Multi-criteria analysis
for the impact assessment of food safety policies:
The case of EU regulation on dietary arsenic

Ragona M.[1], Mazzocchi M.[1] and Alldrick A.J.[2]

[1] Department of Statistics, University of Bologna, Bologna, Italy
[2] Campden BRI, Chipping Campden, UK

maddalena.ragona@unibo.it

Paper prepared for presentation at the 1st AIEAA Conference
'Towards a Sustainable Bio-economy: Economic Issues and Policy Challenges'

4-5 June, 2012
Trento, Italy

Summary

Developments in knowledge concerning the toxicology and occurrence of dietary arsenic suggest that levels of exposure in some groups of the population within the EU are a cause for concern. This applies also in the case of individual Member States where local regulatory limits exist. The situation is such that some foods on the market are already the subject of consumer advice provided by government agencies. In the light of these considerations, some have suggested that Member States' legislation concerning arsenic in foods should be modified and harmonised to reflect such developments.

An evaluation of alternative policy initiatives is considered in this work. It employed a computer-based, fuzzy multi-criteria impact assessment tool for the identification of the preferable policy option. Such a tool, named 'Scryer', includes a rigorously structured qualitative assessment of each type of impact (e.g. public health, costs for businesses, costs for public authorities) for each policy option, a feasibility filter which considers the opportunity to undertake a quantitative estimation for any type of impact, and the comparison of policy options through a fuzzy multi-criteria approach. The transparency of the tool allows also for a weighting of the impacts.

Three policy options concerning regulation of arsenic were evaluated: (1) the status quo option, reflecting the current situation, where levels of arsenic in drinking and mineral waters are governed by EU-wide legislation and that of foods is determined at a Member State level; (2) statutory controls, resulting in the introduction of maximum residue limits for arsenic in foods; and (3) voluntary standards, where the adoption of a policy of self-regulation is expected to reduce levels of inorganic arsenic in the food supply.

Application of the Scryer tool suggests that the preferable policy option would be to replace the status quo with legally enforceable (lower) limits at the EU-wide level, and that voluntary limits would be the least risky choice.

Keywords: multi-criteria analysis; fuzzy; regulatory impact assessment; food safety regulations; arsenic

JEL Classification codes: C02; C65; D81; Q18

Multi-criteria analysis for the impact assessment of food safety policies: The case of EU regulation on dietary arsenic

Ragona M.[1], Mazzocchi M.[1] and Alldrick A.J.[2]

[1] Department of Statistics, University of Bologna, Bologna, Italy
[2] Campden BRI, Chipping Campden, UK

1. Setting the scene: regulatory impact assessment in food safety

The need to improve analysis of the expected impact of policies aimed at responding to market failures in the bio-economy is crucial nowadays. The standard procedure usually employed in regulatory impact assessment (RIA) – i.e. cost-benefit analysis (CBA) – faces major obstacles, especially for food safety policy interventions. The major limitation is that precise information and data are rarely available and of good quality, especially for some key impacts, like the effects on human health. Other obstacles for a proper CBA include (a) the probabilistic nature of some actions, as food hazards may still occur with lower risks; (b) the difficulty in isolating confounding factors (e.g. market forces, weather conditions, etc.); (c) uncertainty in compliance levels; (d) different timing in the occurrence and discounting of costs and benefits (e.g. short-term costs for SMEs vs. long-term health outcomes). Finally, a key point usually overlooked is the relevance of public perceptions and media coverage on foodborne risks for public decision making, especially when policy intervention is sparkled by a major food outbreak.

As Sunstein (2001) pointed out, even with a large amount of scientific data available as support for policy decision making, CBA might only identify a range – often wide – of both benefits and costs. He brings the example of a regulatory proposal on arsenic in drinking water expecting to save as few as 0 lives and as many as 112. When monetization of benefits is made, the range goes from $0 to $560 million. Such figures cannot inform properly decision makers, but can only bring a sense of the potential consequences of various courses of actions. This is the case of many policies targeting at improving consumer health benefits, including food safety regulations.

An alternative to CBA is multi-criteria analysis (MCA), where the ranking of policy options can be based on different measurement scales (including the combination of quantitative and qualitative variables). The extension of MCA to consider fuzzy measurement allows to accompany discrete qualitative impact evaluations with an indication of uncertainty.

The advantage of MCA (over CBA) to consider different criteria which can be estimated in different ways fits well to agricultural, food and environment policy areas, due to their multi-dimensional nature. Therefore, MCA has recently received attention as a support tool to decision making, both at the private (firm or industry) and public (government) levels (see Bartolini and Viaggi, 2010, for a review). For example, it has been used in farm modelling to assess the optimum level of factor use by farmers in order to account for a broader set of objectives than the mere profit maximisation (Gómez-Limón, 2004), or as a support to decisions in lowland irrigated agriculture in order to take into account both economic and sustainability criteria (Tiwari et al., 1999).

In this paper, we explore how fuzzy multi-criteria analysis (FMCA) could address the obstacles associated with the impact assessment of food policies, with a sample application on policy initiatives to address the problem of arsenic in food. This application is still a 'work in progress', so no reliable conclusions should be drawn for real policy making, although analytical consideration can be regarded as realistic. Arsenic is a very global health problem affecting many millions of people. It has both acute and chronic toxicity and exposure mainly occurs through contaminated water, but also through food. On 14 March 2012 Italy notified a level of arsenic exceeding the permitted levels in mineral water and withdrew some product – that was being exported to Portugal – from the market through the Rapid Alert System for Food and Feed (RASFF)[1].

The paper is structured as follows. In the next section the various steps of the fuzzy MCA procedure will be described (with details shown in the Appendix). Section 3 will present the application of the tool on alternative policy options addressing the contamination of arsenic in food and water. The final Section will draw some methodological and applicative consideration.

2. METHODOLOGY

To meet the requirements of the EC-IAGs within the food safety realm, we developed a framework to account for the following aspects: (a) identification of 14 potential impacts, with different levels of detail and aggregation, and a breakdown into sub-impacts where needed; (b) explicit consideration of uncertainty in the qualitative assessment of these impacts; (c) inclusion of an element of 'proportionate level of analysis' to identify those impacts for which quantification (and possibly monetisation) is needed, feasible and affordable; and (d) the possibility of evaluating simultaneously qualitative and quantitative assessments.

The procedure we propose for the ex-ante assessment of food safety regulations consists of three steps. First, all impacts are evaluated on a qualitative scale, based on a clearly defined scoring system. Second, a quantitative evaluation may be entered for some or all of the impacts. The selection of quantified impacts is supported by a feasibility filter which considers costs and constraints of measurement and modelling. Finally, policy options are compared on a pairwise basis and ranked through a fuzzy multi-criteria analysis (FMCA), considering uncertainties in qualitative assessment and estimation error for quantitative evaluations.

In MCA, different 'actions' (policy options, in our case) are assessed regarding a set of criteria (impact categories, in our case) to finally produce a ranking of such actions. Each criterion can be assessed either in qualitative or quantitative terms (but the same criterion needs to be assessed in the same way across the various actions).

Within the fuzzy approach of MCA (see e.g. Meyer et al., 2005, and references therein), it is possible to accompany discrete qualitative impact evaluations with an indication of uncertainty. Thus, when comparing two policy options A vs. B, fuzzy methods allow for all outcomes (e.g. 'A is better than B', 'B is better than A', 'A and B are similar'), although with different grades of 'fuzziness' (e.g. 'A is better than B' is much more credible than 'B is better than A'). To make a rough parallel with stochastic quantification, fuzzy methods take into account the precision (confidence intervals) of policy impact estimates, then

[1] http://ec.europa.eu/food/food/rapidalert/rasff_portal_database_en.htm

associate the dominance of a policy option with a credibility measure which can be paralleled with a probability level. Since the measurement of qualitative variables is 'linguistic' and not probabilistic, fuzzy methods develop the required indicators of credibility, which result as a combination of (1) the degree of dominance of a policy option versus another (pairwise methods) for each individual impact; (2) the degree of uncertainty (fuzziness) in the linguistic measurement. Individual impact assessments and pairwise comparisons are then aggregated to produce a ranking of policy options.

In order to deal with the specificities of food safety policies, we selected and adapted the impacts listed in the EC-IAGs (EC, 2009, pp. 32-37). The impact list was integrated with assessment of societal concerns (public opinion and media, consequences on vulnerable groups, etc.), so that a decision maker may overtly incorporate such considerations in a transparent and rigorous way.

2.1. Qualitative assessment

The first step requires to score each of the 14 impacts for each policy option, based on a qualitative scale. This set of qualitative indicators is obtained indirectly by combining four components of an impact : the direction of the policy impact (whether positive, neutral or negative), its severity (a measure of intensity of the effect on those that are affected the policy), its scale (which proportion of the reference population would be actually exposed to the policy effects), and its likelihood (the probability that the impact actually occurs).

For example, if one considers the (positive) public health effects of a food safety regulation, severity may refer to the type of the expected health outcome (mortality, morbidity or well-being), the scale would measure which percentage of people is exposed to the health risk covered by the regulation, and the likelihood is a measure of the probability that the regulation actually achieves the expected health outcomes, as there may be external uncontrollable factors that affect the outcome with a given level of probability (e.g. specific weather conditions may alter the actual health risk for the exposed population). The scoring procedure converts qualitative statements on the above four components into an ordinal variable (X) which ranges between 1 (strong negative impact) and 9 (strong positive impact), where 5 is neutrality.

Beyond the uncertainty due to external or uncontrollable factors (which is captured by the likelihood component), there is another major element of uncertainty associated with the potential lack of objective information, and the 'internal' uncertainty, i.e. the degree of confidence that the assessors place into their statements. This is captured by another ordinal variable (*u*) which ranges between 1 (maximum uncertainty) and 5 (minimum uncertainty).

The final output of the scoring step is a qualitative impact matrix (*I*) of dimension $14\times(2n)$, where each row is an impact, n is the number of policy options and for each policy option a numerical value for x (the qualitative score) and a value for u (uncertainty in the assessment) are recorded. Formally, the matrix I can be defined as:

$$\mathbf{I} = [\mathbf{X}|\mathbf{U}] \tag{1}$$

Where X is a $14\times n$ matrix whose elements x_{ij} are the ordinal values (between 1 and 9) which measure the qualitative impact of policy j (j: $1,\dots,n$) for the impact category i (i: $1,\dots,14$) and U is also a $14\times n$ matrix whose elements u_{ij} are the corresponding uncertainty assessments (with values between 1 and 5).

2.2. Feasibility filter and quantitative assessment

Once the preliminary qualitative assessment of the policy options for all 14 impacts is performed, a feasibility filter may be used in order to determine the opportunity for a quantitative assessment for any of the 14 categories of impact, by considering:

- the availability of primary and secondary data needed for quantitative evaluation,
- the costs for collecting quantitative necessary to perform the quantitative assessment, in comparison to available resources, and
- the presence or absence of any time constraints to perform a quantitative assessment.

Similarly to the previous qualitative assessment, the filter is rigorously structured. A numerical value is associated to each possible alternative, and from the combination of the numerical values and the previous qualitative assessment (the *x* and *u* of the qualitative impact matrix), a suggestion is given on the opportunity to proceed or not with a quantitative assessment, or to make an ad hoc decision. The details of the scoring procedure are in Appendix A.1. However, a different choice can be made, regardless of the suggestion given by the tool.

For the categories of impacts selected for a quantitative assessment is decided for, the numerical value of the expected quantitative outcome and a variability measure of the quantitative indicator - for each policy option – are requested. The variability measure is considered as the natural expression of quantitative uncertainty, and can be indicated by the standard error of estimates – also in consideration of the fact that assessments are generally model-based.

The measurement scale must be the same for all policy options of a given individual impact (i.e. measurement scales may vary across impacts, but not across policy options for the same impact).

For the categories of impacts estimated quantitatively, the quantitative values will replace the qualitative indicators in the fuzzy calculations of the following step: the comparison of policy options.

2.3. Comparison of policy options

For the final step of policy comparison, we decided to draw on a routine developed by Munda (see e.g. Munda et al., 1995) named NAIADE. The adaptation of NAIADE to the needs and characteristics of food safety policy evaluations required to address some drawbacks and complexities, mainly due to the fact that the procedure was developed for more general purposes and especially for environmental assessment. Here we describe the steps to arrive to the final output, whilst the mathematical details can be found in Appendix A.3:

Qualitative variables are firstly transformed into Gaussian fuzzy sets, based also on the uncertainty indicator. A fuzzy set is fully defined through a membership function, representing the degree of belonging of an element to each of the elements of the fuzzy set. Such a membership degree may range from 0 (no membership) to 1 (full membership). Membership values could be paralleled with probabilities, although they are conceptually different. In the present case, the elements of a fuzzy set are the discrete (ordinal) values between 1 and 9 (representing all possible values that a qualitative assessment may take), so a degree of membership is needed for each of these nine values. As an example, imagine that a qualitative judgment on the public health effects of relaxing the standards on a food contaminant is produced, and such assessment is towards a negative outcome. However, uncertainties makes it difficult to make a sharp qualitative judgement (e.g. that the impact is 3 rather than 2 or 4 on a 9-point scale). Thus, it is preferable to spread such evaluation across all potential values, but with different degree of membership, so that the highest degree of membership might indeed be associated with the qualitative value 3, lower grades of memberships will correspond to 2 and 4 and so forth. The furthest values will show the lowest grades of membership, and possibly zero if such values are ruled out even in presence of uncertainty, for example when the qualitative judgement and the level of uncertainty rule out a positive public health outcome of relaxing food standards.

For impacts where a quantitative assessment is performed, a standardized normal Gaussian distribution is applied.

Then, pairwise distances between fuzzy sets are computed. The distance is calculated for each pair of policy option, for each impact. For quantitative variables, the Hellinger distance is exploited.

Credibility values are generated for a set of 6 preference relations ('much better', 'better', 'approximately equal', 'indifferent', 'worse', and 'much worse') between each pair of policy options and for each criterion, based on the distance and some fixed thresholds.

Credibility values are then aggregated across criteria for each pair of policy options. Through an outranking approach, the outcome is a preference intensity index for each pair of policy options. At this stage, different weights to the various criteria can be applied. Weights can be assigned explicitly by rating the relative importance of each impact.

Finally, two indexes for each policy option are computed. The 2 indexes ('best option' and 'worst option') can be interpreted as the degree of membership to the statements 'this policy option is the best one' and 'this policy option is the worst one', respectively. At this stage, an implicit weighting measure can be introduced, the so-called entropy, which reflects different degrees of consistency among the credibility values of each pair of policy options. Therefore, the final 2 ranking indices can be expressed (a) without any weights, (b) with entropy weights, (c) with explicit weights, and (d) with entropy and explicit weights.

3. RESULTS FROM THE APPLICATION

The occurrence and carcinogenic significance of arsenic has been discussed by IARC (2004). Arsenic is the 20th most common element in the earth's crust; it rarely occurs in the elemental form but rather as tri- or penta- valent salts. Common trivalent forms include arsenic trioxide and sodium arsenite while pentavalent forms include arsenic pentoxide and other arsenates. Although well known for its acute toxicity, in terms of general food-safety arsenic is also significant as a chronic toxicant. Chronic toxicity is of particular concern since long-term exposure to inorganic arsenic compounds (in particular) is associated with the increased incidence of certain human cancers (e.g. of the skin, bladder, and liver). Inorganic arsenic compounds have also been implicated in the incidence of diabetes, cardiovascular disease as well as dermatological and neurological diseases. In terms of food, arsenic's significance lies in the observation that while exposure to arsenic containing compounds can be the result of human activities (e.g. smelting, cigarette smoking and burning of fossil fuels); the major source of exposure is through the diet in terms of both the food and water consumed.

Dietary arsenic presents itself either as inorganic (discussed above) or organic compounds (e.g. arsenobetaine, arseno sugars, dimethylarsinate and dimethylarsonate). In terms of the known toxicity of arsenic and its compounds, while cause and effect has been demonstrated with the inorganic species; the organic forms appear either to present no toxicological concern or have not been sufficiently investigated to determine their true toxicological significance.

The risk to human health of dietary arsenic has recently been assessed by the European Food Safety Authority (2009). In light of the known hazards presented by inorganic arsenic compounds the Authority's assessment focussed on inorganic arsenic and its ability to induce human cancers. It was partly based on over 100,000 occurrence data from 15 European countries. Most of the analytical data provided was for total arsenic and consequently the Authority had to make certain assumptions as to the proportions of inorganic arsenic present.

Occurrence information only provides an insight to the amounts of arsenic present in food and forms the basis of estimating the actual individual contribution to the total amount of arsenic consumed and hence dietary exposure. Dietary exposure is expressed in a terms of a body burden (µg arsenic per kg body weight). In terms of inorganic arsenic the body burden depends on the proportion of total arsenic present in the appropriate molecular form as well as the amount of food actually eaten.

These data were used by EFSA to estimate inorganic arsenic intakes by the European population as a whole. The major dietary source of inorganic arsenic was considered to be cereal-based foods and in particular those foods based on imported rice. Estimates of inorganic arsenic consumption by adults eating an omnivorous diet were estimated to range from 0.13-0.56 µg/kg body weight per day for average consumers and 0.377-1.22 µg/kg body weight per day for high consumers. These estimates increased when considering high rice-consuming adults (0.95 µg/kg body weight per day). Infants (6 months) and young children (< 3years old) were also identified as being of potential interest. Average inorganic arsenic intakes for breast-fed and infant formula fed infants were 0.04 and 0.116 µg/kg body weight per day respectively; while for those fed rice based food the average value was estimated to range from 1.63-1.76 µg/kg body weight per day. In the case of children aged under 3 years, values ranged from 0.74-1.39 (average consumers) to 1.47-2.66 (high consumers).

When considering the food-safety implications of such data, they have to be compared with an index of safe exposure. JECFA had proposed a provisional tolerable weekly intake of 15 µg/kg body weight per week. The European Food Safety Authority considered that this value was no longer appropriate, bearing in mind subsequent evidence concerning the toxicity of inorganic arsenic. They adopted a range of benchmark dose lower confidence limit ($BMDL_{01}$) values of between 0.3 and 8 µg/kg body weight per day depending of the end-point of toxicological concern. From the above it can be seen that average dietary intakes of inorganic arsenic are close to or within the $BMDL_{01}$ values. The Authority therefore concluded that there was little or no margin of exposure and that the possibility of a risk to some consumers could not be excluded.

3.1. Regulatory background and identification of policy options

In terms of inorganic arsenic, Community-wide legislation exists for maximum residues in drinking (tap)[2] and mineral waters[3]. These are expressed as total arsenic and are 10 µg per litre for both types of water. No similar Community-wide regulation exists for arsenic in foods. Within the European Union arsenic contamination of foods is therefore addressed through the provisions of the general food contaminants regulation[4] and those of individual Member States. For example, within the United Kingdom control of the arsenic content of foods is effected through Arsenic in Food Regulations[5], which, with some exceptions, sets a maximum limit for total arsenic in foods of 1mg kg^{-1}.

The growth in knowledge concerning the toxicity of arsenic suggests that levels of exposure in some groups of the population within the EU are a cause for concern. This is even the case in individual Member States where local regulatory limits exist. The situation is such that some foods on the market are already the subject of consumer advice provided by government agencies. One such example concerns 'rice-milk' a product sometimes fed to infants as a cow's milk replacement. Research commissioned by the UK Food

[2] Council Directive 98/83/EC of 3 November 1998 on the quality of water intended for human consumption

[3] Commission Directive 2003/40/EC of 16 May 2003 establishing the list, concentration limits and labelling requirements for the constituents of natural mineral waters and the conditions for using ozone-enriched air for the treatment of natural waters and spring waters.

[4] Council regulation (EEC) No. 315/93 of 8th February laying down Community procedures for contaminants in food

[5] Minister of Agriculture Fisheries and Food (1959) The Arsenic in Foods Order *Statutory Instrument 1959 no. 831*

Standards Agency surveyed the amount of arsenic present in such products. Its findings were that, when found, arsenic levels in the products complied with the UK legal maximum. Nevertheless at the levels found, exposure assessments demonstrated that young children consuming rice milk were at risk of elevated levels of dietary arsenic exposure. As a consequence, the Agency recommended that children should not be fed rice milk. Current legislation at a Member State level therefore does not therefore appear to reflect the risk of the hazard presented by dietary arsenic.

Three policy options concerning regulation of arsenic were evaluated:

1. 'status quo' option, reflecting the current situation, where levels of arsenic in drinking and mineral waters are governed by EU-wide legislation and that of foods is determined at a Member State level
2. 'statutory controls', resulting in the introduction of maximum residue limits for arsenic in foods, and
3. 'voluntary standards', where the adoption of a policy of self-regulation is expected to reduce levels of arsenic in the food supply.

3.2. Impact analysis

The results of qualitative assessments using the Scryer tool for each policy option are presented in Table 1. As already pointed out in the introductory section, we are still revising the qualitative impact assessment for this application, so it will be used to emphasise the main potentialities of Scryer, but cannot be regarded as informative for real policy considerations. The assessments are based on a scoring system for each impact (x and u) for each policy option. In the light of current knowledge the status quo was considered to have the potential for contributing additional costs in the future, consequently the impact was not necessarily considered to be neutral (score = 5). A justification for the scores is presented on a criterion by criterion basis below.

Table 1. Impact assessment of policy options on arsenic.

Impact categories	*Actual assessment*	*Impact relevance*	*Explicit weight*	*Policy option 1 Status quo*		*Policy option 2 Regulatory limits*		*Policy option 3 - Voluntary limits*	
				x	*u*	*x*	*u*	*x*	*u*
Public health	Qualitative	10	0.17	2.00	4.00	8.00	3.33	7.33	3.33
Firm competition	Qualitative	1	0.02	5.00	3.20	3.80	2.80	4.60	2.40
Conduct of businesses/SMEs	Qualitative	5	0.09	4.23	3.54	3.31	3.38	3.15	3.23
Administrative burdens on businesses	Qualitative	5	0.09	5.00	5.00	1.00	5.00	2.00	5.00
Public authorities	Qualitative	3	0.05	1.00	4.00	1.00	4.00	1.00	4.00
Innovation and research	Qualitative	2	0.03	7.33	4.00	8.33	4.00	7.67	4.00
Consumers	Qualitative	5	0.09	5.00	4.25	4.50	3.00	4.50	2.75
International trade	Qualitative	3	0.05	5.00	2.60	3.00	3.20	3.60	3.20
Macroeconomic environment	Qualitative	1	0.02	3.50	2.00	6.50	2.00	5.00	2.00
Labour markets	Qualitative	5	0.09	5.00	3.00	5.00	3.00	5.00	3.00
Environment	Qualitative	3	0.05	5.00	3.00	5.00	3.00	5.00	3.00
Distributive effects - negative	Qualitative	5	0.09	2.63	3.00	3.25	3.00	3.75	3.00
Distributive effects - positive	Qualitative	5	0.09	5.00	3.00	8.40	3.00	6.00	3.00
Societal concerns	Qualitative	5	0.09	3.00	3.00	7.00	3.00	7.00	3.00

Source: own elaboration

Public Health: The European Food Safety Authority has already determined that margins of exposure to arsenic in relation to indices of safe exposure are low or non-existent (EFSA, 2009). These are considered reasonable grounds for intervention. Given the precedent with other toxic metals (e.g. cadmium, lead and mercury) and water regulations, a legislative as opposed to a voluntary approach is considered to have a more positive impact on public health, as it is expected that the number of firms adhering to a voluntary standard will not reach the totality of firms that should comply with a mandatory regulation.

Firm competition: Introduction of any residue limit or additional voluntary action is considered to negatively impact on competition given that one party will find it easier to comply than other. A regulatory limit is considered to provide for a 'more level playing field' giving a common standard to be achieved by all, and a better competitive position of EU firms with respect to foreign firms in the global market.

Conduct of businesses/SMEs: The current situation is expected to imply costs for any businesses eventually incurring in an arsenic scandal. Introduction of an additional mitigation measures (especially in case of new regulatory limits) will inevitably place additional costs on affected food industries. Given that many agri-food businesses are SMEs these might well be additionally affected.

Administrative burden on businesses: Achieving compliance with any residue limit has an administrative cost - this applies whether it is applied through legislation or voluntarily. A voluntary standard is expected, however, to generate less severe administrative costs at the aggregate level, for the lower number of adhering firms.

Public authorities: Dietary arsenic is already a public health issue and considerable investment has been/will be implemented. Introduction of regulatory limits is considered to be a more efficient route than voluntary regulation to ensure a consistent approach within the single market of the 27 Member States.

Innovation & research: The area is already being investigated in order to mitigate dietary arsenic consumption. The introduction of further controls or voluntary action would be expected to have a positive effect by giving still greater impetus to the development or improvement of more efficient analytical methods to detect arsenic in food and water.

Consumers: In terms of material cost and availability, introduction of either regulatory limits or additional voluntary controls would have the potential on increasing final prices for products affected by the limits. Such limits would not imply any change in product taste and other organoleptic characteristics.

International trade: Introduction of regulatory limits or voluntary controls has the possibility of introducing a two-tier market; with product contaminated at low levels being exported and higher contaminated product being kept for home consumption. A second issue may that exporters will no longer be able to export the volumes of the food that they had been able to do historically.

Macroeconomic environment: Establishments of new regimes (voluntary or regulatory) may lead to improved consumer confidence and therefore contribute to the maintenance of economic growth within the Community.

Labour markets: Effects on labour markets are considered to be negligible for all policy options, as it is deemed very unlikely that any intervention on arsenic in food would imply creation or loss of jobs in the food sector. This assessment has a medium level of uncertainty for a knowledge gap.

Environment: Given the geological origin of arsenic contamination in rice and that the issue of concern relates to imports from specific parts of the world outside of the EU. Establishment of legislative or voluntary limits within the Member States is not expected to have a direct environmental impact on the growing areas.

Distributive effects: Overall introduction of new regimes (voluntary or regulatory) would be considered to be more beneficial than the status quo to the most at-risk segments of society (children, elderly and those eating ethnic diets). On the other side, such regimes will impose costs to SMEs.

Societal concerns: Laypeople would see the introduction of either voluntary or regulatory regimes as positively affecting socially sensitive sub-groups of populations, like infants and young children, pregnant women, the elderly, low-income people, and food minorities (e.g. those eating ethnic diets).

3.3. Comparison of policy options and discussion

A ranking of each policy option is shown in Table 2. Regardless of the consideration or not of explicit weights and entropy, the policy option of introducing regulatory limits (option 2) has the highest 'best option' index, while the *status quo* has the lowest one. Interestingly, if we look at the 'worst option' index, the 'voluntary limits' option has the lowest value, to be interpreted as the lowest credibility that it is the worst option.

Table 2. Ranking of policy options on arsenic from fuzzy MCA.

Type of output	*Options*	*Best option index*	*Worst option index*
Not weighted	*Status quo*	0.09	0.32
	Reg. lim.	0.19	0.07
	Vol. lim.	0.13	0.03
Entropy	*Status quo*	0.09	0.31
	Reg. lim.	0.17	0.06
	Vol. lim.	0.12	0.03
Explicit weights	*Status quo*	0.10	0.48
	Reg. lim.	0.27	0.08
	Vol. lim.	0.22	0.03
Explicit weights and Entropy	*Status quo*	0.10	0.48
	Reg. lim.	0.23	0.07
	Vol. lim.	0.20	0.03

Source: own elaboration

Thus, it seems that the real policy choice is between the regulatory option and voluntary limits, which provides a higher guarantee to achieve health benefits. Policy makers with high risk aversion towards the possibility to generate the worst possible outcome would opt for voluntary limits, while those with a higher propensity on the best possible outcome would probably choose to issue a regulation.

Our results are interesting from a methodological point of view, as they show that the index rankings are not necessarily symmetric: when a policy option is ranked as the most credible 'best option', it is not consequential that it ranks as the least credible 'worst option'.

In this application, ranking results are not affected by the inclusion of explicit and/or entropy weights. However, it is worth showing the output resulting from the inclusion of different explicit weights to the impacts (Table 3). For example, we might assign the highest weight to the public health impact, as previously, but a higher importance on the impacts on businesses and public authorities than in the previous assessment.

Table 3. Ranking of policy options with different explicit weights.

Impact categories	*Impact relevance*	*Explicit weight*
Public health	10	0.15
Firm competition	2	0.03
Conduct of businesses/SMEs	9	0.13
Administrative burdens on businesses	9	0.13
Public authorities	8	0.12
Innovation and research	2	0.03
Consumers	9	0.13
International trade	1	0.01
Macroeconomic environment	1	0.01

Labour markets		5	0.07
Environment		1	0.01
Distributive effects - negative		5	0.07
Distributive effects - positive		5	0.07
Societal concerns		1	0.01
Type of output	*Options*	*Best option index*	*Worst option index*
Explicit weights	*Status quo*	0.15	0.36
	Reg. lim.	0.21	0.11
	Vol. lim.	0.16	0.05
Explicit weights and Entropy	*Status quo*	0.15	0.35
	Reg. lim.	0.18	0.11
	Vol. lim.	0.15	0.05

Source: own elaboration

Results do change as the status quo and the voluntary limits options seem to have the same credibility to be the best option, after the regulatory option, which still has the highest index, although with a lower value with respect to the previous assessment. This is just an example of how changing explicit weights may affect final results.

4. CONCLUSIONS

The aim of this paper has been to show the potentialities of an impact assessment tool based on fuzzy MCA specifically developed to food safety policies, that tries to overcome the main challenges of performing CBA.

First, all potential economic, social and environmental impacts that should be considered whenever a food safety RIA has to be undertaken were organised in 14 impact categories, including societal concerns.

Second, the tool has a rigorous structure in order to get a qualitative score for each impact category. This is especially useful for policy areas where quantification of future impacts is often difficult, if not impossible.

Third, a feasibility filter follows the principle of proportionate level of analysis, recommended by the EC-IAGs, in order to get an indication of which impacts are worth being quantified. Time is a very scarce resource in ex ante evaluation, and this step facilitates the evaluator orienteering among the different impact categories.

Fourth, a multi-criteria approach for the comparison of policy options allows for different impact categories to be assessed in different scales (linguistic, qualitative, quantitative, monetary), reflecting the different situations in terms of information and data availability on a case-by-case basis.

Fifth, a fuzzy extension of MCA allows for uncertainty in qualitative indicators to be accounted for, together with quantitative uncertainty. The fuzzy logic guides all steps of MCA pairwise comparison of policy options in terms of credibility degrees (ranging from 0 to 1) to different preference relations (option 1 is much better, better, more or less equal to, …, than option 2) between options.

Sixth, the possibility to assign weights to impacts makes explicit an implicit behaviour of evaluators and policy makers.

Overall, this tool, even though it could appear as too complicated – and so, less acceptable – to policy makers, it is in fact simple for its rigour, and more comprehensive in trying to cover the multiple aspects of food safety policies (and social regulations in general).

Of course, there is much room for improvement, like considering a time dimension (short-run vs. long-run impacts and policy outcomes), adding a procedure for collecting qualitative evaluations from multiple experts and stakeholders or quantitative estimates from models, explicitly and systematically testing the sensitivity of the final rankings to the various elements which drive the procedure, for example the elicitation instruments and scoring system or a formalised algorithm the explore the response of rankings to different weighting sets or other mathematical parameters.

ACKNOWLEDGMENTS

This work was financially supported by the European Commission through the MoniQA Network of Excellence ('Towards the harmonisation of analytical methods for monitoring quality and safety in the food supply chain'), 6th Framework Programme, contract no. FOOD-CT-2006-036337), although the views expressed here are solely those of the authors. For an overview of the socio-economic research carried out within the project, see Mazzocchi et al. (2009).

REFERENCES

Bartolini, F. and Viaggi, D. (2010). Recent developments in multi-criteria evaluation of regulations. *Quality Assurance and Safety of Crops & Foods*, 2: 182-196.

European Commission (2009). Impact Assessment Guidelines. SEC(2009) 92. 15 January 2009.

European Food Safety Authority (2009) Scientific Opinion on Arsenic in Food. *EFSA Journal*, 7(10):1351.

Gómez-Limón, J. A., Riesgo, L. and Arriaza, M. (2004), Multi-Criteria Analysis of Input Use in Agriculture. *Journal of Agricultural Economics*, 55: 541-564.

International Agency for Research on Cancer (2004). Some drinking-water disinfectants and contaminants, including arsenic. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans Volume 84.

Mazzocchi, M., Ragona, M., Giorgi, L., Schiefer, G., Fritz, M., Novelli, E., Krapp, H., Spichtinger, D. and Poms, R. (2009). Assessing the socio-economic dimension of food quality and safety regulations: Research challenges and recent advances. *Quality Assurance and Safety of Crops & Foods*, 1(3): 195-200.

Meyer, P. and Roubens, M. (2005). Choice, ranking and sorting in fuzzy multiple criteria decision aid. in: Figueira, J., Greco, S. and Ehrgott, M. (Eds.), *Multiple criteria decision analysis: State of the art surveys*. Springer, pp. 471-506.

Munda, G., Nijkamp, P. and Rietveld, P. (1992). Comparison of fuzzy sets: A new semantic distance. Serie Research Memoranda. Free University, Amsterdam.

Munda, G., Nijkamp, P. and Rietveld, P. (1995). Qualitative multicriteria methods for fuzzy evaluation problems: An illustration of economic-ecological evaluation. *European Journal of Operational Research*, 82: 79-97.

Ragona, M., Mazzocchi, M., Zanoli, A., Alldrick, A.J., Solfrizzo, M., and Egmond, H.P. van (2011). Testing a toolbox for impact assessment of food safety regulations: Maximum levels for T-2 and HT-2 toxins in the European Union. *Quality Assurance and Safety of Crops & Foods*, 3(1):12-23.

Sunstein, C.R. (2001). The Arithmetic of Arsenic. John M. Olin Law & Economics Working Paper No. 135, The Law School, The University of Chicago.

Tiwari, D.N., Loof, R. and Paudyal, G.N. (1999). Environmental–economic decision-making in lowland irrigated agriculture using multi-criteria analysis techniques, *Agricultural Systems*, 60 (2): 99-112.

APPENDIX

A.1 – Qualitative assessment

For the purposes of our analysis, we developed a specific 'scoring procedure' for food safety regulations, based on the principles of the EC-IAGs. Obviously, there are many alternative routes to scoring impacts, all with a degree of subjectivity, but the adherence to explicit rules guarantees a good degree of transparency and consistency across impact evaluations.

Within Scryer, the qualitative assessment of each impact (or sub-impact) for each policy option is broken down into four components as described in Section 3.1. These component (and the related scores) were defined as follows:

- *Direction of the impact (D)*: negative impact (-1), neutral impact (0), positive impact (+1).
- *Severity of the impact (Se):* 3-level scale (low = 1, medium = 2, high = 3). In the actual data entry form, these three levels can be associated with a qualitative description which depends on the category of impact. For example severity for the public health outcome was defined as 'effects on people well-being' (low), 'effects on morbidity levels' (medium), and 'effects on mortality levels' (high).
- *Scale of the impact (Sc):* 3-level scale (low = 1, medium = 2, high = 3). As for severity, there may be reference figures, like 'The policy option will concern less than 5% of firms' (low), whose values could be adjusted on a case-by-case basis.
- *Likelihood of the impact (L)*: 3-level scale (low = 1, medium = 2, high = 3).

From the values of *D, Se, Sc,* and *L*, the impact score for the *i*-th category of impact and the *j*-th policy is computed on a 1 to 9 discrete scale through the following steps:

1) Compute the magnitude of the individual (sub-)impact:

$$M = f(Se \cdot Sc) = \begin{cases} 1 & \text{if } Se \cdot Sc = 1 \\ 2 & \text{if } Se \cdot Sc > 1 \text{ and } Se \cdot Sc < 6 \\ 3 & \text{if } Se \cdot Sc \geq 6 \end{cases} \tag{A.1}$$

2) Combine magnitude and likelihood into a 4-level variable (*Y*) as follows:

$$Y = f(M \cdot L) = \begin{cases} 1 & \text{if } M \cdot L = 1 \\ 2 & \text{if } M \cdot L = 2 \\ 3 & \text{if } M \cdot L > 2 \text{ and } M \cdot L < 6 \\ 4 & \text{if } M \cdot L \geq 6 \end{cases} \tag{A.2}$$

3) Compute the overall impact score as:

$$x = 5 + D \cdot Y \tag{A.3}$$

In case of an impact category structured in sub-impacts, the overall impact score is obtained as an average of the sub-impact scores, which can be weighted if a different relevance for each sub-impact is defined. In this case, the resulting impact is not necessarily discrete, but equation (2) can still be applied to obtain a discrete fuzzy set from continuous *x* values.

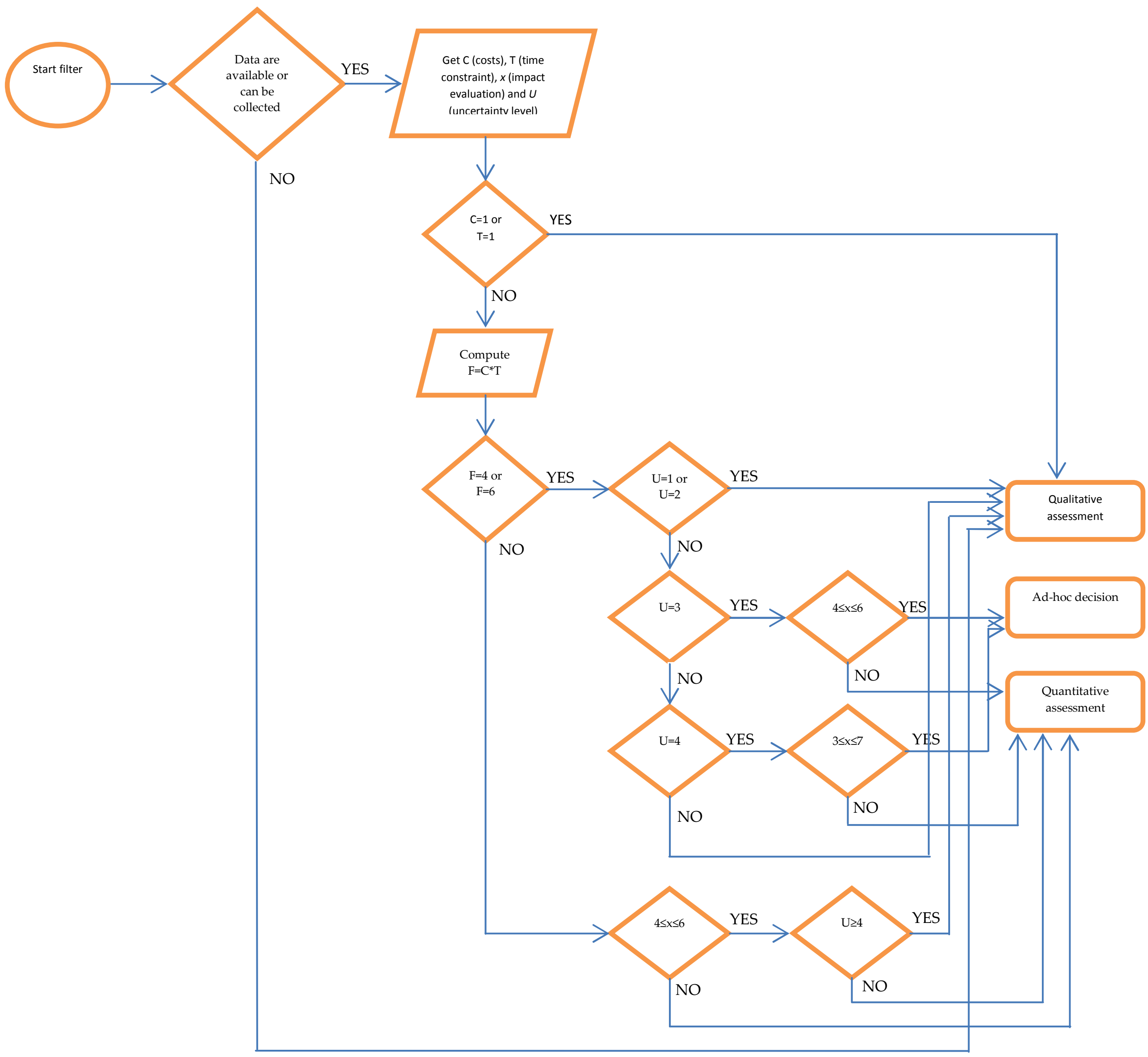
A.2 – Feasibility filter

The judgment on data availability, costs and time to perform a quantitative assessment is guided by the following possible statements:

- *Availability of primary or secondary data for quantitative evaluation (A)*: this is a yes/no question;
- *Costs for collecting quantitative data and producing a quantitative evaluation (C):* 3-level scale ('too expensive compared to resources' = 1, 'affordable costs' = 2, and 'negligible costs' = 3);
- *Time constraints for producing a quantitative evaluation (T)*: 3-level scale ('not feasible due to time constraints' = 1, 'affordable time constraints' = 2, 'minor time constraints' = 3).

The filter produces an indication on whether to pursue quantification through data-collection and modelling, based on *A*, *C*, *T*, and also on the x_{ij} and u_{ij} value obtained from the previous qualitative assessment. More specifically, simple averages $\bar{x}_i$ and $\bar{u}_i$ across the *n* policy options are considered for each impact category. The final indication is based on a hierarchical filter, as described in Figure 1.

Figure 1: The feasibility filter.



Source: Own elaboration

A.3 – Fuzzy multi-criteria comparison of policy options

This part provides the technical details of the steps described in Section 2, which closely follow Munda et al. (1995).

Transformation of qualitative indicators into Gaussian fuzzy sets

Consider the element x_{ij} of the impact matrix I, as defined in equation (1). If x_{ij} is a qualitative score, it first needs to be transformed into a fuzzy set.

In Scryer, the assumption is that these membership functions are Gaussian membership functions, while NAIADE allows for different specifications.

Formally, consider a fuzzy set Sk where k: 1,…,9 are the potential values that x_{ij} may assume, and the actual assessment is x_{ij} =q, where q is a single value between 1 and 9. The membership function is defined as follows:

$$\mu_{S^k}(x_{ij}=q)=\exp\left(-\frac{(k-x_{ij})^2}{2\sigma_k^2}\right) \tag{2}$$

where k is the centre of the fuzzy set S^k and σ^k is the width of the fuzzy set S^k (i.e. a measure of dispersion around the centre). While in NAIADE the dispersion parameter σ^k is fixed, here we allow dispersion to be a function of the centre *k* of each fuzzy set and of the stated uncertainty level *u*. This step allows to consider the uncertainty in qualitative assessments. Intuitively, a larger uncertainty (hence dispersion) generates larger grades of memberships for values that are distant from the centre of the fuzzy set (k), whereas a very low uncertainty level concentrates the membership around the centre of the fuzzy set. The assumption we adopt is that dispersion is larger for k around 5 (neutrality) and for smaller values of uij (i.e. higher subjective uncertainty). Given that the standard deviation for a continuous uniform distribution ranging from 1 to 9 is 2.58, we adopt this value as the maximum variability level. The equation for σ^k is thus the following:

$$\sigma_k=\frac{(6-u_{ij})\cdot 0.2+min(k,10-k)\cdot 0.2}{2}\cdot 2.58=(6-u_{ij}+min(k,10-k))\cdot 0.258$$

with k=1,…,9 (3)

Distance between policy options

For qualitative assessments (fuzzy sets), we exploit the semantic distance concept as in Munda et al. (1992). As in equation (2), consider x_{i1} as the qualitative impact of the first policy for the *i*-th category of impact and x_{i2} as the impact of the second policy for same category. The comparison depends on two fuzzy sets $\mu_S(x_{i1}=q)$ and $\mu_S(x_{i2}=h)$, where *q* and *h* are the two values (between 1 and 9) deriving from the scoring procedure of Appendix A. Their distance can be computed as follows:

1) Rescale the membership functions through a constant *c* so that their integral equals to 1, for example, for $\mu_S(x_{i1}=q)$

$$\int_k c\mu_{S^k}(x_{i1}=q)=1\rightarrow\sum_{k=1}^{9}c\mu_{S^k}(x_{i1}=q)=1\rightarrow c=\sum_{k=1}^{9}\mu_{S^k}(x_{i1}=q) \tag{4}$$

2) Compute the distance as:

$$D=\int_{l_{MIN}}^{l_{MAX}}\int_{m_{MIN}}^{m_{MAX}}|l-m|c_1\mu_{S^l}(x_{i1}=q)c_2\mu_{S^m}(x_{i2}=h)dldm=\sum_{l=1}^{9}\sum_{m=1}^{9}|l-m|c_1\mu_{S^l}(x_{i1}=q)c_2\mu_{S^m}(x_{i2}=h) \tag{5}$$

In short, the semantic distance is a weighted average of all potential distances between the linguistic values, weighted by their membership functions.

For the Scryer application, the minimum distance between two fuzzy sets occurs when $q = h$ and $U=5$ (minimum degree of uncertainty). The value of this distance is 0.112. On the other hand, the maximum distance occurs when $q=1$ and $h=9$ (or vice versa) and $U=5$. In this case, the maximum distance is 0.834.

It is convenient to re-scale the distance in (5) in order to expand the range to the domain [0,1]. Thus the distance equation employed in Scryer is:

$$D = \frac{\sum_{l=1}^{9}\sum_{m=1}^{9}|l-m| c_1 \mu_{S^l}(x_{i1}=q) c_2 \mu_{S^m}(x_{i2}=h) - 0.112}{0.834} \tag{6}$$

When the impact is quantitative, the distance between two impacts can be computed by assuming a normal distribution and exploiting the Hellinger distance:

$$DH = \sqrt{1 - \sqrt{\frac{2s_1s_2}{s_1^2+s_2^2}}\, e^{-\frac{1}{4}\frac{(x_{i1}-x_{i2})^2}{s_1^2+s_2^2}}} \tag{7}$$

where s_1 and s_2 are the standard errors of the estimated impacts x_{i1} and x_{i2}, respectively.

Credibility values for pairwise policy comparison

Credibility values are computed for a set of 'preference relations' between two policy options for each of the c criteria (i.e. impacts). Consider two policy options P_1 and P_2. There are six statements as follows:

>>	P_1 is much better than P_2 (according to criterion i)
>	P_1 is better than P_2
≅	P_1 is more or less like P_2
=	P_1 is identical to P_2
<	P_1 is worse than P_2
<<	P_1 is much worse than P_2

For each of these statements, the *credibility value* is generated, based on the semantic distances (D) and some fixed thresholds (χ). We use Munda's approach (Munda et al., 1995) to compute credibility values for each statement, based on the comparison of the comparison between the two qualitative impacts for the same criterion (e.g. x_{i1} and x_{i2} as defined above):

$$c_{>>,i}(P_1,P_2) = \begin{cases} 0 & \text{if } x_{i1} \le x_{i2} \\ \dfrac{1}{\left(1+\dfrac{\chi_{>>}{}^2\left(\sqrt{2}-1\right)}{D^2}\right)} & \text{if } x_{i1} > x_{i2} \end{cases}$$

(much better)

$$c_{>,i}(P_1,P_2)=\begin{cases}0 & \text{if } x_{i1}\le x_{i2}\\ \dfrac{1}{\left(1+\dfrac{\chi_{>}^{2}}{D^2}\right)} & \text{if } x_{i1}> x_{i2}\end{cases}$$ (better)

$$c_{\cong,i}(P_1,P_2)=e^{-\left(\frac{\ln 2}{\chi_{\cong}}D\right)}$$ (approximately equal)

$$c_{=,i}(P_1,P_2)=e^{-\left(\frac{\ln 2}{\chi_{=}^{2}}D^2\right)}$$ (equal)

$$c_{<,i}(P_1,P_2)=\begin{cases}0 & \text{if } x_{i1}\ge x_{i2}\\ \dfrac{1}{\left(1+\dfrac{\chi_{<}^{2}}{D^2}\right)} & \text{if } x_{i1}< x_{i2}\end{cases}$$ (worse)

$$c_{<<,i}(P_1,P_2)=\begin{cases}0 & \text{if } x_{i1}\ge x_{i2}\\ \dfrac{1}{\left(1+\dfrac{\chi_{<}^{2}\left(\sqrt{2}-1\right)}{D^2}\right)} & \text{if } x_{i1}< x_{i2}\end{cases}$$ (much worse)

Aggregation across criteria and weights (pairwise comparison of policies)

Pairwise comparison of policies, considering all criteria at once, is based on a 'preference intensity index' (see equation (45) in Munda et al., 1995). At this stage one may wish to assign (explicitly) different weights to the various criteria. The equation for the aggregate preference intensity index for each of the 6 preference statements (hereafter generally indicated with a subscript *) can be written as:

$$\mu_*(P_1,P_2)=\frac{\sum_{i=1}^{c}\max(c_{*,i}-\alpha,0)w_i}{\sum_{i=1}^{c}\left|c_{*,i}-\alpha\right|w_i} \qquad (8)$$

where $w_i\in [0,1]$ (with i:1,…,c) are the weights assigned to each criterion, $\sum_{i=1}^{c} w_i=1$. These weights are computed in Scryer as a function of statements on the relative importance of each impact.

Entropy measure of consistency across criteria

Similar preference intensity indices may hide very heterogenous situations, in terms of consistency across the credibility indices for the various criteria. Thus, it is advisable to generate an entropy measure, which can then be applied to 'weigh' the preference intensity indices in the final policy ranking step (see equations (46) to (48) in Munda et al., 1995). First, an adjusted membership function can be defined for each policy comparison, considering a threshold to rule out very small preference intensities:

$$\gamma_{*,i}(P_1, P_2) = \begin{cases} 0 & \text{if } c_{*,i}(P_1, P_2) < \alpha \\ c_{*,i}(P_1, P_2) & \text{if } c_{*,i}(P_1, P_2) \geq \alpha \end{cases} \tag{9}$$

Then, starting from this membership function, entropy is computed (as in Munda et al., 1995) as:

$$H_*(P_1, P_2) = \frac{1}{c}\sum_{i=1}^{c} -[\gamma_{*,i}(P_1, P_2)\cdot \lambda^A_{*,i}(P_1, P_2) + (1-\gamma_{*,i}(P_1, P_2))\lambda^B_{*,i}(P_1, P_2)] \tag{10}$$

where

$$\lambda^A_{*,i}(P_1, P_2) = \begin{cases} 0 & \text{if } \gamma_{*,i} = 0 \\ \log_2 \gamma_{*,i} & \text{otherwise} \end{cases} \tag{11}$$

and

$$\lambda^B_{*,i}(P_1, P_2) = \begin{cases} 0 & \text{if } \gamma_{*,i} = 1 \\ \log_2 (1-\gamma_{*,i}) & \text{otherwise} \end{cases} \tag{12}$$

Each aggregate credibility value for pairwise policy comparison can now be accompanied by an entropy measure *H*, which increases as the basic credibility values concentrate around 0.5 (i.e. uncertainty), whereas tends to 0 when most of the basic credibility values are 0 or 1 (i.e. certainty). The extremes are *H=0* when all basic credibility values are 0 or 1, and *H=1* when all basic credibility values are 0.5.

Ranking of policies

Starting from the pairwise comparison of policies through the preference intensity indices (either weighted or unweighted), it is possible to rank the policies. At this stage entropy can be considered.

$$\phi^{+}(P_i) = \frac{\sum_{j\neq i}^{p} \mu_{>>}(P_i,P_j)\cdot[1-H_{>>}(P_i,P_j)] + \mu_{>}(P_i,P_j)\cdot[1-H_{>}(P_i,P_j)]}{\sum_{j\neq i}^{p} 2-H_{>>}(P_i,P_j)-H_{>}(P_i,P_j)} \quad (13)$$

$$\phi^{-}(P_i) = \frac{\sum_{j\neq i}^{p} \mu_{<}(P_i,P_j)\cdot[1-H_{<}(P_i,P_j)] + \mu_{<<}(P_i,P_j)\cdot[1-H_{<<}(P_i,P_j)]}{\sum_{j\neq i}^{p} 2-H_{<}(P_i,P_j)-H_{<<}(P_i,P_j)} \quad (14)$$

Where p is the number of policy options being considered. The same indicators can be computed without the entropy weighting, by omitting the terms in square brackets and using *2(p-1)* as the denominator.

$$\phi^{+}(P_i) = \frac{\sum_{j\neq i}^{p} \mu_{\gg}(P_i,P_j) + \mu_{>}(P_i,P_j)}{2(p-1)} \quad (15)$$

$$\phi^{-}(P_i) = \frac{\sum_{j\neq i}^{p} \mu_{<}(P_i,P_j) + \mu_{\ll}(P_i,P_j)}{2(p-1)} \quad (16)$$

For each policy option, the above equations aggregate the much better (much worse) and better (worse) preference intensity index, to generate an aggregate preference index for the best (worst) policy option. They can be interpreted as a degree of membership to the statements that 'Policy alternative i is the best policy option' and 'Policy alternative j is the worst policy option'. All policy alternatives can now be ranked according to Φ^{+} and Φ^{-}, which range between 0 and 1.