


<b>Subject:</b>	<b>ORGAPET Section C4: Expert Judgement Annex C4-3: Guidelines for the use of a modified Nominal Group Technique for eliciting expert opinion on the impact of organic farming and alternative agri-environmental schemes on policy objectives, illustrated by a case study of Wales.</b> <b>(Version 4)</b>	
<b>Author(s):</b>	<b>Ian Jeffreys</b> University of Wales, Aberystwyth (UWA), UK	<b>Specific Support Action project:</b> <b>European Action Plan for organic food and farming</b>
<b>Date:</b>	<b>October 31, 2006</b>	<b>FP 6 – 006501</b>

## 1 Introduction

This Annex presents the Nominal Group Technique (NGT – see ORGAPET [Section C4](#) for background) used to elicit expert opinion regarding the performance of the Welsh Organic Farming Scheme and the Tir Gofal countryside stewardship scheme. The performance of the policies is evaluated with respect to a set of indicators that reflect agri-environmental and rural development policy objectives. The focus of this Annex is the method for eliciting expert opinion regarding the performance of schemes rather than how the schemes perform against the indicators. This paper also provides a brief description of the evaluation indicators, detailed discussion of these are presented elsewhere.

## 2 The basics of NGT

The basis for using this approach is that, when other data sources are not available, expert opinion can be used to evaluate the options. Opinion can be gathered for each of the criteria from a single expert or many experts. When eliciting opinion from a single expert the person with greatest level of expertise is asked to evaluate all the options against the criteria they are most familiar with, different persons would be asked to evaluate different criteria. Alternatively multiple experts can be used. With NGT, an expert panel is convened with knowledge in all or a sub-set of the criteria, this panel then provides the evaluations. The panel members often have divergent opinions and a mechanism to address this is required.

The technique involves, in a workshop context, the private generation of ideas, followed by a discussion of these ideas, and finally a review of the ideas and the development of a consensus view. In the process of evaluating an effects table (see ORGAPET [Section C3](#)), the experts would score the effects table privately, the results would be collated, and presented to the group. Scores which diverge greater than a agreed limit would be discussed, assumptions stated and unified, then a re-score completed. A networked computer system is used to collate and present opinion and discussions are then focused on the areas of diverse opinion. The experts then re-score the table, if there is still divergence opinion and time allows a second discussion session occurs. The

computer requirements for this technique include a PC for each panellist plus ideally two PCs for systems operator, these PCs need to be networked and running MS Excel.

The presence of all the experts together in a workshop means that with NGT, the process can be completed in a considerably shorter time-frame, but requires all the participants to be present and there is also no time available for further data to be gathered outside the workshop context.

### 3 Scoring regimes

The data gathered using expert opinion can take a number of forms, but often the form of score out of X, typically 5 or 10 points are used. These can be interpreted in a number of number of ways, commonly the mid-point is described as no change from the status-quo to highest score the maximum amount of improvement in state and the minimum value being the maximum amount of degradation in state. These may be expressed in the terms of qualitative statements using plus and minus signs. This technique is amongst those offered in the DEFINITE software.

“The – – –/+++ scale is often used to describe qualitative effects. This is a useful method, as long as the results are interpreted with a degree of caution. It is important to ensure that the scores are not interpreted out of context. Generally the pluses and minuses mean the following:

– – – : very big negative effect

– – : big negative effect

– : small negative effect

0 : no effect

+ : small positive effect

++ : big positive effect

+++ : very big positive effect” Janssen *et al.* 2003

The scoring technique should be chosen to facilitate ease of evaluation for the experts. All are standardised to single dimensionless utilities before aggregation. In the case of the – – –/+++ scale it may be preferable to use a scale of one to seven, where a score of one indicates the most negative effect, four indicates no effect and seven indicates the most positive effect.

The scoring regimes are statements that define the impact that an option should achieve regarding a particular criterion to receive a particular score for that criterion. Depending on the data sources available for each criterion this will take various forms. For measured or modelled data Objective Value Functions are used to define the relationship between the raw data and the criteria score. These statements are of greatest importance when opinion is used to score the effects table. Typical statements would include a brief description of the criterion and the level of performance that would achieve each score. Below is a criterion from an analysis of landholder environmental monitoring indicators. The analysis assessed the usefulness of each indicator using expert opinion in this example an indicator that would score highly would be consistent with national and international monitoring goals.

Consistent with other regions, States/Territories and nations standards

(consistent methodology and units)

- 10 Indicator consistent internationally
- 9 Indicator consistent nationally, at State level, and by all regions
- 7-8 Indicator consistent at State level, and by other regions
- 5-6 Indicator consistent at regional level only, but consistent with other regions
- 3-4 Indicator consistent in one region only
- 1-2 Indicator consistent at local level only

## **4 The workshops**

NGT workshops are undertaken using a computer based group decision support system (GDSS). This system allows the experts to privately record their opinion regarding the performance of a policy measure against an indicator, the system then aggregates the opinion of all the experts which can then be presented to the whole group. The system highlights areas where there is a divergent opinion amongst the experts. These areas are then discussed, discussions start with experts stating their assumptions in making the assessment. If the assumptions differ a common set of assumptions are agreed upon. The differences in the evaluations are then discussed and differing opinions regarding impacts of the policy measures regarding the indicators are discussed. After the discussions have been completed a second private evaluation is undertaken. In this study two private evaluations and one discussion round are completed, and a level of greater 40% divergence in opinion was used to trigger further discussions. These levels were used in this example as a large number of evaluations were undertaken (550 evaluations undertaken over two days). As the number of evaluations are likely to be substantially fewer in this application (52 evaluations took place in the Welsh case study) a lower divergence may be used and an additional discussion and scoring round could be added.

It may be possible, but probably less effective, to organise the workshop as a teleconference. The teleconference would proceed in a similar way to the workshop. The spreadsheets used to create the shared software environment for the capture and presentation of opinion will be replaced by a website with similar functionality. The participants will enter their opinion privately on a section of the website, with the aggregated opinion presented on a separate section of the website. This will require the functionality of the spreadsheet ([Annex C4-5](#)) to be programmed in to the website. (In practice, the spreadsheet supplied for illustrative purposes would need to be reprogrammed to meet the specific needs of the evaluation to be conducted.)

### **4.1 Organisation of the expert panel workshops**

There are a number of considerations that need to be addressed in running the expert panel workshops. Specifically, the participants – who will comprise the panel, the workshop venue and computer facilities and the facilitation of the workshop and the driving the GDSS ([Annex C4-5](#)).

The expert panel needs to include persons with the skills and knowledge to evaluate all the criteria. For each criterion there should be at least one panellist with particular

expert knowledge (expertise level 5), and a number of others with knowledge and interest (expertise level 3 or 4).

Regarding the venue and computer facilities, the venue needs to be suitably large to accommodate the participants and the computer facilities. The seating needs to be arranged around a single table with good visibility to the projected data. Computer teaching rooms are rarely suitable as the machines tend to be situated around the edge of the room or in rows. The computer requirements include a networked PC for each panellist and two networked PCs for the facilitator and system operator. Ideally notebook computers should be used as normal monitors tend to be large and obscure visibility of the panellists.

The current version of the GDSS (presented below in section 6.3, with the spreadsheets presented in [Annex C4-5](#)) is not fully automated and requires a person other than the facilitator to operate it. The system operator needs to observe the progress of the panellists using one PC and present the aggregated panellist input on a second PC. This allows the system operator to observe the progress of the panellists without presenting this data to the whole group. Panellists are required to complete several tasks in each evaluation round, the process can not continue until all panellists have completed all the tasks. The system operator observing the panellists progress needs to inform the facilitator when panellists have failed to complete all their tasks. In this situation the facilitator needs to quietly prompt the panellist to ensure the tasks are complete. In addition it would be useful for notes to be taken of this workshop, notes need to be referenced to the option and criteria set being discussed and the evaluation round being completed. The option and criteria set can be defined prior to the workshop and a suitable note template can be produced.

Details of the Welsh case study are presented below to illustrate these points.

## **5 Aggregating the effects table**

Once the expert evaluations have been completed, the final stage is to resolve the effects table into an overall measure of performance. This process involves aggregating the evaluations presented in the effects table. This process can be split into two separate stages. First, relative weights are placed on the decision criteria by which the options are evaluated. This recognises that some of the criteria are more important to the stakeholders than others, and should attract greater weights in the analysis. Second, the scores that the options receive against each criterion judging their performance against that criterion (which form the effects table) are aggregated. That is, the scores are combined to create an overall score for each option. A wide variety of techniques are available for weighting and aggregating criteria (see Appendix 2).

The EU-CEE-OFP study employed the Rank Order Centroid (ROC) weighting technique and the weighted summation and Electr  aggregation techniques in combination. The ROC technique was chosen for this study because it is easy to elicit the rank orders from the stakeholders and this technique has a strong theoretical basis (Edwards and Barron 1994). The weighted summation technique provides a sound measure of the overall performance is logically simple and easily gains the acceptance of stakeholders. It is a compensatory aggregation technique. Compensatory techniques provide an overall measure of performance but tend towards the mean; that is, the result of using a weighted average tends to produce results that cluster around the mid-point of the score

range. High performance in the majority of criteria will compensate for poor performance in one or more criteria. An option that has high performance in all criteria may have the same aggregate performance as an option that has exceptionally high performance in a large number of criteria and exceptionally poor in some criteria.

Electré is a non-compensatory aggregation technique which checks for poor performance in a number of criteria, by using the discordance test. Poor performance in one or more criteria amongst high performance in the other criteria will be reflected in poor aggregate performance. An option with sound to high performance in all criteria will outperform the option with exceptionally high performance against the majority of criteria and exceptionally poor performance against a few criteria. Electré is a complicated technique and is not easily communicated to, or understood by, persons without an advanced understanding of mathematics. Consequently, the technique has not received wide application to natural resource management outside of theoretical studies.

When there is general agreement about ranking of options using the two aggregation techniques, no further investigation is required. If there is disagreement between the techniques and the performance of one or more options is relatively low using the Electré analysis compared to weighted summation, it can be concluded that these options have one or more poor performing criteria. These criteria can be identified by examining the effects table. The poor performance according to these criteria should then be judged as to importance in land-use decisions. If the criteria have been identified as highly important to the stakeholders then the options should be considered as undesirable options. If the criteria have been identified as having low importance to the stakeholders the options may still be considered as desirable.

As regards computer software, the package Definite (Janssen 1991 and Janssen *et al.* 2001) is a high-end multi-criteria analysis tool, modular in structure and with high functionality. The problem definition module includes the ability to construct criteria hierarchies and the effects table, as well as evaluate the options by using various methods, and can undertake correlation analyses on the criteria. The multi-criteria analysis module includes many aggregation and weighting techniques. Definite was chosen for this application because it includes all the multi-criteria analysis discussed, including the Electré and weighted summation aggregation techniques and the Rank Order Centroid weighting technique. Definite also includes functionality for cost-benefit analyses and sensitivity analyses.

### **The rank order weighting method**

Rank order weighting techniques are commonly used for assigning relative weights to a set of criteria. These weighting techniques assign weights to individual criteria according to a statement of importance or rank. These are usually in the form 'Criterion 1 is more important than Criterion 2, which is more important than Criterion 3, which is more important than Criterion 4', and so on. . Given these statements, a set of feasible weights can be defined, as well as the probability density function for all the feasible weights. The Rank Order Centroid (ROC) employed in this study uses the centroid value as the weight for the criteria.

### **The weighed summation method of aggregation**

This is the simplest and most commonly used aggregation technique. The value accorded to an option is the sum of the value the option receives for each criterion multiplied by the weight given to that criterion. The weights for the criteria are assigned

by the decision-maker or stakeholders to reflect the relative importance they place on the criteria. Weights may be directly assigned or assigned using a ranking process as applied in this study using the ROC technique. Whilst other MCA techniques have a stronger theoretical basis they are rarely used because they are complicated and time consuming.

### **The Electré method of aggregation**

The Electré (*E*limination et choix traduisant la *r*éalité or Elimination and choice corresponding to reality) method uses pairwise comparison of options with regards to the measured value of each criterion. Each pairwise comparison assesses the criteria using a concordance and a discordance measure. The concordance index is derived from the differences between all criteria for the two options. The discordance index reflects the maximum difference between the scores given to the same criterion in the two options. Threshold values are set for concordance and discordance. For example, when comparing option 1 with option 2 for a given criterion, option 1 is said outperform option 2 if the following statement is true:

*The difference between the value of option 1 compared to that of option 2 is greater than the concordance threshold and smaller than the discordance threshold.*

The concordance index measures the relative overall performance of the options, while the discordance index measures poor performance in individual criteria. The rationale of the discordance index is that overall good performance in most criteria cannot compensate for poor performance in an individual criterion.

Detailed descriptions of these techniques can be found in Appendix 6.

## **6 The Welsh Case study**

### **6.1 Indicators**

The following set of indicators is the revised set developed incorporating feedback from the Welsh case-study.

1. Capital investment on-farm
2. Diversification of farm enterprises
3. Diversification of rural economy
4. Fragmentation and other farm structure issues
5. Implementation costs (scheme)
6. Farm income
7. Uptake of regulated production systems
8. Biodiversity impacts
9. Control of climate change
10. Control of pollutants
11. Forestry
12. Landscape impacts
13. Natural resource conservation

14. Energy use
15. GM traceability
16. Animal welfare
17. Employment
18. Food quality and safety
19. Agricultural demographic
20. Public Health impacts
21. Occupational health
22. Knowledge and skills development
23. Rural community wellbeing
24. Social justice and equality (gender, intergenerational, international))
25. Rural infrastructure (including transport, housing)
26. Local consumption

## **6.2        *The process***

The process applied in the Welsh case study used a modified version of the NGT and a computer-based Group Decision Support System (GDSS). The modified version of the NGT allowed for private evaluations of the policy option followed by an open discussion of the evaluations. The GDSS was constructed commonly available academic or business computing hardware and software resources. Participating in the expert panel for the Welsh case-study were nine panellists, a facilitator, a recorder and system operator. The nine panellists provided the expert knowledge to evaluate the policies, the facilitator introduced the techniques to be used in the workshop, facilitated the discussions and the interactions with the GDSS, the recorder made notes on the discussions and the system operator ensured each panellist was using the system correctly recording when evaluations were complete and ensured the correct indicator was being presented in the data projector.

## **6.3        *The Group Decision Support System***

The application of NGT for evaluating the MCA effects table used a Group Decision Support System constructed from commonly available business computing hardware and software resources. Each of the 10 panellists and the system operator had a computer. Twelve desktops personal computers and a data projector were used, one for each panellist, one for the recorder and two for the system operator one connected to the data projector for presenting the summary of the opinion and one to monitor the process (a full specification of the computers used is presented in appendix 3). The system operators PC connected to the data projector was also used a server, all the spreadsheets used were saved on the hard drive of this PC.

The computer used by the system operator was connected to the data projector and all the computers were networked using a local area network. The computer used by the system operator was used as the server with all the spreadsheets used in the workshop stored on the hard drive of this computer.

Each panellist had a personal spreadsheet open on their computer in which they recorded their opinion as to the performance of the monitoring indicators and their self-assessed level of knowledge and their level of expertise (see appendix 4) in making these evaluations. An example of a panellist spreadsheet is displayed in Figure C4.3-1.

Microsoft Excel - Panelist01.xls

File Edit View Insert Format Tools Data Window Help

Type a question for help

Arial

12

**B**

</

**Figure C4.3-1**, Input spreadsheet for a panellist 1

The evaluations provided by the panellists in batches of three indicators were collated on the main spreadsheet controlled by the system operator. Figure C4.3-2 displays an example of the main spreadsheet. The process of collation was controlled by a simple macro in the main spreadsheet. The main spreadsheet contained references to all panellist spreadsheets. In Figure C4.3-2 the cell C3 is selected, this cell contains the reference to the evaluation provided by Panellist 1 as to the performance of the Tir Gofal scheme against the indicator “Capital investment on-farm”, the updating macro updates this and all the other links using the VBA (Visual Basic for Applications) function ‘UpdateLink’. To ensure the panellists’ spreadsheets contained the most recent assessments the panellists had to save their spreadsheets, this was required because the link in the mainsheet can only refer to the copy on the server not the active copy on the panellists PCs. A copy of the VBA code of the updating macro can be found below.



Microsoft Excel - main sheet 9 panellists.xls

File Edit View Insert Format Tools Data Window Help

Type a question for help

Arial 10 B I U

C3 =([Panellist01.xls]Sheet1!\$C\$2)

Main Sheet		Tir Gofal			Organic farming scheme		
		First assessment	Seocnd assessment	Third assessment	First assessment	Seocnd assessment	Third assessment
3	Capital investment on-farm	0	0	0	0	0	0
4		0	0	0	0	0	0
5		0	0	0	0	0	0
6		0	0	0	0	0	0
7		0	0	0	0	0	0
8		0	0	0	0	0	0
9		0	0	0	0	0	0
10		0	0	0	0	0	0
11		0	0	0	0	0	0
12		Maximum	0	0	0	0	0
13	Minimum	0	0	0	0	0	0
14	Average	0.00	0.00	0.00	0.00	0.00	0.00
15	Difference	0	0	0	0	0	0

Sheet1 / Sheet2 / Sheet3 /

Ready NUM

**Figure C4.3-2,** The Main Spreadsheet displaying the reference to the Panellist 1 spreadsheet

As the panellists made their evaluations they were requested to fill in their own personal spreadsheets. Figure C4.3-3 displays the first assessment of Panellists 1 as to the relative performance of the schemes against the capital investment, diversification and fragmentation indicators. In this example the Organic farming scheme receives a score of 7 thus indicating that it has a high level of performance against the capital investment indicator and receives a score of 1 against the fragmentation indicator thus indicating that it has a low level of performance against this indicator. When the panellist had completed their assessment they were requested to save their spreadsheets.

At this point the panellists are asked to consider their level of expertise in making these evaluations. The self-assessment of expertise is rated on a five-point scale. A score of one indicates that the expert is unfamiliar with the subject and a score of five indicates a high level of expertise.

Microsoft Excel - Panellist01.xls

File Edit View Insert Format Tools Data Window Help

Type a question for help

Arial 12 B I U

A1 =Panellist 1

Panelist 1		Tir Gofal	Organic farming scheme
Capital investment on-farm	First assessment	5	7
	Second assesment		
	Third assessment		
	Expertise	4	4
Diversification of farm practice and products	First assessment	4	4
	Second assesment		
	Third assessment		
	Expertise	4	3
Fragmentation and other farm structure issues	First assessment	5	1
	Second assesment		
	Third assessment		
	Expertise	3	2

Sheet1 / Sheet2 / Sheet3 /

Ready NUM

**Figure C4.3-3**, Input Spreadsheet for Panellist 1 displaying the first assessment and expertise evaluations

Once all the panellists had completed and saved their first assessments all the assessments were collated and presented back to the group for discussion using the data projector. The Main Sheet was then updated using the updating macro “Macro1”, this can be found in Main Sheet.xls (see Annex C4-5). The updating could also be completed using the “update now” function in Excel. Figure C4.3-4 show a section of Main Sheet.xls with the first collated assessments of the panellists shown in the yellow cells.

Main Sheet		Tir Gofal			Organic farming scheme		
		First assessment	Seocnd assessment	Third assessment	First assessment	Seocnd assessment	Third assessment
1		5	0	0	7	0	0
2		5	0	0	6	0	0
3		6	0	0	5	0	0
4		4	0	0	6	0	0
5		6	0	0	5	0	0
6		4	0	0	5	0	0
7		3	0	0	5	0	0
8		5	0	0	4	0	0
9		4	0	0	5	0	0
10		6	0	0	7	0	0
11		3	0	0	4	0	0
12	Maximum	6	0	0	7	0	0
13	Minimum	3	0	0	4	0	0
14	Average	4.67	0.00	0.00	5.33	0.00	0.00
15	Difference	3	0	0	3	0	0

**Figure C4.3-4**, The Main Sheet spreadsheet including the panellists’ first assessments

Owing to the nature of the spreadsheet (Main Sheet.xls) all of the evaluations provided by the panellists are displayed in order, i.e. the top value is that of panellist 1, then panellist 2 and so on. In the interest of keeping the evaluations anonymous the panellist numbering was not identified to the participants and they were placed in a random order around the room. Main Sheet.xls displays a summary of all the score, the maximum value, the minimum value, the average value and the difference. Indicators where there were more than two points of difference were automatically highlighted in orange. These were the indicators where there was the greatest divergence in opinion between the panellists and these were discussed first.

During the discussions the panellists were invited to present their rationales for their scores including any specific data to which they are privy and any assumptions they had made. Upon discover of varying assumptions the facilitator lead discussion upon which is what would be assumed in the particular example.

After the discussions were complete a second private evaluation was undertaken. Following the second assessment the panellists were requested to save their spreadsheets. Figure C4.3-5 displays a panellist’s spreadsheet with the second assessment completed.

Once all the second assessments have been completed and the panellists' spreadsheets saved the main sheet can then be updated. Figure C4.3-6 displays the second assessments collated in the mainsheet. The second assessment to the right of the first assessment in this example the opinions of the panellists have converged to within two points of difference for all criteria. In the cases where there was still convergent opinion, that is more than two points of difference in the second assessment further discussion would be required and a third assessment needs to be made.

Panelist 1		Tir Gofal	Organic farming scheme
Capital investment on-farm	First assessment	5	7
	Second assessment	5	6
	Third assessment		
	Expertise	4	4
Diversification of farm practice and products	First assessment	4	4
	Second assessment	4	4
	Third assessment		
	Expertise	4	3
Fragmentation and other farm structure issues	First assessment	5	1
	Second assessment	6	3
	Third assessment		
	Expertise	3	2

**Figure C4.3-5,** Input Spreadsheet for Panellist 1 displaying the first assessment, second assessment

Main Sheet		Tir Gofal	Organic farming scheme
		First assessment	Second assessment
Capital investment on-farm	First assessment	5	7
	Second assessment	5	6
	Third assessment	0	0
	Expertise	4	4
Diversification of farm practice and products	First assessment	4	4
	Second assessment	4	4
	Third assessment	0	0
	Expertise	4	3
Fragmentation and other farm structure issues	First assessment	5	1
	Second assessment	6	3
	Third assessment	0	0
	Expertise	3	2

**Figure C4.3-6,** The Main Sheet spreadsheet including the panellists' first and second assessments

In the example presented in figure C4.3-6 opinion regarding the Tir Gofal scheme is still at two points of difference, in the Welsh case-study it was decided that two points was a satisfactory level of agreement, considering there was limited time for further discussions. In a situation where more time is available it may be desirable to continue

the discussions until there is one or no points of difference. A trade-off needs to be made in each case study. Achieving consensus on all scores will be very time consuming but will increase the ease and robustness in analysing the data. Accepting a degree of divergence will reduce the length of the expert panel workshop but will increase the complexity of the analysis.

The evaluations provided in the second or third assessment were then used to assess the performance of the options. The evaluations in the second assessment and the evaluations of expertise elicited using these spreadsheets were passed to a separate spreadsheet for analysis. The analysis spreadsheet used risk analyses and compensatory and non-compensatory multi-criteria analyses to evaluate the performance of the options.

#### **6.4      *Feedback of the Welsh workshop***

The panellists in the Welsh study were asked to complete a feedback questionnaire. A summary of comments is presented below (Appendix 1). This section presents some of the preliminary comments.

Most respondents commented that time limited the discussion and this limited discussion and debate. One respondent also commented that the limited time encouraged compromise and increased interest in the workshop. It was suggested that the Welsh workshop should have been extended by 30 minutes from 3.5 hours to four hours. The Welsh workshop was limited to 3.5 hours so it could be completed before lunch.

Respondents commented that achieving consensus relied on gaining a common interpretation of the issues and reaching agreement on assumptions. It was noted that agreement was not always achieved but understanding of different opinions increased. In reviewing the discussion notes a common set of assumptions is achieved in all except the one indicator (the capital investment indicator). In cases where there are strongly held and divergent opinions both should be recorded and noted in the analysis.

Respondent commented that the assessments were made with incomplete knowledge and felt this resulted in an understatement of the impacts. It was also noted that experts made some evaluations on the basis of their best guess however they noted that in these cases they rated their expertise as low. These opinions with associated low rating expertise will be given little weight in the evaluation, but there were concerns in the feedback that this would not entirely remove this bias. It was commented that the indicators cover both direct and indirect effects of the schemes, these were defined as effects that the scheme was designed to address and those it was not. Knowledge of the indirect effects was less than knowledge on the direct effects as they are less likely to have formally assessed. Similarly this assessment aims to address both the on-farm and off-farm effects of the schemes. The knowledge of on-farm effects will be greater than that of the off-farm effects.

Several comments for the refining the indicator set were made in the feedback as well as in the discussions. These were included in the revised set of indicators attached. Other comments included ensuring that participants were happy to share their thought processes in reaching their evaluations. The possibility of completing the workshop using a web-based system and telephone conference was raised.

## 7 Summary

This paper presents a process for eliciting expert opinion using the nominal group technique and a computer-based decision support system. This system encourages debate regarding the performance of farming policy against a suite of agri-environmental and rural development indicators. By eliciting opinion privately the process seeks to avoid undue social pressure from other members of the group.

For future application the following points should be considered:

1. Evaluating two schemes against 25 indicators will take four hours, evaluating more schemes will increase the time proportionally.
2. All participants should be willing to share their thoughts in open discussions (in this respect this process is not strictly an NGT process).
3. Notes of discussion must be recorded, with special care taken to record assumptions made in undertaking the assessments.

## 8 References

- Armstrong, J., (1999), 'Introduction to paper and commentaries on the Delphi technique', *International Journal of Forecasting*, 15(4): 351-352
- Barron, F. and Barrett B. (1996), 'Decision quality using ranked attribute weights', *Management Science*, 42(11): 1515-1523.
- Dalkey, N. and Helmer, O., (1963), 'An Experimental Application of the Delphi Method to the Use of experts', *Management science*, 9(3): 458-467.
- Delbecq, A., Van de Ven, A. and Gustafson, D., (1975), *Group Techniques for Program Planning: a guide to the delphi and nominal group processes*, Scott, Foresman and Company, Glenview.
- Edwards W. and Barron F. (1994), 'SMARTS and SMARTER: improved simple methods for multiattribute utility measurement', *Organizational Behavior and Human Decision Processes*, 60(3): 306-325
- European Commission (1999), *Evaluating Socio-economic Programmes: MEANS Collection Vols. 1-6*, Office for Official Publications of the European Communities, Luxembourg.
- Goicoechea, A., Hansen, R. and Duckstein, L. (1982), *Multiobjective Decision Analysis with Engineering and Business Applications*, Wiley, New York.
- Janssen R. (1991), *Multiobjective Decision Support for Environmental Management*, Kluwer Academic Publishers, Dordrecht.
- Janssen R., van Herwijnen M. and Beinat E., (2003), *DEFINITE 3.0: case studies and user manual*, Institute for Environmental Studies, Vrije Universiteit, Amsterdam
- Keeney R. and Raffia H. (1976), *Decisions with Multiple Objectives*, Cambridge University Press, Cambridge.
- Linstone and Turoff (1975) *The Delphi method: techniques and applications*, Addison-Wesley, Reading.
- Liou, Y., (1998), 'Expert System Technology: Knowledge Acquisition', in *The Handbook of Applied Expert Systems*, Liebowitz, J. (ed), CRC Press, Boca Raton.
- Lootsma F. (2000), 'Distributed multi-criteria decision making and the role of the participants in the process', *Journal of Multi-Criteria Decision Analysis*, 9(1-3):45-55.
- Lovebridge, D., (2001), *Who is an expert?*, Ideas in Progress – Paper number 22, University of Manchester, Manchester, [http://les.man.ac.uk/PREST/Download/Denis/Who\\_is\\_an\\_Expert\\_wp22.PDF](http://les.man.ac.uk/PREST/Download/Denis/Who_is_an_Expert_wp22.PDF), accessed 6<sup>th</sup> Nov. 2003.
- Matsatsinis, F. and Samaras, A.. (2001), 'MCDA and preference disaggregation in group decision support systems', *European Journal of Operational Research*, 130(2): 414-429.

- Meyrick, J., (2001), *The Delphi Method and Health Research*, Macquarie Business Research Papers Number 11/2001, Macquarie University, Sydney.
- Rowe, G. and Wright, G., (1999), 'The Delphi technique as a forecasting tool: issues and analysis', *International Journal of Forecasting*, 15(4): 353-375
- Rowe, G., Wright, G. and Bolger F., (1991), 'Delphi: A Reevaluation of Research and Theory', *Technological Forecasting and Social Change*, 39(3): 235-251
- Roy B. (1991), 'The outranking approach and the foundations of ELECTRE methods', *Theory and Decision*, 31(1): 49-73
- Simpson, L. (1996), 'Do decision makers know what they prefer?: MAVT and ELECTRE II', *The Journal of the Operational Research Society*, 47(7): 919-929

## **Appendix 1: Feedback Questionnaire with results from the Welsh organic farming policy analysis workshop**

This section contains a summary of the scores and comments from the five panellists who completed the feedback questionnaire. The questionnaire included a number of questions where the panellist were asked to rate their experience. In these questions a score of five indicates a positive experience and one a negative experience. The average score is presented against these questions. Other questions were limited to comments only.

### **1. Were you happy with the workshop timing and duration (i.e. would you have preferred a longer workshop, or a number of shorter workshops, or...)?**

Average score: 4 – comments included:

- Happy with the timing from a personal point of view, but the subject matter would have benefited from greater debate
- Not enough time for discussions
- Push on time sometimes means people are more willing to compromise and keeps it interesting

### **2. Did you find the process useful for building consensus?**

Average score: 4 – comments included:

- Some of the questions had numerous possible interpretations so consensus depended on firstly accepting an interpretation.
- The process helped to refine definitions and interpretations. Participants will not always agree, but it did make a difference to be able to see where different participants were coming from. In some cases this changed scores when participants better understood what the issue was, or gained information from other participants.

### **3. Do you feel the process captured how well the policy options perform on ground?**

Average score: 3 – comments included:

- The policy options have an impact beyond those holdings in Tir Gofal or OFS. In assessing performance on the ground this should be reflected.
- Lack of knowledge / monitoring data probably resulted in understatement of impacts
- Depends on how much people know about schemes / rural development issues, where their information has come from, and how accurate that information is. There is an element of 'best guess' and wishful thinking, especially on topics on which participants don't have firm information on, so there will inevitably be an element of bias. This can be reduced by using the 'expertise' scores to weight results, but this will not entirely overcome bias.

### **4. Did you learn much from the workshop?**

Average score: 3.2

### **5. Do you feel your ideas were adequately incorporated in to the discussions?**

a. In general?

Average score: 4

b. When you assessed your level of knowledge and understanding as high?

Average score: 3.8

c. When you assessed your level of knowledge and understanding as low?

Average score: 3.2 – comments included:

- Limited time was a factor here
- Other not aware of one persons comments (as they were sited in the corner)

**6. Would you happily agree to undertake a similar process again?**

Average score: 4.4

**7. Did the workshop cover an adequate range of farming policy?**

Average score: 4.2 – comments included:

- The range of indicators seemed to adequately cover the range of policy objectives in the RDR. As mentioned elsewhere, it is important that anyone doing the analysis is aware of the context (ie. Looking at schemes from a perspective of both direct and indirect outcomes). The indirect effects (ie. those issues that the schemes are not specifically designed to deliver) will be more difficult to judge, and may not have been formally assessed – so information on them will be more speculative

**8. Were the indicators used adequate to evaluate the policy options?**

Average score: 3.4 – comments included:

- Needs more knowledge of purpose to judge! Needs more time than I have to properly assess this question.
- Well being of farmer family could be more specific target
- The indicators seemed to work well for the UK schemes – though some further refinement may be needed (as per discussion at the workshop) and consistency between topic, question and notes needs to be checked. At the EU-wide level some further checking may be required to see that the indicators are compatible.

**9. Are there other indicators that should be added (or removed) from the evaluation?**

- Lack of time
- Some re-wording needed

**10. How did you find the quality of the workshop facilitation?**

Average score: 3.6 – comments included:

- Fast way of making data; well done!
- Some issues usefully identified as pilot – need to keep process moving

**11. What do you see as the strengths of this process?**

- Rapid assessment, diverse respondents
- Sharing of expertise
- The strength was the ability to discuss differences and then to re-score, this definitely helped where the differences were as a result of different understanding of the issues. It worked well for a small group, but would be more difficult with a larger group.

**12. What do you see as the weaknesses of this process?**

- Technology took a long time
- Lack of time – another 30 minutes would have made a difference
- Time consuming (number of people \* hours)
- Not all issues can be resolved, so aiming for consensus on all issues will not be possible. In a very diverse group, especially where there are entrenched opinions, it may be impossible to get consensus. In itself, consensus perhaps should not always be the aim. Recording the range of



views and explaining opposing viewpoints, may be valid. Politically motivated opinions may be difficult to accommodate through the process.

**13. How could we improve that process in the future?**

- Ensure participants are willing to share thought processes on enrolment
- Web / phone conference option?
- Perhaps where no consensus is reached (ie differences in scores of 3 or more points), there should be a sub-process that explores and classifies the reasons why these differences occur. An example from the workshop was a difference in opinion on capital funding – at least one participant felt that Tir Gofal payments were largely for management options, and not for capital items. Others agreed that they were justifiably capital works. Being able to record this in the analysis might be helpful.

**14. Would you agree to your comments from the workshop and this questionnaire being used anonymously in future publications?**

- All respondents agreed

**15. Do you have any other comments?**

- It was interesting and enjoyable; on balance a worthwhile exercise; thanks

**Thank you, for attending the workshop and completing this questionnaire.**

## Appendix 2: Details of the weighting and aggregation techniques

This section presents the algorithms of the MCA weighting and aggregation techniques, namely the Rank Order Centroid Weighting technique, and the Electré and weighted Sum aggregation techniques.

### The Rank Order Centroid Weighting Technique

This weighting technique converts the rank order of criteria importance into a set of ordinal weights following a simple algorithm described by Edwards and Barron (1994). Given that criterion 1 is more important than criterion 2, and criterion 2 is more important than criterion 3 ( $Cr_1 \geq Cr_2 \geq Cr_3$ ), the weight assigned to each criterion would follow a similar rule  $w_1 \geq w_2 \geq w_3$ , where  $w_j$  is the possible weight applied to criterion  $j$ .

If  $Cr_1 \geq Cr_2 \geq \dots \geq Cr_j$   
then  $w_1 \geq w_2 \geq \dots \geq w_j$   
If  $k$  is the number of criteria  
then  $w_1 = (1 + 1/2 + 1/3 + \dots + 1/k)/k$   
 $w_2 = (0 + 1/2 + 1/3 + \dots + 1/k)/k$   
 $w_3 = (0 + 0 + 1/3 + \dots + 1/k)/k$   
:  
 $w_j = (0 + \dots + 0 + 1/k)/k$

More generally, the weight of criterion  $j$  is:

$$w_j = (1/k) \sum_{i=1}^j (1/i) \quad \text{Equation 1}$$

### The Weighted Summation Aggregation Technique

The weighted summation aggregation technique takes the form (Janssen 1991, Barron and Barrett 1996):

$$V(w, v) = \sum_j w_j v_j$$

where  $V$  = the weighted value or overall score for a given option

$w_j$  = the weight for a given criterion  $j$

$v_j$  = the value or score for a given option with respect to criterion  $j$

$$\text{Equation 2}$$

## The Electr  II Aggregation Technique

The first step in the Electr  II technique, as described by Goicoechea *et al.* (1982), Roy (1991) and Simpson (1996), is to establish a complete rank order of the options using a number of tests. The first three of these are the indifference test, the preference test and the veto test. These are evaluated between two options for each of the criteria. These tests are now outlined.

The indifference test is designed for testing that Option A is indifferent to another Option B (notation *AIB*) for a given criterion. This allows for options with a small variation in score to be considered equal. In Electr  II the decision-maker sets the indifference measure  $q$  for each criteria.

$$AIB \text{ if } |a_j - b_j| \leq q_j$$

Equation 3

where  $a_j$  is the performance (the score it receives in the effects table) of option A on criterion j,  $b_j$  is the performance of Option B on criterion j, and  $q_j$  the indifference measure. The indifference measure will vary depending on the score range of the criteria.

The preference test evaluates whether an Option A is preferred to another Option B (notation *APB*), that is Option A outperforms or outranks Option B. This determines a strong or strict preference for Option A; that is, the score received by option A for criterion j is considerably better than that for option B. The preference threshold  $p_j$  is the level at which the decision-maker considers Option A to strongly outperform Option B.

$$APB \text{ if } a_j > b_j + p_j$$

Equation 4

where  $a_j$  is the performance of option A on criterion j,  $b_j$  is the performance of option B on criterion j, and  $p_j$  is the strong preference threshold. This test is applied to all the criteria and the results from all the criteria are considered further in the analysis.

The veto test evaluates whether an Option A is not preferred to another Option B; that is, the score received by option A for criterion j is considerably worse than that for option B (notation *BPA*). The Veto level  $v_j$  is the level at which Option A is not preferred to Option B, for any criterion. If the option A performs at a level that is lower than the performance of option B by a greater than the Veto level for any criterion, Option A is always not preferred to Option B, regardless of the scores in the other criteria.

$$BPA \text{ if } a_j + v_j < b_j$$

Equation 5

where  $a_j$  is the performance of option A on criterion j,  $b_j$  is the performance of option B on criterion j, and  $v_j$  is the veto measure. This test is applied to all the criteria. Failing this test in one criterion will result in Option A always being considered not preferred to Option B.

The performance of a set of two options for each criterion can be assigned. This can be categorised into four groups. The first is indifferent *I* group, the comparisons that fulfil equation 3, the second the strongly preferred *P* group, the comparisons that fulfil equation 4 or equation 5. The third group *Q* is a group of weak preference all those that

fall between the indifferent set and the strong preference set. A fourth group **R** is created from the residual – those that failed all of the above tests – and these are deemed to be incomparable.

Using these threshold tests and the subsequent group Electr  II technique builds an outranking relationship  $S_j$  for each criterion. This group (notation  $aS_jb$ ) is the set for each criterion where option A is at least as good as Option B. Therefore for criterion j:

$$aS_jb: A \text{ is at least as good as } B \text{ on criterion } j \text{ if } a_j \leq b_j - q_j$$

$$\{aS_jb\} = \{aP_jb\} \cup \{aQ_jb\} \cup \{aI_jb\}$$

Equation 6

Using the groups *I*, *P*, and *Q*, Electr  II specifies values known as the concordance and discordance coefficients. The concordance coefficient is a measure of the strength of the argument *A is at least as good as B*,  $c(A,B)$ . The discordance coefficient is the strength of the counter argument; that is *A is not as good as B*,  $d(A,B)$ . In many examples the criteria do not carry equal weight. Weights are applied to each criterion using the ROC technique. These weights are included in defining the Concordance and Discordance indices.

$$c(A,B) = \sum P(w_j) / \sum (w_j)$$

Using ROC weighting technique  $\sum (w_j) = 1$ , so, in this example

$$c(A,B) = \sum P(w_j)$$

Equation 7

The next stage of the Electr  II technique requires the decision-maker to set concordance and discordance thresholds. The level of the concordance threshold  $c^*$  and the discordance threshold  $d^*$  are set according to the aims of the analysis. High levels are likely to filter out all but the better performing options, while lower levels will identify acceptable and unacceptable groups of options. That is, when the concordance threshold  $c^*$  and the discordance threshold  $d^*$  are set to a high level, relative to the range of score received by the options, the most preferred options will have substantially higher performance in most criteria and no critically poor performing criteria.

The following decision rules are applied to the pairwise comparison of all the options against each other.

$$\begin{array}{lll} \text{A outranks B if} & c(A,B) \geq c^* & \text{and} \\ & d(A,B) \leq d^* & \end{array}$$

$$\begin{array}{lll} \text{A outranks B if} & c(A,B) > c^* & \text{and} \\ & c(B,A) \geq c^* & \text{and} \\ & d(B,A) \leq d^* & \end{array}$$

A and B are incomparable if

$$\begin{array}{l} c(A,B) \geq c^* \text{ and } d(A,B) > d^* \text{ or} \\ c(A,B) > c^* \text{ and } c(B,A) < c^* \text{ or} \\ c(A,B) \leq c^* \text{ and } c(B,A) < c^* \text{ and } d(B,A) < d^* \end{array}$$

Statements of outranking can be drawn from the outcomes of the decision rules. The statements of outranking can then be interpreted as an ordinal ranking of the options. Options that are incomparable or outrank the same group of options should be given equal ranking. These ranks can be displayed as lists of position or as bar charts.

## Appendix 3: Computer specifications

The hardware resources used in this study were:

- Twelve desktop computers of various specifications (minimum of an Intel Celeron processor, 500MHz and 256 MB of RAM) with networking hardware installed
- A Local Area Network
- A Data Projector

The software resources used were:

- Microsoft Windows XP operating system
- Microsoft Excel 2003 spreadsheet
- Networking capacity using Microsoft Windows networking

### ***Update macro***

This section contains the Visual Basic for Applications code of the macro used to update the links in the NGT GDSS. This imports all the expert evaluations into the main sheet for presentation to the panel.

```
Sub Macro1()
```

```
' Macro1 Macro
```

```
' Macro recorded 02/09/2005 by David Ian Jeffreys
```

```
' Keyboard Shortcut: Ctrl+q
```

```
Windows("main sheet 9 panellists.xls").Activate
```

```
FilePath = ActiveWorkbook.Path & "\"
```

```
ActiveWorkbook.UpdateLink Name:=FilePath & "Panellist01.xls",_
```

```
Type:=xlExcelLinks
```

```
ActiveWorkbook.UpdateLink Name:=FilePath & "Panellist02.xls",_
```

```
Type:=xlExcelLinks
```

```
ActiveWorkbook.UpdateLink Name:=FilePath & "Panellist03.xls",_
```

```
Type:=xlExcelLinks
```

```
ActiveWorkbook.UpdateLink Name:=FilePath & "Panellist04.xls",_
```

```
Type:=xlExcelLinks
```

```
ActiveWorkbook.UpdateLink Name:=FilePath & "Panellist05.xls",_
```

```
Type:=xlExcelLinks
```

```
ActiveWorkbook.UpdateLink Name:=FilePath & "Panellist06.xls",_
```

```
Type:=xlExcelLinks
```

ActiveWorkbook.UpdateLink Name:=FilePath & "Panellist07.xls",\_

Type:=xlExcelLinks

ActiveWorkbook.UpdateLink Name:=FilePath & "Panellist08.xls",\_

Type:=xlExcelLinks

ActiveWorkbook.UpdateLink Name:=FilePath & "Panellist09.xls",\_

Type:=xlExcelLinks

End Sub

## Appendix 4: Expertise levels

The workshop panellists were asked to provide a self-assessment of their knowledge and experience in evaluating each indicator against each of the decision criteria according to the following ratings. These ratings have been adapted from those developed by Lovebridge (2001).

1. **Unfamiliar** with the topic.
2. **Casually acquainted**; you have read or heard about the topic in the popular media or other popular presentations.
3. **Familiar** with the topic; you know most of the arguments advanced for and against some of the issues surrounding it and you have read about it and have formed some opinions about the topic.
4. **Knowledgeable** in the topic
  - (a) if you understand this topic and use this knowledge in land-use management;
  - (b) if you are in the process of becoming an expert, but still have some way to go to achieve mastery of the topic;
  - (c) if you work in a neighbouring field and occasionally draw upon or contribute to the development of this topic; or
  - (d) if you were an expert in it some time ago but feel somewhat rusty now.
5. **Expert** if you consider yourself to belong to that community of people who currently dedicate themselves to the topic matter and if you are in the technical field you are likely to have presented, written up and/or published the outcomes of your work.