined expert has been formed, we can also measure the calibration and information of this combined expert. For more detail see Cooke[(5)].

In the classical model calibration and information are combined to yield an overall or combined score with the following properties:

[1] In particular, the calibration score is the p-value of a standard Chi square goodness of fit test.
[2] The overall information score is the mean of the information scores for each variable. This is proportional to the information in the expert's joint distribution relative to the joint background measure, under the assumption of independence. Independence in the experts' distributions means that the experts would not revise their distributions for some variables after seeing realisations for other variables. Scoring calibration and information under the assumption of independence reflects the fact that expert learning is not a primary goal of the study.

1. Calibration dominates over information, information serves to modulate between more or less equally well calibrated experts,
2. The score is a long run proper scoring rule, that is, an expert achieves his/her maximal expected score, in the long run, by and only by stating his/her true beliefs. Hence, the weighting scheme, regarded as a reward structure, does not bias the experts to give assessments at variance with their real beliefs, in compliance with the principle of neutrality.
3. Calibration is scored as ‘statistical likelihood with a cut-off’. An expert is associated with a statistical hypothesis, and the seed variables enable us to measure the degree to which that hypothesis is supported by observed data. If this likelihood score is below a certain cut-off point, the expert is unweighted. The use of a cut-off is driven by property (2) above. Whereas the theory of proper scoring rules says that there must be such a cut off, it does not say what value the cut-off should be.
4. The cut-off value for (un)weighting experts is determined by optimising the calibration and information performance of the combination.

A fundamental assumption of the Classical model (as well as Bayesian models) is that the future performance of experts can be judged on the basis of past performance, as reflected in the seed variables. Seed variables enable empirical control of any combination schemes, not just those which optimise performance on seed variables. Examples of expert judgement studies using seed variables are available(6,7,8,9). Therefore, choosing good seed variables is of general interest, see Goossens et al(10) for background and detail.

6 RESULT OF PERFORMANCE IN THE EC/USNRC STUDY

The classical model in the eight expert panels shown in Table 1.The experts for each panel are internationally recognised in their fields, and were selected according to the method described in the subsequent EC/USNRC-reports(11,12,13,14,15,16,17) . Seed variables were available for all panels except for study (13)[3]. The seed variables for the Late Health Effects panel are defined in terms of the follow-up of the Nagasaki and Hiroshima survivors, to be published in 2001. Hence the values of these variables are not available at present. For the other panels seed variables were queried. Table 1 shows the performance based combination and the equal weight combination for the other seven panels. For each panel, Table 1 shows the calibration score (1 is maximal, 0 is minimal), the mean information score (0 is minimal), and the 'virtual weight'. Virtual weight is the weight that the combination would receive if added to the expert panel as an additional virtual expert. A virtual weight of one half or more indicates that the combination would receive more weight than the real experts cumulatively.

CASE	WEIGHTING	Calibr.	Mean inform	Number seed	virtual weight
DISPERSION	Perform	0.90000	1.024	23	0.80545
	Equal	0.15000	0.811	23	0.33166
DRY DEPOSITION	Perform	0.52000	1.435	14	0.50000
	Equal	0.00100	1.103	14	0.00168
WET DEPOSITION	Perform	0.25000	1.117	19	0.93348
	Equal	0.00100	0.793	19	0.07627
ANIMAL	Perform	0.75000	2.697	8	0.50000
	Equal	0.55000	1.778	8	0.19204
SOIL/PLANT	Perform	0.00010	1.024	31	0.13369
	Equal	0.00010	0.973	31	0.12779
INTERNAL DOSE	Perform	0.85000	0.796	55	0.52825
	Equal	0.11000	0.560	55	0.09217
EARLY HEALTH	Perform	0.23000	0.216	15	0.98749
	Equal	0.07000	0.165	15	0.94834
LATE HEALTH	Equal	*******	0.280	0	0

Table 1 Performance based and equal weight combinations

Apart from the SOIL/PLANT case, the performance based combination performs well; the calibration scores are not alarmingly low, and the virtual weight is high. The equal weight combination sometimes returns good calibration and high virtual weight, but these scores are lower than those of the performance based combination. In the case of SOIL/PLANT, we must conclude that the evidence

[3] This study’s constraints precluded the collection of seed variables.

gathered from the seed variables does not establish the desired confidence in the results.[4] In DISPERSION, ANIMAL and INTERNAL DOSE, the results of equal weighting are not dramatically inferior to the performance based combination. In such cases, a decision maker giving priority to *political* rather than *rational* consensus might apply equal weight combination without raising questions of performance. In the other cases the evidence for degraded performance in the equal weight combination, in our opinion, is strong. Table 2 shows the individual expert scores for the results in Table 1.

DISPERSION

Expt	Cal	Mean Inf	# seed
1	0.0001	2.078	23
2	0.0001	1.594	23
3	0.0010	1.504	23
4	0.1300	1.286	23
5	0.0300	1.092	23
6	0.0050	1.590	23
7	0.0100	1.508	23
8	0.0200	1.840	23
Prf	0.9000	1.024	23
Eq	0.1500	0.811	23

DRY DEPOSITION

Expt	Cal	Mean Inf	# seed
1	0.0001	1.953	14
2	0.5200	1.435	14
3	0.0010	1.702	14
4	0.0010	1.732	14
5	0.0001	1.792	14
6	0.0010	2.234	14
7	0.0010	1.695	14
8	0.0005	1.985	14
Prf	0.5200	1.435	14
Eq	0.0010	1.103	14

WET DEPOSITION

Expt	Cal	Mean Inf	# seed
1	0.0001	2.638	19
2	0.0100	1.979	19
3	0.0010	1.009	19
4	0.0001	1.028	19
5	0.0010	1.565	19
6	0.0001	1.946	19
7	0.0001	1.252	19
Prf	0.2500	1.117	19
Eq	0.0010	0.793	19

ANIMAL

Exprt	Cal.	Mean Inf	# seed
1	0.00100	2.658	8
2	0.00100	2.730	8
3	0.09000	1.689	8
4	0.75000	2.697	8
5	0.01000	2.835	6
6	0.64000	2.888	8
7	0.02000	2.821	7
Prf	0.75000	2.697	8
Eq	0.55000	1.778	8

SOIL/PLANT

Expt	Cal	Mean Inf	# seed
1	0.0001	2.376	31
2	0.0001	1.309	31
3	0.0001	1.346	31
4	0.0001	1.607	31
Prf	0.0001	1.024	31
Eq	0.0001	0.973	31

INT. DOSIMETRY

Expt	Cal	Mean Inf	# seed
1	0.0010	1.671	39
2	0.7300	0.822	55
3	0.0001	2.003	50
4	0.0001	2.366	39
5	0.0001	1.205	39
6	0.0050	0.838	28
Prf	0.8500	0.796	55
Eq	0.1100	0.560	55

EARLY HEALTH

Expt	Cal	Mean Inf.	# seed
1	0.0001	0.834	15
2	0.0001	1.375	15
3	0.0001	1.008	15
4	0.0001	0.966	15
5	0.0001	1.115	15
6	0.0001	0.573	15
7	0.0001	0.410	15
Prf	0.2300	0.216	15
Eq	0.0700	0.165	15

LATE HEATH

Expt	Cal	Mean Inf	# seed
1	*****	0.440	0
2	*****	1.379	0
3	*****	1.024	0
4	*****	0.507	0
5	*****	0.836	0
6	*****	0.599	0
7	*****	0.616	0
8	*****	0.988	0
Eq	*****	0.280	0

Table 2. Individual scores

The mean information of the performance based combination is usually slightly lower than that of the least informative experts, and the calibration score is typically substantially higher. This reflects the dominance of calibration over information in this weighting scheme. The equal weight combination has wider confidence bands still, and the calibration is typically lower than the best calibrated experts. Inspecting the data in Table 2, we see that the performance based combination for DRY DEPOSITION and ANIMAL, actually coincides with one of the experts. In other words, performance is optimised by assigning weight one to a single expert. This naturally raises the question of robustness with regard to expert choice. How much would the results differ if this one expert happened not to be available? One way to address this question is to repeat the analyses, leaving this expert out. If the differences between the original and the 'perturbed' combination are smaller than the differences among the experts themselves and if the performance is still acceptable, then there is no strong indication that the results are not robust against choice of experts. Table 3 shows the results of these comparisons. Experts are excluded one at a time and the performance based combination is recalculated. Columns 2 and 3 show the mean information and calibration of the 'perturbed' combination. Column 4 shows the relative information of the ‘perturbed’ expert with respect to the original combination. The differences of the experts among themselves are reflected in the last column, which shows the relative information of each expert with respect to the equal weight combination.

---

[4] Although it might be argued that 31 seed variables constitutes a rather sever test of calibration, reducing the effective number of seed variables to 10 still yields poor performance (calibration scores 0.04 and 0.01 for the performance based and equal weight combinations respectively). In general, the number of effective seed variables is equal to the minimum number assessed by some expert. Hence the effective number in INTERNAL DOSIMETRY is 28 and in ANIMAL is 6. Experts are scored on the basis of the effective number of seed variables; lowering this number is comparable to lowering the power of a statistical test. Thus we cannot directly compare calibration scores of different panels without first setting the effective number of seed variables equal.

ROBUSTNESS ON EXPERTS: ANIMAL

Expert Excluded	Mean Inf.	Calibration	Rel.Inf Original	Rel.Inf Eq.Wgt
None	2.697	0.750	0	
1	2.697	0.75000	0	1.084
2	2.697	0.75000	0	0.987
3	2.045	0.75000	0	0.374
4	2.695	0.64000	0.569	0.719
5	2.697	0.70000	0	0.835
6	2.690	0.75000	0	0.818
7	2.697	0.75000	0	0.988

ROBUSTNESS ON EXPERTS: DRY DEPOSITION

Expert Excluded	Mean Inf	Calibration	Rel.Inf Original	Rel.Inf Eq.Wgt
None	1.435	0.520	0	
1	1.435	0.52000	0	0.852
2	1.245	0.05000	0.858	0.420
3	1.435	0.52000	0	0.555
4	1.435	0.52000	0	0.608
5	1.435	0.52000	0	0.651
6	1.446	0.52000	0	1.137
7	1.431	0.52000	0	0.618
8	1.435	0.52000	0	0.860

Table 3. Robustness on experts, performance based combination.

We see from Table 3, that the robustness on experts for ANIMAL is satisfactory in the sense that the largest entry in column 4 is smaller than all but one entry of column 5. Robustness for DRY DEPOSITION is marginal. Lack of robustness is always a danger when performance is optimised. The equal weight combination is almost always more robust, but the price of course is lower performance.

Finally, Table 4 compares cancer risks at various cites of the EC-USNRC study with those of other studies, for high dose, high dose-rate. These results are obtained from the LATE HEALTH panel and hence reflect the equal weight combination.

	EC-USNRC (+90% confidence)[5]	BIER V[6]	ICRP 60[7]	UNSCEAR[8]	COSYMA[9]
BONE	0.035 (<0.001, 0.88)	-	-	-	0.01
COLON	0.98 (0.011, 3.35)	-	3.24	0.6	2.24
BREAST	0.78 (0.11, 3..78)	0.35	0.97	1.0	0.80
LEUKEMIA	0.91 (0.026, 2.33)	0.95	0.95	1.1	0.52
LIVER	0.086 <0.001, 2.02)	-	-	1.2	-
LUNG	2.76 (0.59, 8.77)	1.70	2.92	2.50	0.90
PANCREAS	0.17 (<0.001, 1.26)	-	-	-	-
SKIN	0.039 (<0.001, 0.37)	-	0.03	-	0.01
STOMACH	0.30 (<0.001, 4.01)	-	0.51	1.4	-
THYROID	0.059 (,0.001, 0.71)	-	-	-	0.17
ALL OTHER	2.60 (<0.001, 10.8)	-	-	-	-
ALL CANCERS	10.2 (3.47, 28.5)	7.90	12.05	12.0	5.02

Table 4. Comparison of elicited high dose and high dose-rate lifetime low LET cancer risks for a general EU/US population with those derived from other sources ($10^{-2}Gy^{-1}$)

Although the median values of the EC-USNRC study generally agree with the values from the other studies in Table 4, the 90% central confidence intervals are sometimes significantly wider than the spread of values from these studies. Indeed, the spread of assessments in the last four columns of table 4 is *not* an assessment of uncertainty.

6 CONCLUSIONS

We collect a number of conclusions regarding the use of structured expert judgement.

[5] Radiation exposure-induced deaths (REID) for joint current EU-US population (high dose, high dose-rate)
[6] BIER V calculates excess cancer deaths for current US population
[7] ICRP calculates REID average of risks for current UK and US populations.
[8] UNSCEAR calculates REID for current Japanese population.
[9] COSYMA default values (low dose and low dose-rate)

1. Experts' subjective uncertainties may be used to advance rational consensus in the face of large uncertainties, in so far as the necessary conditions for rational consensus are satisfied.
2. Empirical control of experts' subjective uncertainties is possible.
3. Experts' performance as subjective probability assessors is not uniform, there are significant differences in performance.
4. Experts as a group may show poor performance.
5. A structured combination of expert judgement may show satisfactory performance, even though the experts individually perform poorly.
6. The performance based combination generally outperforms the equal weight combination.
7. The combination of experts' subjective probabilities, according to the schemes discussed here, generally has wider 90% central confidence intervals than the experts individually; particularly in the case of the equal weight combination.

We note that poor performance as a subjective probability assessor does *not* indicate a lack of substantive expert knowledge. Rather, it indicates unfamiliarity with quantifying subjective uncertainty in terms of subjective probability distributions. Experts were provided with training in subjective probability assessment, but of course their formal training does not (yet) prepare them for such tasks.

ACKNOWLEDGEMENTS

The development of the Procedures Guide has been made possible by the support of the CEC under contract ETNU-CT93-0104-NL for Directorate-general XI (Environment and Nuclear Safety). The authors very much appreciate the stimulating discussions with others involved in the expert judgement exercises, in particular with Dr. G.N. Kelly responsible for the CEC/USNRC joint project on Uncertainty Analysis of Probabilistic Accident Consequence Codes (DG-XII-F-6). We gratefully acknowledge the enthusiastic and wholehearted participation of all the experts in this study. All experts provided written rationales explaining how they arrived at their assessments. These rationales, reproduced in the published reports, are highly recommended to those readers who wish deeper insight into the sources of uncertainty in these issues. We also gratefully acknowledge the help of many institutes, in particular the National Radiological Protection Board, the Forschungszentrum Karlsruhe and the Energy Centrum Netherlands, in developing seed variables.

REFERENCES

(1) Cooke RM and Goossens LHJ. Procedures guide for structured expert judgement. Report EUR 18820. Brussels-Luxembourg. To be published 1999
(2) Savage L. The foundation of statistics. Dover, New York, 1954
(3) Hogarth R. Judgement and choice. Wiley, New York, 1987.
(4) Granger Morgan M, Henrion M. Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge University Press, Cambridge, 1990
(5) Cooke RM. Experts in uncertainty, Cambridge, 1991
(6) Cooke RM, Mendel M, Thys W. Calibration and information in expert resolution: a classical approach. Automatica 1988: **24** 87-94
(7) Goossens LHJ, Cooke RM, Woudenberg F, van der Torn P. Expert judgement and lethal toxicity of inhaled chemicals. J Risk Res 1998: **1**(2) 117-133
(8) Cooke RM and Jager E. Failure frequency of underground gas pipelines: methods for assessment with structured expert judgement. Risk Analysis 1998: **18**(4) 511-527
(9) Goossens LHJ and Cooke RM. Applications of some risk assessment techniques: Formal expert judgement and accident sequence precursors. Safety Science 1997: **26**(1/2) 35-48
(10) Goossens LHJ, Cooke RM, and Kraan, BCP, Evaluation of weighting schemes for expert judgement studies, Final report prepared under contract Grant No. Sub 94-FIS-040 for the Commission of the European Communities., Directorate General for Science, Research and Development XII-F-6, Delft University of Technology, Delft, the Netherlands, 1996.
(11) Harper FT, Goossens LHJ, Cooke RM, et al. Probabilistic accident consequence uncertainty analysis: Dispersion and deposition uncertainty assessment. Report NUREG/CR-6244, EUR 15855, Washington/USA, and Brussels-Luxembourg, 1995
(12) Cooke RM, Goossens LHJ, Kraan BCP. Methods for CEC\USNRC accident consequence uncertainty analysis of dispersion and deposition - Performance based aggregating of expert judgements and PARFUM method for capturing modelling uncertainty. Report EUR 15856, 1995
(13) L.H.J. Goossens, J. Boardman, F.T. Harper, B.C.P. Kraan, M.L. Young, R.M. Cooke, S.C. Hora and J.A. Jones, *Probabilistic accident consequence uncertainty analysis: Uncertainty assessment*

for deposited material and external doses, Prepared for U.S. Nuclear Regulatory Commission and Commission of European Communities, NUREG/CR-6526, EUR 16772, Washington/USA, and Brussels-Luxembourg, (Volumes 1 and 2), 1997.

(14) J. Brown, L.H.J. Goossens, F.T. Harper, E.H. Haskin, B.C.P. Kraan, M.L. Abbott, R.M. Cooke, M.L. Young, J.A. Jones, S.C. Hora and A. Rood, *Probabilistic accident consequence uncertainty analysis: Food chain uncertainty assessment,* Prepared for U.S. Nuclear Regulatory Commission and Commission of European Communities, NUREG/CR-6523, EUR 16771, Washington/USA, and Brussels-Luxembourg, (Volumes 1 and 2), 1997.

(15) F.E. Haskin, L.H.J. Goossens, F.T. Harper, J. Grupa, B.C.P. Kraan, R.M. Cooke and S.C. Hora, *Probabilistic accident consequence uncertainty analysis: Early health uncertainty assessment*, Prepared for U.S. Nuclear Regulatory Commission and Commission of European Communities, NUREG/CR-6545, EUR 16775, Washington/USA, and Brussels-Luxembourg, (Volumes 1 and 2), 1997.

(16) M. Little, C.M. Muirhead, L.H.J. Goossens, F.T. Harper, B.C.P. Kraan, R.M. Cooke, and S.C. Hora, *Probabilistic accident consequence uncertainty analysis: Late health uncertainty assessment,* Prepared for U.S. Nuclear Regulatory Commission and Commission of European Communities, NUREG/CR-6555, EUR 16774, Washington/USA, and Brussels-Luxembourg, (Volumes 1 and 2), 1997.

(17) L.H.J. Goossens, J.D. Harrison, F.T. Harper, B.C.P. Kraan, R.M. Cooke and S.C. Hora, *Probabilistic accident consequence uncertainty analysis: Internal dosimetry uncertainty assessment*, Prepared for U.S. Nuclear Regulatory Commission and Commission of European Communities, NUREG/CR-6571, EUR 16773, Washington/USA, and Brussels-Luxembourg, (Volumes 1 and 2), 1998.