

A review of risk prioritisation schemes of pathogens, pests and weeds: principles and practices

Jaakko Heikkilä

MTT Economic Research, Latokartanonkaari 9, 00790 Helsinki, Finland

e-mail: jaakko.heikkila@mtt.fi

Society's resources are scarce, and biosecurity actions need to be targeted and prioritised. Various models have been developed that prioritise and rank pests and diseases according to the risks they represent. A prioritisation model allows utilisation of scientific, ecological and economic information in decision-making related to biological hazards. This study discusses such models and the properties associated with them based on a review of 78 prioritisation studies. The scope of the models includes all aspects of biosecurity (human, animal and plant diseases, and invasive alien species), but with an emphasis on plant health. The geographical locations of the studies are primarily North America, Europe, Australia and New Zealand. Half of the studies were conducted during the past five years. The review finds that there generally seems to be several prioritisation models, especially in the case of invasive plants, but only a select few models are used extensively. Impacts are often accounted for in the model, but the extent and economic sophistication of their inclusion varies. Treatment of uncertainty and feasibility of control was lacking from many studies.

Key-words: biosecurity, pest, pathogen, weed, prioritisation, model, review

Introduction to prioritisation

Society's resources are scarce. The state cannot, for instance, provide perfect biosecurity, and resources need to be targeted and prioritised. In practice, prioritisation is often based on opinions and views of different agents rather than on risk assessment. It is nonetheless possible to render prioritisation transparent and assist in optimal targeting of societal resources by using economic tools and principles as well as risk assessment.

This study reviews and discusses 78 studies where biological hazards are ranked according to the risk that they present. The aim of this study is to gain an overview of the types of studies conducted, to assess the strengths and weaknesses of prioritisation, and to note good practices for conducting such studies. The scope of the models discussed includes all aspects of biosecurity (human diseases, animal diseases, plant pests and diseases, and invasive alien species), but with an emphasis on plant health. Although the outcome is usually time and place specific, and thus cannot be directly applied elsewhere, we can learn a great deal from the prioritisation models themselves.

Morgan et al. (2000) noted that ranking of environmental risks in general has increased, particularly in the United States, Canada and New Zealand. There is also political demand for prioritisation. For example, within the European Union there is a desire to produce a scheme that can be used to classify and prioritise animal diseases and their management based on their health and economic impacts (European Commission 2007). Similar plans exist in relation to plant health and invasive species, and risk-based surveillance is becoming increasingly popular (McKenzie et al. 2007). In relation to surveillance of communicable human diseases, the World Health Organization (WHO) considers it to be useful to ensure that planning and resource targeting is rational, explicit and transparent. They point out that since surveillance systems have developed over time, new diseases have been added to the lists without any old ones being removed (WHO 2006).

Australia and New Zealand have officially been using a prioritisation tool for weed risk assessment since the late 1990s (Gordon et al. 2008a; Weber et al. 2009) and Canada has undertaken systematic prioritisation of human diseases for surveillance over about the same period. In Great Britain human diseases were prioritised by Public Health Laboratory Service (in 1995, 1997 and 1999), but the Health Protection Agency, which replaced it in 2003, discontinued the practice (DEFRA 2006). In relation to their new animal health strategy, the United Kingdom has been developing a methodology for surveillance based on transparent prioritisation of risks and impacts, aiming at more efficient use of public resources (DEFRA 2003).

Notwithstanding these few initiatives, in a literature review carried out as a part of the UK animal disease prioritisation, it was noted that many organisations do not use such methods to support decision-making (Gibbens et al. 2006). Fox and Gordon (2004) reviewed 113 action lists of invasive plants from 16 countries, and found that in about 17% of cases there was no explanation as to why the species were on the list, in 67% there were single sentences that used vague terms such as "harmful" or "causes harm", and in the rest (16%) there were several criteria for inclusion. Ranking was done on fewer than 10% of the lists and 14% used choice questions that determined which species were selected for inclusion on the list (Fox and Gordon 2004).

The problem in the absence of prioritisation studies is that resource allocation is likely to be inefficient. For instance, Virtue (2007) pointed out that in Australia too many species have been declared weeds in comparison with the resources available for their effective management or eradication. He further pointed out that species have been targeted for different reasons in different areas, including long histories of control, substantial visibility, political pressure, suspected impacts, help received in management, knowhow related to the species, and pressure from agriculture. Economics-based prioritisation has not impacted on the choices in any significant way (Virtue 2007). Similarly, risk does not seem to be the primary determining factor in control of many animal diseases in Finland (Rosen-

gren and Heikkilä 2009). DEFRA (2006) pointed out that in the private sector cost-benefit analysis guides investments, but due to multiple simultaneous objectives the situation is more challenging in the public sector, which is also led by political objectives and public opinion. However, in the academic literature various models and application frameworks have been developed to rank pests and diseases according to the risk they represent. Ideally, resources would be distributed so as to provide maximum social welfare, the objective being to get the best value-for-money by controlling the most harmful and most manageable hazards.

Prioritisation can be undertaken in relation to various targets: it is possible to prioritise individual animals (for disease sampling), farms (for surveillance), persons (for vaccination) or geographical regions (for surveillance and control). This review addresses prioritisation of different kinds of biological hazards, which can be carried out at two hierarchical levels. Firstly, prioritisation can be done across different categories of biological hazards (e.g. across animal and plant diseases). This kind of prioritisation is relatively uncommon. The more common type of prioritisation is completed within a single type of biological hazard – for instance within plant diseases and pests, or within individual families (e.g. ants). Such prioritisation often follows the basic structure of Covello-Merkhofer risk assessment (Covello and Merkhofer 1993), and is composed of separate criteria, including the probability of entry (invasion, introduction, outbreak), the probability of establishment (spread, invasiveness), and the likely impacts on various processes, which may or may not be measured in monetary terms. These criteria are then subdivided into sub-criteria and individual questions, where different answers attract a specified number of points. These points are aggregated and possibly weighted to form the total score (see, e.g., Doherty 2000).

The terms used in this review may differ from those that the reader is familiar with, because the use of terminology differs somewhat in the different areas of biosecurity. Some of the models included are traditionally viewed as screening tools (for instance the Australian Weed Risk Assessment, WRA), but the term prioritisation model is here

understood as any structured system that places biological hazards into a ranking order by asking the evaluator a series of questions.

The rest of the paper is organised as follows. The general details and basic components of the reviewed studies are discussed next, followed by a discussion of the benefits and challenges of prioritisation. Finally, some conclusions are provided.

Models for prioritisation

75 distinct prioritisation studies were reviewed for this article, a comprehensive list and data table can be found as supplementary material on the publisher's web site. In some studies more than one evaluation framework was used, and hence 78 cases are summarised here.

The geographical location of the studies included primarily Europe (26 studies), North America (23), and Australia and New Zealand (19). Most of the studies concentrated on environmental health, human health and plant health, but the division between environmental and plant health is sometimes ambiguous. Food safety (7) and animal health (7) were the least represented areas of biosecurity. Half of the studies (39) were conducted during the past five years (2006–2010). Two thirds of the studies were reported in reviewed scientific publications (journals, conference proceedings or theses). The maximum number of ranked organisms in an individual study was 851 (Hayes and Sliwa 2003), while the mean was 83 and the median 37 organisms. About 15% of the studies did not apply the model that was developed.

There generally seems to be a fair amount of prioritisation models available for practical application, especially in the case of invasive plants. Of the 78 studies, 55% were new model developments, 26% were straightforward applications of existing models, 6% were comparative tests of several models and 13% were applications or comparisons that involved further model development. However, only a select few models are used extensively, and these deal mostly with invasive

plants. Some of the most popular models are listed in Table 1.

The evaluation panel constitution and size also varied widely, and was often not reported. At least in a quarter and probably in about half of the cases it was the authors themselves who acted as the panellists. When reported, the size of the evaluation panel varied from 1 (Pheloung et al. 1999, Gordon et al. 2008b) to 1174 (More et al. 2010), with a median panel size of 10.

Entry of an organism was considered in 27%, establishment in 71% and impacts in 90% of the studies. The general lack of entry assessments reflects the fact that many of the studies were for species that were proposed for intentional introduction, in which case entry becomes irrelevant. Although impact was accounted for in most of the studies, the extensiveness and sophistication of the questions varied substantially. Of the different impact types, social and trade effects were mostly lacking from the studies, whereas impacts on human health, agriculture and the environment were mostly included. The impacts were often consid-

ered in one system only, although for instance in the Australian WRA the impacts on the environment and agriculture were considered separately.

Number of questions

The number of questions in the reviewed studies varied. 24% of the studies had fewer than 10 questions, 26% between 10 and 19 questions, 17% between 20 and 29 questions, 9% between 30 and 48 questions and 19% had 49 questions (the number of questions in the widely applied Australian WRA). The mean number of questions was 22 and the median was 17. The Australian WRA model involves a relatively large number of questions so as to reduce the need for subsequent evaluation (Parker et al. 2007). Generally, it can be argued that the higher is the number of questions, the more precise is the outcome of the model, and the more difference there will be in the total scores of the organisms.

Table 1. Some widely used prioritisation models.

| System | Original application | Used also for/in |
|---------------------------------------|---|---|
| Australian Weed Risk Assessment (WRA) | Weeds in Australia (Pheloung 1995) | Weeds in Australia (Cunningham et al 2004), Australia and New Zealand (Pheloung et al. 1999), Hawaii (Daehler et al. 2004; Buddenhagen et al. 2009), Florida (Gordon et al. 2008b), Chicago (Jefferson et al. 2004), Spain (Gassó et al. 2010; Andreu and Vilà 2010), Italy (Crosti et al. 2010), Czech Republic (Křivánek and Pyšek 2006), Japan (Nishida et al. 2009; Kato et al. 2006), and Tanzania (Dawson et al. 2009). Also applied to fish in the UK (Copp et al. 2005; Copp et al. 2009; Tricarico et al. 2010). |
| Reichard-Hamilton decision tree | Weeds in North America (Reichard and Hamilton 1997) | Plant health in Hawaii (Daehler and Carino 2000), Chicago (Jefferson et al. 2004), Iowa (Wirdlechner et al. 2004) and the Czech Republic (Křivánek and Pyšek 2006). |
| Risk Ranger | Food safety in Australia (Ross and Sumner 2002) | Food safety in the EU (Mataragas et al. 2008) and in Australia (Sumner and Ross 2002; Sumner et al. 2005). |
| Weber and Gut model | Weeds in central Europe (Weber and Gut 2004) | Weeds in Spain (Andrey and Vilà 2010). |

There has been some discussion regarding whether the number of questions (or the number of questions answered) affects the outcome of the model. Daehler et al. (2004) noted that there is a statistically significant linear relationship between the number of questions answered and the WRA score of the evaluated organisms, but the fit of the regression is low ($R^2=0.08$). The case is similar to that reported by Kato et al. (2006), whereas Dawson et al. (2009) found no statistical connection. In other words, although there may be a statistical connection, the number of questions only explains a small proportion of the evaluation outcome.

Although it has been noted in different studies that the ranking is not greatly affected when the number of questions is reduced (e.g. Daehler and Carino 2000), it is difficult to state *a priori* which questions determine the outcome and which could be left out. Models with few questions may reliably predict the occurrence of harmful pests, but also tend to predict harmless species to be harmful (e.g. Gordon et al. 2008b). Hence, the sensitivity of a model with very few questions is high, but its specificity suffers. Sensitivity refers to proportion of true positives (e.g. pests) correctly identified, and specificity to proportion of true negatives (e.g. non-pests) correctly identified (Altman and Bland 1994). This property has also been reported in other studies (Reichard and Hamilton 1997, Krivánek and Pyšek 2006, NWRAS Review Group 2006). Gordon et al. (2008b) suggested that questions which have been found to predict the outcome reliably could be used for pre-screening, in which case a ‘yes’ answer in an import risk assessment would lead to outright denial of entry, whereas all other answers would lead to conducting the full assessment.

Increasing the number of questions also increases the resources required to evaluate the organism. The reported time estimate for the Australian WRA is from 5 hours (Kato et al. 2006) to 1-2 days per species (Jefferson et al. 2004), and for the Hawaiian WRA 5-8 hours per species (Daehler et al. 2004). Of course, the time taken depends on the evaluators, the mode of data acquisition, and the existence of biological and other relevant data.

Point scales, score aggregation and weighting

Of the studies reviewed, approximately 5% used a numerical scale combined with a binary yes/no scale, while the remainder used semi-quantitative scales. While these semi-quantitative scales contain numbers, the numbers are not an exact measure of performance, but are instead means of translating qualitative information into quantitative data. For instance, agricultural damage may be measured on a scale (e.g. 1-5) rather than in absolute monetary units. Such a model is faster to apply than a fully quantitative approach, as often the information available is not accurate enough to allow full quantification. On the negative side, the scores and their aggregation are arbitrary, which may not be fully transparent (McKenzie et al. 2007) and causes difficulties when assigning the points for each response option.

Letters have also been used (e.g. A-E in Invasive Species Assessment Protocol, Morse et al. 2004). Since these are converted to numbers (1-5) when score aggregation is conducted, the use of letters probably more concerns psychology and allowing the evaluators to concentrate on the descriptions of the available choices. Nonetheless, whether using letters or numbers, it is important that the options have clear verbal descriptions in order to reduce the scope for differing interpretations by different evaluators. The clearer the descriptions are, the more trustworthy is the outcome of the scoring (Ryan 2006). Some models (e.g. Australian WRA) have specific guidance on how to answer the questions (Gordon et al. 2010), whereas for others no such guidance may be available.

Many studies used the Likert-type scale, which typically ranges from 1 to 5. However, McKenzie et al. (2007) chose a scale from 1 to 4 in order to avoid having a central value (3), and hence forcing the evaluators to choose whether the property is more or less likely. It has been found (Dawes 2008, see also Makowski and Mittinty 2010) that there is no difference in the results produced by 5- and 7-point scales, but results from a 10-point scale are statistically somewhat lower. Moreover, it has been

argued that having a scale running from 1 to 9 is unclear, as each alternative is unlikely to be clearly defined (MacLeod and Baker 2003). For instance, the European and Mediterranean Plant Protection Organization (EPPO) advocates such a scale, and in some studies (e.g. Copp et al. 2005) that have used the EPPO criteria, the authors have actually modified the scale to reduce the number of options. Evaluation on a scale 1-3 (low, medium, high) or 1-5 (very low, low, medium, high, very high) is more likely to yield similar results from different evaluators, and hence be more objective. On the other hand, if there are few questions, a scale with only three available options could result in little difference to the total score of the organisms, hence making ranking more difficult. In for instance the Australian WRA most questions are on a 3-point scale, but due to a relatively large number of questions the total scores may differ substantially. In the reviewed studies a vast scale range was evident, including [-1, 0, 1], [0-1], [0-2], [0-4], [0-5], [0-6], [0-10], [0, 0.5, 1], [0, 1, 3, 6, 10], [1-3], [1-4], [1-5], [1-9], [1-10], and so forth. Scales [-1, 0, 1], [0-1], [0-5] and [1-5] were somewhat more common than others, however.

The overall rank or risk score is typically obtained by either multiplying the different criteria values by each other or by summing them. One property of the multiplicative approach is that the total score approaches zero if any of the individual criteria does so. For instance, if the probability of entry is zero, then the total score is zero as well. In an additive model it is possible that the total score is still relatively high, although for instance habitat suitability might be very low (Parker et al. 2007, Cox 2009). Makowski and Mittinty (2010) simulated the outcome of several scoring systems and showed that multiplication-based systems performed better than sum-based systems. Despite the appeal of the rationale behind multiplicative score aggregation, the majority of the reviewed studies applied additive score aggregation (62% of studies), while in only 14% was the score aggregation multiplicative. In a further 5% the aggregation was by a matrix, in 5% by a decision tree and in 6% there was no score aggregation. This breakdown remains qualitatively similar when all the reviewed

model frameworks are included only once (i.e. applications are excluded). There were also studies where the results were aggregated by criteria (e.g. Ciotti 2003; Weber and Gut 2005). In the Australian WRA (Pheloung 1995) the outcome can be separated by whether the impact is on the environment or on agriculture without having to combine the impacts (although the model also produces a combined score).

In many cases in score aggregation the mean of the points is taken at some stage, such as when forming a criteria or sub-criteria score from individual questions. This is problematic since taking the mean has the tendency to make the scores more alike between organisms. Holt (2006) questioned how well such a tendency towards the mean reflects the actual risk, as it may overestimate low risk and underestimate high risk. However, the ranking order remains unaffected, and this may not be a serious issue if there is only interest in the order of the organisms in the ranking, not in how much more serious a threat one organism represents compared with another.

As models often consist of separate criteria (e.g. entry, establishment, impacts) and criteria consist of sub-criteria or individual questions, it has to be decided in score aggregation whether each criteria or each question within the assessment carries the same or a different weight in determining the final score. Criteria or questions may be weighted to give a greater importance for certain components in predicting the outcome of the evaluation. In such a case for instance chi-square analysis can be used to evaluate the predictive power of certain questions, and those having high importance can be weighted more heavily. Similarly logistic regression can be used to determine the weights given to different sub-categories. Such assessment of weights was rare in the reviewed studies. Weighting can also be done to better reflect social preferences, for instance agricultural impacts may be weighted more heavily than impacts on recreation. For instance, in the Invasive Alien Plant Program's Species Scoring Algorithm (Garry Oak Ecosystems Recovery Team 2007), the impacts were evaluated such that impact on human health received a score of 5, animal

health, and natural or agricultural environment 4, native plants 3, recreational use 2 and aesthetics 1.

The weights can be determined by the researchers or they may be subjected to evaluation by experts or a panel; possibly the same panel that does the actual evaluation (e.g. Darin 2008). Krause et al. (2008) set the weights by having experts put the criteria in the order of importance, and then taking the mean value of the rank as the weight. They also suggested that weighting should be done separately from the actual analysis, or at least before it, in order to make it more objective. Similarly, Cook and Proctor (2007) allowed each expert in the panel to design their own weighting factors, and the weights used in the assessment were derived from the means of these values. However, they also used the distribution of the weightings among experts to measure the level of uncertainty related to ranking.

59% of the reviewed studies applied weighting. In the model weighting can be applied by assigning different point scales for different criteria or questions. 22% of the reviewed studies used several (four or more) scales, even to the extent that in some studies almost every question was scored on a different scale, meaning that different questions had different weights in the calculation of the total score. The number of questions in different criteria can also be used for weighting. This would be the case when, for example, each question is scored on a scale 1 to 5, and there are different numbers of questions in different criteria. The criteria then automatically gain different weights. Weighting can also be applied directly using transparent weights by which the score of criteria or questions is multiplied. When weighting is completed separately from the actual scoring, it is easy to apply any desired weighting, whereas if it is inbuilt in the model, the whole scoring system needs to be changed to change the weights (Krause et al. 2008). In the reviewed papers weighting was primarily applied such that it is inbuilt in the scoring system: different questions or criteria get a different number of points. In only a few cases (e.g. Darin 2008, Randall et al. 2008, Ou et al. 2008, Ward et al. 2008) were specific weights explicitly assigned to criteria, making the assessment more transparent.

Uncertainty and validation

In prioritisation there are several types of uncertainty associated with the different impacts, the quality and existence of current information, and regarding panel members' information (Ryan 2006). Put simply, uncertainty can be related to data inputs or data outputs. Data input uncertainty refers to uncertainty regarding the information needed for the evaluation, for instance uncertainty regarding some characteristics of the organism. Output uncertainty refers to how reliable the outcome of the prioritisation model is. Dealing with uncertainty regarding the model outcome can be regarded as validation of the model (see, e.g., Caley et al. 2006; Gordon et al. 2008a; Hughes and Madden 2003).

There are several ways one can account for these uncertainties (Table 2). Still, input uncertainty was not taken into account in 47% of the studies. When it was included, it was primarily by the model being built such that it was not necessary to answer a question if the answer was not known. In these cases the treatment of “do not know” answers differed markedly. For instance, McKenzie et al. (2007) used the worst case scenario when no response was made, whereas Petersen et al. (1996) allocated mid points, while Branquart et al. (2007) gave no score at all for no information. As for output uncertainty, 36% did not include treatment of output uncertainty (validation) and in a further 12% there was no application, so it was not explicitly tested whether the model actually produces sensible results. When it was included, the most common validation (36% of studies) was to test the model with species whose pest status is known, and assess whether the model produces acceptable results.

Feasibility of control and management

The basic criteria (entry, establishment and impact) can be augmented with a measure for feasibility of control. As mentioned earlier, the objective of the prioritisation exercise should be to maximise social wellbeing through allocation of the resources such

Table 2. Methods for including uncertainty.

| Methods for including input uncertainty | Methods for including output uncertainty |
|--|---|
| 1. Provide a score for uncertainty related to each answer or for the reliability of information used in answering the question (e.g. journal, observation, anecdotal) (Risk Assessment and Management Committee 1996; Parker et al. 2007; Warner et al. 2003). | 1. Second round of panel evaluation, where the results of the first round are fed back to the panel (or other experts) (Horby et al. 2001; Doherty 2006; Weinberg et al. 1999). |
| 2. It is not necessary to answer all the questions or can answer "do not know" (Pheloung 1995; Pheloung et al. 1999; Petersen et al. 1996). | 2. Model sensitivity tested by changing the input values by a given percentage and calculating the impact on the final results (Parker et al. 2007). |
| 3. Can answer multiple choices (i.e. a scale of answers) instead of a single choice (Morse et al. 2004). | 3. Hazards with moderate risk scores are assigned to an "evaluate further" category and a second round of evaluation (Pheloung 1995; Pheloung et al. 1999; Daehler et al. 2004). |
| 4. Use more than one evaluator (Copp et al. 2009; Ciotti 2003; Horby et al. 2001; Cook and Proctor 2007). | 4. Results are given as a distribution instead of simple point scores (NWRAS Review Group 2006). |
| 5. Specifically ask about uncertainty related to the hazard (SZEID 2006). | 5. Results are validated by comparing some of them against known hazards (Smallwood and Salmon 1992; Weber and Gut 2004; Gassó et al. 2010; Mataragas et al. 2008), by subjecting them to expert evaluations (Daehler et al. 2004; Pheloung et al. 1999) or by comparing against results of other models (Champion and Clayton 2000; Daehler and Carino 2000; Parker et al 2007). |
| 6. Do not include the hazard if there is uncertainty; not recommended in most cases (Hayes and Sliwa 2003). | |

that the investments produce maximum net benefits. Since entry, establishment and impact potential can be affected by human actions, not including those human actions in a social prioritisation exercise is likely to produce a sub-optimal outcome. Because of this, it is not only the risk but also the controllability of the organisms that should be accounted for in prioritisation. Virtue (2007) augmented the risk measure by a containment feasibility measure, which is the product of control costs, current distribution and persistence. Score is then obtained by dividing the risk measure by the control feasibility measure. Hiebert and Stubbendieck (1993) used a slightly different methodology and calculated the score separately for impact and control, and then plotted the impact against the effectiveness of control, giving four possible outcomes: serious threat, difficult to control; serious threat, easy to control; small threat, difficult to control; and small threat, easy to control.

Control can comprise preventability or treatability or both. It is clear that for some diseases or species one or the other may be more important. Only in one study (Krause et al. 2008) were both preventability and treatability included. In 36% of the studies feasibility of control was not accounted for in any way. Even when it was included, it was often based on only one or few questions. This can be argued to be the main shortcoming of the prioritisation approach from the perspective of resource allocation. More constructively, this is an area where further research and model development is required.

Discussion

Prioritisation tools are designed to deal with multiple species in a relatively short period of time. In other

words the resource requirements for evaluation are not excessive. In the reviewed studies, when mentioned, it took 1 or 2 days to find the required information and evaluate a species. In a thorough risk assessment an evaluation may take months or years, whereas the current methods used in the United States take about 2-8 weeks per species (Parker et al. 2007). Further, prioritisation models evaluate the organisms using the same set of questions. This, albeit being a constraint in some sense, allows comparison between the species and subsequently their prioritisation. Caution is required if the organisms are grouped for evaluation: for instance, the results are likely to vary depending on whether H1N1 and H5N1 influenzas are grouped together with common influenza, or whether it is “sexu-

ally transmitted diseases” or “HIV/AIDS” that are evaluated (Horby et al. 2001; Morgan et al. 2000).

There is no objectively correct way to carry out prioritisation. A number of ideal methodological properties have been put forward in the literature, including the following (after Daehler et al. 2004, Virtue 2007, Ciotti 2003): 1) components have a scientific basis that is mathematically simple but logical; 2) the scheme is fully transparent; 3) the questions are understandable and generic enough to allow application to a range of circumstances; 4) the evaluation process minimises the impact of subjective views and is repeatable such that two persons evaluating the same organism reach a similar outcome; 5) there are as few questions as possible, but the comparison is robust; and 6) there is a possibility to use all available data. Further sug-

Table 3. Suggested good practices for prioritisation studies.

| Suggested good practices | |
|--------------------------|--|
| 1. | Establish the aims and objectives of the prioritisation exercise: why are you ranking the organisms? |
| 2. | Consider whether grouping of the organisms is required, or whether you can evaluate each organism separately. |
| 3. | The number of questions should be sufficient (for the precision of the model) but not excessive (to keep the evaluation resource requirements moderate) – more than 10 but less than 50 could serve as a reasonable rule of thumb. |
| 4. | Use moderate point scales – no more than seven alternative answers per question, and preferably about five – naturally depending on the question. |
| 5. | Provide clear descriptions as to what each alternative answer to each question means. |
| 6. | Include entry, establishment and impacts in your model, unless you have a specific reason to exclude some of them. |
| 7. | In the case of impacts, consider all relevant impacts. |
| 8. | Include feasibility of control (preferably both prevention and treatment) in your model, unless you have a specific reason not to do so. |
| 9. | Include input uncertainty in your model, preferably using more than one method of those listed in Table 2. |
| 10. | Allow for weighting in your model, at least at the level of different criteria, even if you do not use it yourself. Do not build weights into the scoring system, but rather use transparent multiplicative weights that can be easily modified. |
| 11. | Consider whether you want score aggregation to be additive or multiplicative. If you choose additive score aggregation, consider whether some critical sections should still be multiplicative. |
| 12. | Apply and test your model. Test for internal correlations within the model. |
| 13. | Validate the model, preferably using at least one method of those listed in Table 2. |
| 14. | Document the model properly for others to review, apply and develop. |

gestions and “good practices” based on this study are collected in Table 3.

In this review, it was found that there generally seems to be a fair amount of prioritisation models available for practical application, especially in the case of invasive plants, but only a select few models are used extensively. Many studies employ previously developed models, particularly the Australian WRA. The studies have been conducted primarily during the last five years. Impacts, including specific economic impacts, are often accounted for in the model, but the level of their inclusion varies and, for instance, social and trade impacts are mostly missing.

Treatment of uncertainty was lacking in about 40% of the studies and feasibility of control was not accounted for in over one third of the studies. Among the 78 models, there were two that included entry, establishment, impacts, control feasibility, input and output uncertainty and allowed weighting, these being the U.S. Weed Ranking Model (Parker et al. 2007) and the Smallwood and Salmon Rating System (Smallwood and Salmon 1992). A further three models included all but one of the above elements (Australian WRA by Pheloung 1995 and its derivatives; AWRAM model by Champion and Clayton 2000, and the Chinese Weed Risk Assessment by Ou et al. 2008).

The results of the review are representative of plant health prioritisation studies, which did not differ significantly from the other studies. There were relatively fewer plant health prioritisation applications in Europe (and more in the USA and Canada), and the plant health studies evaluated relatively greater number of organisms per study. The Australian WRA dominates as the main model in the case of plant health, but there are also other models available – generally much more than in the case of animal health or food safety.

Vall-Ilosera and Sol (2009) argued that prioritisation models such as those reviewed here are not based on statistics, are qualitative, based on expert opinions and require a vast amount of information. The nature of questions and the final score are arbitrary, and are not always based on scientific information. They further argued that ranking systems are sensitive to missing information because

they give the highest value to missing information. Although their criticism is to an extent warranted and sound, this literature review has shown that it is possible to disagree with their statement on several counts. First, giving the highest available number of points to missing information is a property of the model to which Vall-Ilosera and Sol (2009) compare their own model. As noted, missing information can be dealt with in various ways. Second, the models were usually not qualitative but semi-quantitative, but can also be fully quantitative, although the information requirements in that case are naturally much higher. Third, the whole idea behind prioritisation is to incorporate scientific information into the design, and this certainly can be done through the framework. If ranking is not based on scientific information (when such is available), it is not a weakness of the approach but of the application. In fact, Hiebert and Stubbendieck (1993) noted that one reason for analytical ranking is to involve scientists in the process. If the framework used is consistent and logical, scientists can be involved without endangering scientific credibility in the face of uncertain information. This can be done either through an expert panel or individual evaluation (see WHO 2006 for advantages and disadvantages). Lacking an analytical model, the decisions would be based on an opinion of an individual or a group, or on what has been done before. History, Hiebert and Stubbendieck (1993) argued, is partially correct, but is not based on established criteria, its basis is not documented, and it cannot be used to ensure that all important aspects have been taken into account.

However, assessment of risks is an uncertain business with often imprecise or inadequate information. From the perspective of economic theory, the prioritisation models are not without problems. For instance, scoring scales are often non-linear, that is to say they are on an ordinal rather than interval scale. A score of 2 for a particular criterion may represent a species less harmful than a score of 4, but not necessarily a species that will inflict half as much damage. Hence direct score aggregation is challenging. For direct score aggregation to be theoretically sound, a change in score from, say, 1 to 2 should have the same effect on social

welfare as a change from, say, 4 to 5 (Ryan 2006). Moreover, a change of one point in score should have the same effect on welfare regardless of the criteria in which it occurs.

The correlation among different criteria or questions should be assessed, and hence it should be determined whether the same property is being counted several times (e.g. Horby et al. 2001). For instance, Gordon et al. (2008b) noted that in the Australian and Hawaii WRAs the total score reflects more the probability of entry of an organism than the impacts or the spread of the organism. However, such internal correlation may also be by design, if for instance some property can be evaluated by two measures, and information on one measure is more readily available for some organisms, and information on the other measure more readily available for other organisms (although in such a case it would make sense to pose these questions as alternatives to each other).

One issue is that rankings often do not link interventions to consequences. In other words, they assume nothing is done about the problem or that doing something has similar cost and implications for all hazards. If nothing is done to assess the interventions and the consequences of those interventions, allocating resources on the basis of the priority lists is misleading. Of course, this all depends on the base case to which the invasion is compared to: if nothing will be done regarding the organism in any case (do nothing scenario is the base case), excluding feasibility of control makes no difference – although in such a case also the total score of the organism would probably be fairly low. If something has to be done after the invasion (by the state or by private agents), as is the case for many human diseases or serious agricultural pests, then feasibility of control matters. There may also be instances where control is not such an important factor, for instance if it is specifically the entry potential that is being assessed. However, looking at the organisms holistically may give better results than assessing their potential element by element. For instance, if we have two species that are otherwise identical, but one of them is controllable post-border whereas the other one is not, and we have resources to add only one species to the

port exclusion list, it should be the one that is not controllable. In such a case looking at the entry potential alone would not provide an optimal solution – unless port exclusion can be relied to be perfectly effective.

A further issue related to control is the potential correlation in the outcome of control (Cox 2009). In other words, corrective control actions may simultaneously target a range of organisms – at the extreme it may be the case that controlling any one of them by itself is not socially beneficial, but since the same control deals with all of them, the control action becomes socially beneficial. Such joint production of control cannot be taken into account in prioritisation models.

Conclusions

Ranking and prioritisation are carried out every day. Cox (2009) has argued that priority scores can never do better than optimisation methods, which is probably true. However, equally true is that prioritisation models are likely to do better than unstructured individual opinions or other aspirations, as discussed earlier. The benefits of properly constructed prioritisation models are likely to reside in their structured, holistic and transparent mechanisms as well as their ease of application. The Australian WRA, for instance, is in official use in Australia and New Zealand, and has been used to rate over 2800 species in Australia (Weber et al. 2009) as well as been applied widely elsewhere in academic studies.

Cook and Proctor (2007) found that the list produced by their prioritisation model ended up with a very different distribution of resource use than what the existing funding priorities were. Resource reallocation is naturally difficult: plant health officials cannot be transformed into animal health officials overnight, but adaptation should occur over time according to the relative risks (Ryan 2006), and the current level of resourcing should be taken into account. As noted by Horby et al. (2001), an organism high in the ranking may already be well

resourced, whereas an organism ranked further down may be, relative to its position, very much under-resourced. Hence having an organism high in the ranking does not automatically mean that more resources should be invested in it than is currently the case.

Ranking tools do not provide an absolute hierarchy. Rather, they are a basis for decision-making and for detecting hazards that require attention. Prioritisation does not directly tell us how much to spend on each hazard. It is a way of thinking through the problem analytically and systematically in the face of uncertainty in order to achieve a better overall allocation of scarce societal resources. Ryan (2006) summarised the benefits of prioritisation as 1) more efficient resource allocation; 2) transparent basis for decision-making; 3) conceptualisation of the problem; and 4) a quantitative aid to decision-making when there are conflicting objectives that are measured in different units. Prioritisation tool should also be used correctly and by able operators (Hiebert and Stubbendieck 1993). Finally, prioritisation is not a static exercise, since the risk represented by the organisms varies in time and space, and prioritisation should be regularly reviewed. Prioritisation frameworks are designed to support decision-making, and when the entire framework is carefully designed, taking into account the challenges and best practices highlighted in this paper, they can help us in attaining a more efficient allocation of societal resources.

References

- Altman, D.G. & Bland, J.M. 1994. Diagnostic tests 1: sensitivity and specificity. *British Medical Journal* 308: 1552.
- Andreu, J. & Vilà, M. 2010. Risk analysis of potential invasive plants in Spain. *Journal for Nature Conservation* 18: 34–44.
- Branquart, E. (co-ordinator). 2007. Guidelines for environmental impact assessment and list classification of non-native organisms in Belgium. ISEIA Guidelines, Harmonia information system (version 2.5). 1–4.
- Buddenhagen, C.E., Chimera, C. & Clifford, P. 2009. Assessing biofuel crop invasiveness: a case study. *PLoS ONE* 4 (4): e5261
- Caley, P., Lonsdale, W.M. & Pheloung, P.C. 2006. Quantifying uncertainty in predictions of invasiveness, with emphasis on weed risk assessment. *Biological Invasions* 8: 1595–1604.
- Champion, P.D. & Clayton, J.S. 2000. Border control for potential aquatic weeds - Stage 1: Weed risk model. *Science for Conservation* 141. Department of Conservation. Wellington, New Zealand.
- Ciotti, M. 2003. Establishing a list of priority diseases for the countries of the Dubrovnik Pledge. In: WHO. 2003. *The Dubrovnik pledge on surveillance and prioritization of infectious diseases*. Report on a WHO Meeting. Bucharest, Romania. 21–23 November 2002. WHO Regional Office for Europe, Copenhagen.
- Cook, D. & Proctor, W. 2007. Assessing the threat of exotic plant pests. *Ecological Economics* 63: 594–604.
- Copp, G.H., Garthwaite, R. & Gozlan, R.E. 2005. Risk identification and assessment of non-native freshwater fishes: concepts and perspectives on protocols for the UK. *Cefas Science Series*. Technical Report no. 129.
- Copp, G.H., Villizzi, L., Mumford, J., Fenwick, G.V., Godard, M.J. & Gozlan, R.E. 2009. Calibration of FISK, an invasiveness screening tool for nonnative freshwater fishes. *Risk Analysis* 29: 457–467.
- Covello, V.T. & Merkhofer, M.W. 1993. Risk assessment methods: approaches for assessing health and environmental risks. Plenum Press. New York, United States. 319 p.
- Cox, L.A. Jr. 2009. What's wrong with hazard-ranking systems? An expository note. *Risk Analysis* 29: 940–948.
- Crosti, R., Cascone, C. & Cipollaro, S. 2010. Use of a weed risk assessment for the Mediterranean region of Central Italy to prevent loss of functionality and biodiversity in agro-ecosystems. *Biological Invasions* 12: 1607–1616.
- Cunningham, D.C., Barry, S.C., Woldendorp, G. & Burgess, M.B. 2004. A framework for prioritizing sleeper weeds for eradication. *Weed Technology* 18: 1189–1193.
- Daehler, C.C. & Carino, D.A. 2000. Predicting invasive plants: prospects for a general screening system based on current regional models. *Biological Invasions* 2: 93–102.
- Daehler, C.C., Denslow, J.S., Ansari, S. & Kuo, H.-C. 2004. A risk-assessment system for screening out invasive pest plants from Hawaii and other Pacific Islands. *Conservation Biology* 18: 360–368.
- Darin, G.S. 2008. *Prioritizing weed populations for eradication at a regional level: the California Department of Food and Agriculture's A-rated weeds*. Master of Science Thesis. University of California, Davis.
- Dawes, J. 2008. Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *International Journal of Market Research* 50: 61–77.
- Dawson, W., Burslem, D.F.R.P. & Hulme, P.E. 2009. The suitability of weed risk assessment as a conservation tool to identify invasive plant threats in East African rainforests. *Biological Conservation* 142: 1018–1024.
- DEFRA. 2003. Partnership, priorities and professionalism: a strategy for enhancing veterinary surveillance in the UK. Department for Environment, Food and Rural Affairs. London.
- DEFRA. 2006. Approaches to the prioritisation of resources: a brief review of selected public sector organisations in the UK and abroad. Department for Environment, Food

- and Rural Affairs. London.
- Doherty, J.-A. 2006. Final report and recommendations from the national notifiable diseases working group. Canada Communicable Disease Report. 1 October 2006. 32: 211–25.
- Doherty, J.-A. 2000. Establishing priorities for national communicable disease surveillance. *Canadian Journal of Infectious Diseases* 11: 21–24.
- European Commission. 2007. *A new animal health strategy for the European Union (2007-2013) where "prevention is better than cure"*. Office for Official Publications of the European Communities. Luxembourg. http://ec.europa.eu/food/animal/diseases/strategy/docs/animal_health_strategy_en.pdf
- Fox, A.M. & Gordon, D.R. 2004. Criteria for listing invasive plants. *Weed Technology* 18: 1309–1313.
- Garry Oak Ecosystems Recovery Team. 2007. General decision process for managing invasive plant species in Garry Oak and associated ecosystems (GOEs). http://www.goert.ca/documents/General_Decision_Process_revised.pdf
- Gassó, N., Basnou, C. & Vilà, M. 2010. Predicting plant invaders in the Mediterranean through a weed risk assessment system. *Biological Invasions* 12: 463–476
- Gibbens, J., Houston, C., Gibbons, L., Williams, K., Lawton, S., Kelly, L. & Lysons, R. 2006. *Prioritisation of animal health and welfare issues for government intervention*. Presentation at the 11th Conference of the International Society for Veterinary Epidemiology and Economics (ISVEE 11). 6-11 August 2006, Cairns, Australia. <http://www.defra.gov.uk/foodfarm/farmanimal/diseases/vet-surveillance/documents/isvee-prioritisation.pdf>
- Gordon, D.R., Onderdonk, D.A., Fox, A.M. & Stocker, R.K. 2008a. Consistent accuracy of the Australian weed risk assessment system across varied geographies. *Diversity and Distributions* 14: 234–242.
- Gordon, D.R., Onderdonk, D.A., Fox, A.M., Stocker, R.K. & Gantz, C. 2008b. Predicting invasive plants in Florida using the Australian Weed Risk Assessment. *Invasive Plant Science and Management* 1: 178–195.
- Gordon, D.R., Mitterdorfer, B., Pheloung, P.C., Ansari, S., Buddenhagen, C., Chimera, C., Daehler, C.C., Dawson, W., Denslow, J.S., LaRosa, A., Nishida, T., Onderdonk, D.A., Panetta, F.D., Pyšek, P., Randall, R.P., Richardson, D.M., Tshidada, N.J., Virtue, J.G. & Williams, P.A. 2010. Guidance for addressing the Australian Weed Risk Assessment questions. *Plant Protection Quarterly* 25: 56–74.
- Hayes, K.R. & Sliwa, C. 2003. Identifying potential marine pests - a deductive approach applied to Australia. *Marine Pollution Bulletin* 46: 91–98.
- Hiebert, R.D. & Stubbendieck, J. 1993. Handbook for ranking exotic plants for management and control. Natural Resources Report NPS/NRMWRO/NRR-93/08. United States Department of the Interior, National Park Service, Midwest Regional Office. Denver, Colorado.
- Holt, J. 2006. Score averaging for alien species risk assessment: A probabilistic alternative. *Journal of Environmental Management* 81: 58–62.
- Horby, P., Rushdy, A., Graham, C. & O'Mahony, M. 2001. PHLS overview of communicable diseases 1999. *Communicable Disease and Public Health* 4: 8–17.
- Hughes, G. & Madden, L.V. 2003. Evaluating predictive models with application in regulatory policy for invasive weeds. *Agricultural Systems* 76: 755–774.
- Jefferson, L., Havens, K. & Ault, J. 2004. Implementing invasive screening procedures: the Chicago Botanic Garden Model. *Weed Technology* 18: 1434–1439.
- Kato, H., Hata, K., Yamamoto, H. & Yoshioka, T. 2006. Effectiveness of the weed risk assessment system for the Bonin Islands. In: Koike, F., Clout, M.N., Kawamichi, M., De Poorter, M. & Iwatsuki, K. (eds). 2006. *Assessment and control of biological invasions risks*. Kyoto, Japan: Shoukadoh Book Sellers and Gland, Switzerland: IUCN. p. 65–72.
- Krause, G. and the working group on prioritisation at the Robert Koch Institute. 2008. Prioritisation of infectious diseases in public health - call for comments. *Eurosurveillance* 13: 1–6
- Křivánek, M. & Pyšek, P. 2006. Predicting invasions by woody species in a temperate zone: a test of three risk assessment schemes in the Czech Republic (Central Europe). *Diversity and Distributions* 12: 319–327
- MacLeod, A. & Baker, R.H.A. 2003. The EPPO pest risk assessment scheme: assigning descriptions to scores for the questions on entry and establishment. *EPPO/OEPP Bulletin* 33: 313–320.
- Makowski, D. & Mittinty, M.N. 2010. Comparison of scoring systems for invasive pests using ROC analysis and Monte Carlo simulation. *Risk Analysis* 30 (6): 906–915.
- Mataragas, M., Skandamis, P.N. & Drosinos, E.H. 2008. Risk profiles of pork and poultry meat and risk ratings of various pathogen/product combinations. *International Journal of Food Microbiology* 126: 1–12.
- McKenzie, J., Simpson, H. & Langstaff, I. 2007. Development of methodology to prioritise wildlife pathogens for surveillance. *Preventive Veterinary Medicine* 81: 194–210.
- Morgan, M.G., Florig, H.K., DeKay, M.L. & Fischbeck, P.S. 2000. Categorizing risks for risk ranking. *Risk Analysis* 20: 49–58.
- More, S.J., McKenzie, K., O'Flaherty, J., Doherty, M.L., Cromie, A.R. & Magan, M.J. 2010. Setting priorities for non-regulatory animal health in Ireland: results from an expert policy Delphi study and a farmer priority identification survey. *Preventive Veterinary Medicine* 95: 198–207.
- Morse, L.E., Randall, J.M., Benton, N., Hiebert, R. & Lu, S. 2004. An invasive species assessment protocol: evaluating non-native plants for their impacts on biodiversity. Version 1. NatureServe. Arlington, Virginia, USA.
- Nishida, T., Yamashita, N., Asai, M., Kurokawa, S., Enomoto, T., Pheloung, P.C. & Groves, R.H. 2009. Developing a pre-entry weed risk assessment system for use in Japan. *Biological Invasions* 11: 1319–1333
- NWRAS Review Group. 2006. Review of the National (Australia) Weed Risk Assessment System. Weeds Australia.
- Ou, J., Lu, C. & O'Toole, D.K. 2008. A risk assessment system for alien plant bio-invasion in Xiamen, China. *Journal of Environmental Sciences* 20: 989–997.
- Parker, C., Caton, B.P. & Fowler, L. 2007. Ranking nonindigenous weed species by their potential to invade the United States. *Weed Science* 55: 386–397.
- Petersen, K.E., James, W.O., Thaler, A.M., Ragland, R.D. & Hogue, A.T. 1996. Use of a priority rating process to sort meatborne zoonotic agents in beef. *Journal of Agromedicine* 3: 17–36.

- Pheloung, P.C. 1995. *Determining the weed potential of new plant introductions to Australia*. A report to the Standing Committee on Agriculture and Resource Management, Australia.
- Pheloung, P.C., Williams, P.A. & Halloy, S.R. 1999. A weed risk assessment model for use as a biosecurity tool evaluating plant introductions. *Journal of Environmental Management* 57: 239–251.
- Randall, J.M., Morse, L.E., Benton, N., Hiebert, R., Lu, S. & Killefer, T. 2008. The invasive species assessment protocol: a tool for creating regional and national lists of invasive nonnative plants that negatively impact on biodiversity. *Invasive Plant Science and Management* 1: 36–49.
- Reichard, S.H. & Hamilton, C.W. 1997. Predicting invasions of woody plants introduced into North America. *Conservation Biology* 11: 193–203
- Risk Assessment and Management Committee. 1996. Generic nonindigenous aquatic organisms risk analysis review process (for estimating risk associated with the introduction of nonindigenous aquatic organisms and how to manage for that risk). Report to the Aquatic Nuisance Species Taskforce.
- Rosengren, H. & Heikkilä, J. 2009. Risk-based classification of animal diseases: the case of poultry. Poster at 12th International Symposium on Veterinary Epidemiology and Economics (ISVEE XII). 10–14 August 2009, Durban, South Africa.
- Ross, T. & Sumner, J. 2002. A simple, spreadsheet-based, food safety risk assessment tool. *International Journal of Food Microbiology* 77: 39–53.
- Ryan, M. 2006. *Allocating Vote: Biosecurity - towards an 'economics-based' approach for setting priorities for the importation of goods*. Presentation at New Zealand Agricultural and Resource Economics Society Conference. 25–27 August 2006. <http://ageconsearch.umn.edu/bitstream/31949/1/cp06ry01.pdf>
- Smallwood, K.S. & Salmon, T.P. 1992. A rating system for potential exotic bird and mammal pests. *Biological Conservation* 62: 149–159.
- Sumner, J. & Ross, T. 2002. A semi-quantitative seafood safety risk assessment. *International Journal of Food Microbiology* 77: 55–59.
- Sumner, J., Ross, T., Jenson, I. & Pointon, A. 2005. A risk microbiological profile of the Australian red meat industry: risk ratings of hazard-product pairings. *International Journal of Food Microbiology* 105: 221–232.
- SZEID. 2006. Documentation for prototype AHW prioritisation decision support tool. V.1.1.14 p.
- Tricarico, E., Vilizzi, L., Gherardi, F. & Copp, G.H. 2010. Calibration of FI-ISK, an invasiveness screening tool for nonnative freshwater invertebrates. *Risk Analysis* 30: 285–292.
- Vall-Ilosera, M. & Sol, D. 2009. A global risk assessment for the success of bird introductions. *Journal of Applied Ecology* 46: 787–795.
- Virtue, J. 2007. Weed risk management in South Australia. Presentation at 2nd International Weed Risk Assessment Workshop (IWRAW). Perth (WA), Australia, 14–15 September 2007. <http://www.hear.org/iwraw/2007/presentations/iwraw2007virtue.pdf>
- Ward, D.F., Stanley, M.C., Toft, R.J., Forgie, S.A. & Harris, R.J. 2008. Assessing the risk of invasive ants: a simple and flexible scorecard approach. *Insectes Sociaux* 55: 360–363.
- Warner, P.J., Bossard, C.C., Brooks, M.L., DiTomaso, J.M., Hall, J.A., Howald, A.M., Johnson, D.W., Randall, J.M., Roye, C.L., Ryan, M.M. & Stanton, A.E. 2003. *Criteria for categorizing invasive non-native plants that threaten wildlands*. <http://www.cal-ipc.org/ip/inventory/pdf/Criteria.pdf>. California Exotic Pest Plant Council and Southwest Vegetation Management Association. 24 p.
- Weber, E. & Gut, D. 2005. A survey of weeds that are increasingly spreading in Europe. *Agronomy for Sustainable Development* 25: 109–121.
- Weber, E. & Gut, D. 2004. Assessing the risk of potentially invasive plant species in Central Europe. *Journal for Nature Conservation* 12: 171–179.
- Weber, J., Panetta, F.D., Virtue, J. & Pheloung, P. 2009. An analysis of assessment outcomes from eight years' operation of the Australian border weed risk assessment system *Journal of Environmental Management* 90: 798–807.
- Weinberg, J., Grimaud, O. & Newton, L. on behalf of the charter group. 1999. Establishing priorities for European collaboration in communicable disease surveillance. *European Journal of Public Health* 9: 236–240.
- Widrechner, M.P., Thompson, J.R., Iles, J.K. & Dixon, P.M. 2004. Models for predicting the risk of naturalization of non-native woody plants in Iowa. *Journal of Environmental Horticulture* 22: 23–31.
- WHO. 2006. Setting priorities in communicable disease surveillance. WHO/CDS/EPR/LYO/2006.3. World Health Organization.