
# BIODIVERSITY SECTOR

# ECOLOGICAL IMPLICATIONS OF GMOS

## Robust methodologies for ecological risk assessment

## Best practice and current practice in ecological risk assessment for Genetically Modified Organisms

**Keith R. Hayes**[1]

[1] Centre for Research on Introduced Marine Pests, CSIRO Division of Marine Research, GPO Box 1538, Hobart, Tasmania, 7001, Australia. Tel: (+61) 3 6232 5260, Fax: (+61) 3 6232 5485, Email: keith.hayes@marine.csiro.au

## Executive summary

This report compares current practice in ecological risk assessment for genetically modified (GM) plants and microorganisms, as evidenced by eight transnational and national frameworks, with what might reasonably be considered best practice. Best practice is defined for the scientific principles, hazard identification, risk calculation, social appraisal and monitoring stages of an ideal ecological risk assessment, and summarised in the following ten points:

1. Carefully define measurement and assessment endpoints for environmental values for each stage of a genetically modified organism (GMO) release;

2. Construct good qualitative models of all hazard scenarios using structured deductive and inductive hazard assessment techniques;

3. Consider the influence of cognitive bias, framing effects, anchoring and sample size on qualitative decisions;

4. Consider the full spectrum of ecological models from simple (screening) to detailed ecosystem models;

5. Recognise that even simple models can incorporate uncertainty and be useful in ecological risk assessment;

6. It is essential to include a transparent analysis of uncertainty;

7. When information is sparse use probability bounding analysis to express uncertainty;

8. Examine opportunities to promote appropriate and ongoing stake-holder participation in the risk assessment;

9. Adopt a precautionary approach to high consequence, but highly uncertain, hazards; and,

10. Consider statistical power, effect size and model based sensitivity analysis and other remedies to hidden conventional pitfalls in monitoring.

Most of the frameworks reviewed here provide evidence of best practice in the scientific principles and frameworks of ecological risk assessment. All of them, however, rely on simple checklists in the hazard identification stage, and only one discusses inductive techniques. Hazard identification as currently practiced is therefore largely restricted to prescriptive deductive techniques. Analysts will identify a larger range of potential hazards, and gain a better understanding of the event chains associated with these hazards, if they used inductive hazard identifications techniques.

Well-corroborated quantitative techniques exist for some of the potential hazards associated with GMO field release. However, despite the rich scientific literature on quantitative techniques and models, only one framework bridges the divide between science and regulation by identifying specific experimental techniques and models in the regulatory process. Some of the regulatory frameworks recognise that quantitative approaches are possible in certain circumstances, but neither the circumstances (i.e. which hazards) nor available techniques are identified. For the main part it is not clear when and how quantitative techniques are expected of the applicant.

Regulators can assist quantitative risk assessment by helping proponents identify models and analysis techniques relative to specific GMO hazards. Regulators should insist that proponents obtain the necessary data and information in order to achieve best practice and to reduce areas of significant uncertainty. Current field trials only appear to gather information on crop performance. These trials are an ideal opportunity to gather the types of data needed to improve the science of GMO risk assessment.

Quantitative techniques are not currently available for all of the potential hazards associated with GMOs. There are important gaps in the following areas: food-wed and trophic interactions, the transfer of viral genetic material to other viruses, increases in the host range of viruses, fungi and other pathogens, altered farm practice and physical habitat changes. National regulatory authorities should encourage data collection and research in these areas. High consequence, high uncertainty impacts (such as the creation of new viruses) are unlikely to be satisfactorily addressed by quantitative techniques in the near future. More rigorous qualitative techniques, however, including a wider social discourse and directed research, are achievable in the near term.

The degree of practicality, reliability and acceptance of quantitative techniques for less uncertain hazard scenarios varies from model to model. In general terms simple models are the most widely accepted and, when used in conjunction with a rigorous analysis of uncertainty, can provide meaningful answers for risk assessment purposes. Qualitative assessments are often recommended as an initial screen to eliminate low risk events from a potentially lengthy assessment process. This review, however, suggests quite the opposite: simple quantitative techniques should be used wherever possible to screen high and low risk scenarios—qualitative assessments become most important for highly uncertain but potentially high impact scenarios. To be successful these qualitative assessment should include a strong element of social appraisal including, for example, the use of systematic hazard identification techniques to capture the imagination and intuition of non-scientific ‘experts’.

Uncertainty analysis is the very rationale of risk assessment, and yet this is by far the weakest component of current practice. None of the frameworks reviewed here, bar two, require a formal analysis of uncertainty as part of the risk assessment process. This is arguably the biggest gap between current practice and best practice in ecological risk assessment for GM plants and microorganisms. Well-established statistical techniques exist to describe random measurement error and environmental variability. Model error can be approached by ground-truthing risk assessment predictions and testing alternative model formulations. Techniques also exist that bridge the divide between qualitative and quantitative approaches to risk assessment, and thereby facilitate a progression from one to the other.

All of the frameworks reviewed here discuss or at least mention monitoring but none point to best practice in this area. All of the frameworks could be improved by drawing the analyst’s attention to power calculations for typical monitoring strategies. Monitoring strategies will need to continue well beyond the usual period needed to assess the efficacy of the phenotype in order to detect potential ecological impacts. It is important that these strategies test the predictions of prior risk assessments and provide information that will inform future risk assessments, thereby “closing the regulatory loop”. These strategies must explicitly include an appropriate power analysis to avoid blindness to Type II error.

# Contents

# 1 Introduction

## *1.1 Genetically Modified Organisms*

Field releases of Genetically Modified Organisms (GMOs) in member nations of the OECD have grown enormously since the first field trial was held in 1986 (OECD, 1993). Field trial approvals almost doubled every year between 1988 and 1994, peaking in 1998 at 2,312 (Figure 1). The apparent decline in 1999 and 2000 is not real but due to under-reporting (*pers comm*. Tetsuya Maekawa, OECD, December 2001). The total acreage of GM crops and trials around the world grew from 1.7 million hectares in 1996 to 44.5 million hectares in 2000, mostly in the USA (68%), Argentina (23%) and Canada (8%) (Pretty, 2001), to 52.6m hectares in 2001 (Financial Times, 11th January, 2002).

To date, the vast majority of releases (98.4% in the OECD) involve GM plants. Bacteria, fungi, viruses and animals account for only 1.0%, 0.1%, 0.3% and 0.2% of releases respectively (OECD, 2000a). Around the world the principal GM crops are soybean, corn and cotton. Most GM crops contain a single transgene (and sometimes a selectable marker) that modifies the plant for herbicide tolerance and or insect resistance. Less common are GM plants with traits that are expected to influence virus resistance, crop quality, male sterility and disease resistance (OECD, 1992). .

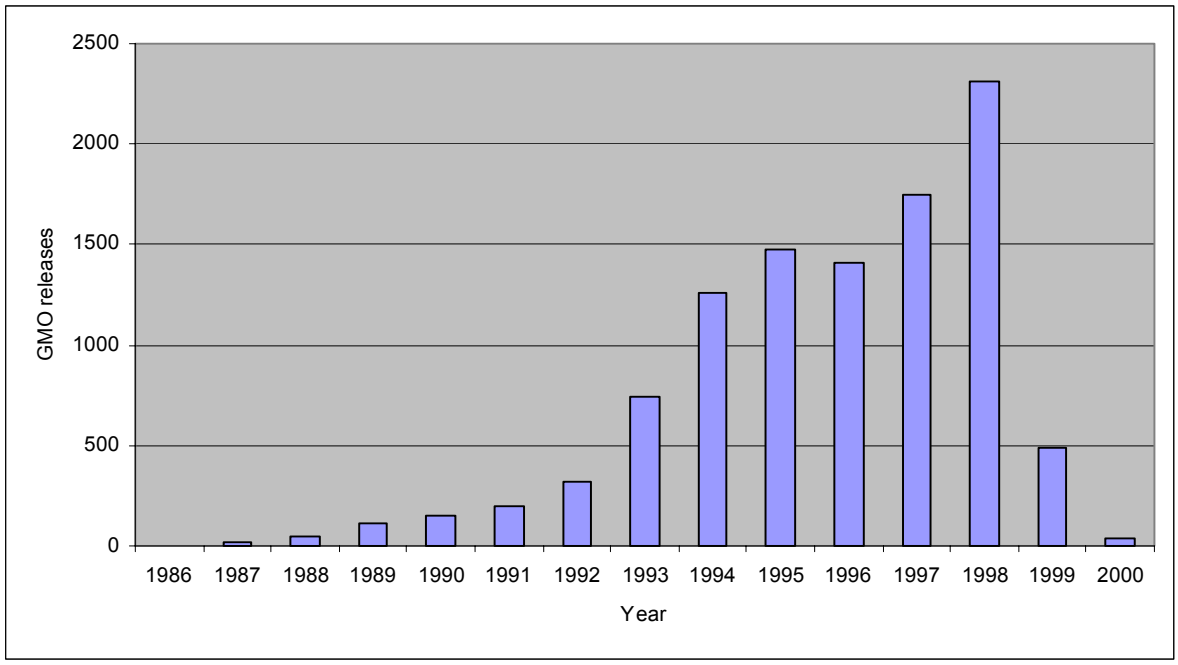
GM plants and microbes represent the first generation of GMOs to be released and tested across the world. GM fish are likely to be the next generation. To date, 28 species of fish have been successfully engineered for traits such as improved growth and cold tolerance (Royal Society of Canada, 2001). The first application for the commercial production of a transgenic fish (growth enhanced Atlantic salmon) occurred in the United States in early 2000. Shellfish, aquatic plants and farm animals are waiting in the wings.

Most developed nations producing or releasing transgenic products have, or are developing, systems of regulatory oversight based on risk assessment. For example, the United States, Canada, New Zealand, Australia and member nations of the European Union have enacted legislation that requires an assessment of the ecological and human-health risks associated with the contained use and unconfined release of GMOs.

Some of these regulatory systems call for “state of the art” risk assessment, or the use of “best practice” without defining what this means relative to environmental risks. This report aims to:

- identify what is “best practice” and “state of the art” relative to ecological risk assessment for GMOs;
- review the ecological risk assessment frameworks developed by various national and transnational organisations to manage GMOs; and,
- thereby compare current practice with best practice.

Figure 1 The number of approvals for field trials of GMOs in member nations of the Organisation for Economic Cooperation and Development (OECD) between 1986 and 2000.




## *1.2 Terms of reference*

Most approaches to GMO risk assessment distinguish between confined and unconfined releases. The former usually refer to research and development within an enclosed laboratory or greenhouse, although references can be found in the literature to "confined field trials" (Canadian Food Inspection Agency, 2001b). The scope of this review is limited to risk assessment frameworks applicable to unconfined release. In this report unconfined is interpreted to mean any release to the outside environment via experimental field trials or full commercial release.

Table 1 lists fifteen risk assessment frameworks that have been developed or advocated for the unconfined release of GMOs. This paper formally evaluates eight of these that have been recognised or referenced by a national regulatory authority in the last 10 years. These eight frameworks are summarised in Appendix A. Table 1 lists five "academic" risk assessment frameworks proposed by individuals or research organisations. These are not included in this review because none are explicitly recognised and/or referenced by a national regulatory authority. The frameworks developed by the US National Research Council (NRC) and Office of Technological Assessment (OTA) are also excluded because they have been reviewed elsewhere (Hayes, 1997) and both are over ten years old.

The scientific principles of risk assessment and hazard identification, and the methods of uncertainty analysis that are discussed here are relevant to all GMOs, and indeed to all ecological stressors. The specific models and methods for likelihood and consequence analysis, however, are restricted to GM plants and microbes for the sake of brevity and because to date these are the only GMOs that have actually been released.

Table 1 Ecological risk assessment frameworks for the unconfined releases of GMOs

| NAME | DESCRIPTION | REFERENCE |
|---|---|---|
| Annex III of the Cartagena Protocol on Biosafety | Transnational - CBD | Secretariat of the Convention on Biological Diversity, 2000 |
| Directive 2001/18/EC on the deliberate release into the environment of GMOs | Transnational - EU | European Union, 2001 |
| International technical guidelines for safety in biotechnology | Transnational - UNEP | United Nations Environment Programme, 1995 |
| Safety considerations for biotechnology | Transnational - OECD | Organisation for Economic Cooperation and Development, 1986; 1992 |
| Canadian Food Inspection Agency decision framework for GM plants | National – Canada - Plants | Canadian Food Inspection Agency, 2001a; 2001b |
| Guidance on principles of risk assessment and monitoring for the release of GMOs | National - United Kingdom | Department of Environment, Transport and Regions, 1999 |
| Office of the Gene Technology Regulator - Risk analysis framework | National - Australia | Office of the Gene Technology Regulator, 2001 |
| Identifying risks for applications under the HSNO Act 1996 | National - New Zealand | Environmental Risk Management Authority, 1999; 2000 |
| NRC risk assessment framework for GM plants and microorganisms | National - United States | National Research Council, 1989a; 1989b |
| OTA risk assessment framework for environmental introductions | National - United States | Fiksel and Covello, 1985 |
| Decision support system for safely conducting research with GM fish and shellfish | Academic - Fish | Hallerman et al., 1999 |
| Two stage safety evaluation for GMO release | Academic | Kappeli and Auberson, 1998 |
| Manual for assessing ecological and human health effects of GMOs | Academic | Scientists Working Group on Biosafety, 1998 |
| 3 stage risk assessment framework | Academic | Rissler and Melon, 2000 |
| Risk assessment for the release of biotechnology products | Academic | Strauss, 1991 |

## 1.3 Risks, hazards and quantitative risk assessment

Risk is not difficult to define, but is usually presented in specific contexts, leading to a plethora of definitions in the scientific literature. Traditional (engineering) risk is defined as

$$Risk = \frac{Event}{Time} \times \frac{Consequences}{Event} = \frac{Consequences}{Time} \quad , \qquad [1]$$

leading to risk functions that describe accidental events in terms of the frequency of consequences. These consequences usually refer to human injuries or fatalities (the risk assessment endpoints). Thus, an engineering risk assessment might express the risk of an activity as 0.01 fatalities per annum, or a 1 in 100 chance of dying each year due to the activity in question. Kaplan (1997) emphasises that risk is defined not as a number, a curve, or a vector, but by three questions: What can happen? How likely is that to happen? If it does happen, what are the consequences? The answer to these questions constitutes a triplet $[S_i, L_i, X_i]$ where $S_i$ denotes individual risk scenarios, $L_i$ denotes the likelihood of the this scenario and $X_i$ the consequences of this scenario. This framework includes an implicit time horizon within which the risk scenarios are evaluated.

Uncertainty about the likelihood of risk scenarios, together with uncertainty about their consequences, means that these components should be expressed in probabilistic terms, denoted $[S_i, p(\varphi_i), p(X_i)]$. The definition of risk is completed by identifying the complete set of possible risk scenarios (c) such that:

$$Risk = \left\{ \left\langle S_i, p(\varphi_i), p(X_i) \right\rangle \right\}_c \quad . \qquad [2]$$

This approach encourages a broader interpretation of risk that is better suited to ecological risk assessment, where the events in question may not be 'accidental' in any sense, nor the endpoints restricted to human fatality or injury.

It is important to recognise that any individual's interpretation of risk is intimately linked to the assessment endpoint (defined below), which in itself is simply an expression of value (Hayes, 1997). Stakeholders, scientists and regulators usually hold different values and will therefore understand and perceive the significance of risks, particularly ecological risks, very differently (Trevors et al., 1994). Consequently, decisions about the acceptability of risk should be taken out of the risk assessment process and made part of a wider socio-economic debate—the analyst should not define risk in terms of what is acceptable or unacceptable during the risk assessment. Instead, the risk should be assessed, together with its associated uncertainty, and then the acceptability, or otherwise, should be gauged by stakeholders and those who will bear the consequences. This does not, of course, exclude the risk analyst from making recommendations (e.g. accept or reject) but these should not form the only basis on which risk decisions are made.

Hazard can be defined as a situation that in particular circumstances could lead to harm (The Royal Society, 1983) or as a substance's or activity's propensity for harm. Hazard is often perceived as a function of a substance's intrinsic properties but, as emphasised in the definition above, it is more usefully conceptualised as a function of both the intrinsic properties of a substance and circumstance. A simple if somewhat contrived example will make this clear. Oxygen in air would not ordinarily be considered as a hazardous substance, but when compressed with air and used by divers at depth, it can be poisonous.

Thus, a substance's intrinsically hazardous properties can often only be realised under a very specific set of circumstances. Any expression of hazard should properly acknowledge both the intrinsic

properties and the circumstances required in order for harm to be realised. These circumstances are often referred to as the conditions of exposure or exposure pathways (Environmental Risk Management Authority, 2000) and usually embody a conceptual model about how an ecological system works. Risk is simply a measure of the likelihood of these circumstances and the magnitude of the subsequent harm. Put another way, hazard becomes risk only when there is a finite probability of a manifestation of the hazard (Beer and Ziolkowski, 1995).

Hazard assessment must address the substance's intrinsic properties and the circumstances required for the manifestation of harm as a result of these properties. This is particularly true for biological stressors (including GMOs) because ecological risks depend on the characteristics of the organism (its intrinsic properties) and when, where and how it is introduced into the environment (the circumstances).

Risk assessment is a general term that is used (often loosely) to describe an array of methodologies and techniques concerned with estimating the likelihood and consequences of undesired events.[2] Risk assessments can be qualitative, semi-quantitative or quantitative, and can be a valuable decision aid if completed in a systematic and rigorous manner.

There is no universally applicable procedure for conducting ecological risk assessment despite considerable efforts by a number of national agencies, most notably the United States Environmental Protection Agency, to create one. The subject area is characterised by a multiplicity of techniques and methods. This is due in part to the relative immaturity of ecological risk assessment as a discipline, but also in part to the complexity of environmental management issues, the variety of possible stressors and endpoints, and thus the widely different types of assessment that are required. Indeed, it would be surprising if a single procedural framework could be developed to cover every conceivable application of ecological risk assessment.

For the purposes of this report, Quantitative Risk Assessment (QRA) is defined as the quantitative evaluation of the likelihood of undesired events and the likelihood of harm or damage being caused as a result of these events, together with value judgements concerning the significance of the results. Again harm or damage is expressed through the assessment endpoints. QRA is a five-stage procedure consisting of:

1. Hazard identification—what can go wrong (identify the events) and why;
2. Frequency analysis—how often do the events go wrong (events/time);
3. Consequence analysis—how much harm is caused by the event (consequences/event);
4. Risk calculation—frequency x consequence; and,
5. Uncertainty and significance analysis—how sure are we of the risk estimate and how important is this type of risk.

[2] In the United States, the term risk assessment refers to the actual calculation of likelihood and consequence, whilst risk analysis describes the wider process including risk management, risk perception, etc. By contrast, in Australia, risk analysis is widely used to describe the calculation component, whilst risk assessment is understood to be the wider process. These semantic differences are unnecessary and serve only to confuse the issues. Here I use the terms interchangeably to mean a process that includes all of the components identified in Figure 2.

Borrowing from this, quantitative ecological risk assessment might be defined as the quantitative evaluation of the frequency and consequences (expressed in terms of environmental harm) of undesired events, together with value judgements concerning the significance of these events. This definition, however, has its own difficulties. While we may at first sight empathise with the meaning of 'harm to the environment', and the desirability of avoiding it, in practice it is much more difficult to delineate and quantify this phenomenon, let alone value it or define what constitutes acceptable harm.

## *1.4 Ecological endpoints*

Endpoints are an expression of the values that the analyst is trying to protect by undertaking the risk assessment. They distinguish ecological risk assessment (ecological endpoints) from human health risk assessment (human fatality or injury endpoints). Risk analysts often distinguish between assessment endpoints—what they are trying to protect—and measurement endpoints—what they can actually measure, extrapolating from one to the other for the purposes of the risk assessment. For example, a regulator may wish to protect beneficial insects from cotton modified to express one of the many *Bacillus thuringiensis* (Bt) toxins. The analyst, however, cannot assess all beneficial insects and may therefore measure the toxicity of Bt cotton on a select group of non-target insects that, for example, inhabit or feed at the location where the toxin is expressed.

Ecological endpoints are most easily expressed in terms of impacts on species—reducing the abundance of commercially valuable or endangered species, for example, or increasing the abundance of weeds. Endpoints can be expressed, however, at various levels of biological organisation—from the individual to the landscape—and can include impacts on species that are of no direct value to man, or impacts on fundamental ecosystem processes (Asian Development Bank, 1990; Suter, 1990).

Table 2 lists assessment and measurement endpoints that are typically used in ecological risk assessment. It is important to recognise that selecting assessment and measurement endpoints in ecological systems is not a trivial process because of the complexity of these systems and the large number of potential candidates. Assessment endpoints must therefore be chosen carefully and, ideally, should:

- be biologically relevant;
- have an unambiguous operational definition;
- be accessible to prediction and measurement; and,
- be exposed to the hazard(s) (Barnthouse et al., 1986; Suter, 1993).

In many instances, however, the analyst may not know what to protect or measure, and indeed may be relying on the risk assessment process to identify important endpoints. This is possible so long as rigorous and systematic hazard identification techniques are used during the risk assessment process (see below).

**Best practice recommendation #1: Carefully define measurement and assessment endpoints for environmental values for each stage of a GMO release.**

Table 2 Possible assessment and measurement endpoints for ecological risk assessment

| ASSESSMENT | MEASUREMENT |
|---|---|
| Individuals | |
| Change in metabolism | Respiration rate, assimilation efficiency |
| Change in behaviour | |
| Inhibition or induction of enzymes | Liver enzymes |
| Increased susceptibility to pathogens | Frequency of individual morbidity |
| Decreased growth | Age/weight ratio |
| Death | |
| Populations | |
| Decreased genotypic and phenotypic diversity | Occurrence |
| Decreased biomass | Population size |
| Increased mortality rate | Population size |
| Decreased fecundity | Age structure |
| Decreased recruitment of juveniles | Age structure |
| Increased frequency of disease | Population morbidity |
| Decrease yield | Biomass |
| Decreased growth rate | Age/weight ratio |
| Increased abundance of harmful organisms | Frequency of blooms or pest outbreaks |
| Species | |
| Commercial extinction of species | Yield/production, CPUE |
| Actual extinction of any species | Numbers/density |
| Creation of new harmful species (virus) | Occurrence |
| Communities | |
| Decreased biodiversity | Diversity indices |
| Decreased food web diversity | Species diversity |
| Decreased productivity | Species evenness |
| Ecosystem | |
| Decreased community diversity | Diversity indices |
| Altered bio- and geo-chemical cycles | Salinity, carbon, nitrogen, phosphorus flux |
| Loss of rare or unique ecosystems | Extent and area |
| Landscape | |
| Physical processes (floods, fires, erosion, flows) | Frequency of floods, fires, low flows |
| Resource quality (air, water, soil) | Pollutant concentrations |

# 2 Ideal risk assessment—principles, frameworks and methods

## *2.1 Scientific principles and framework*

The ecology of GMOs is identical to the ecology of any other organism (Crawley, 1990). It is accepted practice to examine the construct, i.e. the modified organism, on a case-by-case basis, rather than the construction method (OECD, 2000b). The risk assessment should be rigorous, systematic, repeatable and transparent. Important subjective judgements take place during a risk assessment—for example, when defining the scope, the choice of models, the degree of caution exercised when handling uncertainty and the acceptable standards of proof. Each of these judgements should be clearly stated and justified.

GMOs represent a new technology. Risk assessments for a new technology must initially be inductive, particularly in the hazard identification stage. As experience with the technology grows, deductive techniques play an increasingly important role in the analysis. Risk assessment for GMOs should therefore demonstrate a mixture of inductive and deductive techniques. The depth of experience with the particular organism in each environmental setting should determine the balance of techniques.

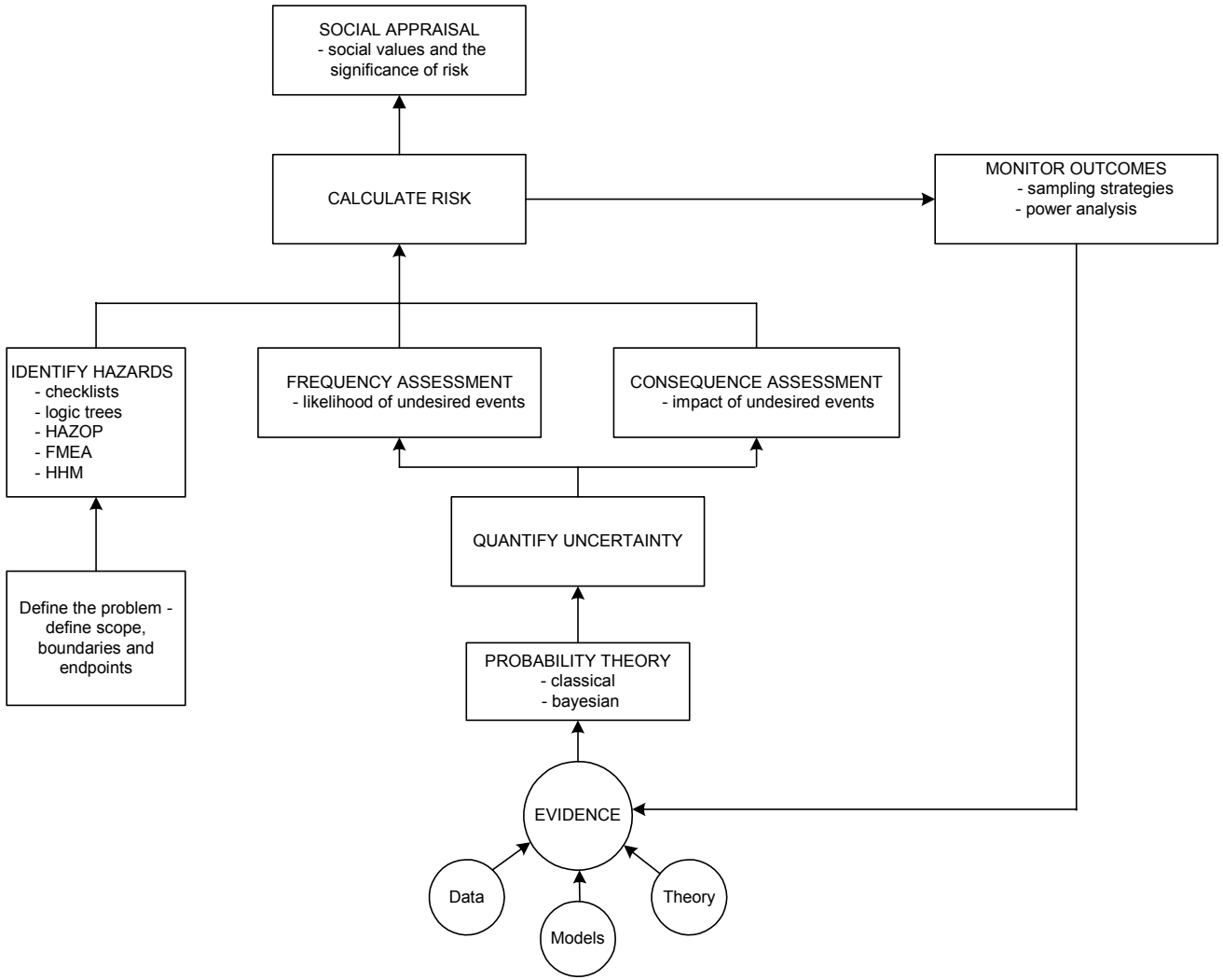
Conventional agricultural and aquaculture practices are not risk free. The assessment should therefore compare the novel risks associated with the release of GMOs against the risk of "no action" —i.e. the background risk associated with existing practice, projected over the expected lifetime of the release. It is important to recognise that GMOs may offer substantial risk reductions over existing practices, but also that existing practices are continually evolving and may be able to achieve similar risk reductions in different ways.

The number and variety of approaches to ecological risk assessment are bewilderingly large (Hayes, 1997). While there is no standard framework, each assessment should include all of the components identified in Figure 2, namely:

- a rigorous and systematic hazard analysis;
- data, theory and models collected and analysed in a manner that addresses the uncertainty regarding the likelihood and consequence of potential hazards identified in the first step;
- an estimate of risk based on the likelihood and consequence of events, that reflects the levels of uncertainty in the assessment process. Individual risk estimates are made for each assessment endpoint. A single risk assessment may address multiple endpoints and may therefore make multiple risk estimates;
- a monitoring system that tests the assumptions and predictions of the assessment in a statistically valid manner, over the lifetime of the project, and;
- a social appraisal of risk, including issues such as significance, duration, controllability, geographical scope, distribution across society, background risk and reversibility.

Each of these components should be framed by explicit spatial and temporal boundaries that recognise the extent and duration of exposure to each of the hazards identified in the analysis. They must also be linked in an iterative manner whereby the initial risk estimates are compared to observations and continually updated as and when additional information becomes available.

Figure 2 The components of an ideal risk assessment framework. Individual approaches and techniques may vary depending on the context of the assessment, but without all of these components the assessment is incomplete.



## 2.2 Hazard assessment

Ecological hazards may manifest in natural, arable and marginal environments and cut across all levels of biological organisation. Table 3 lists the ecological hazards typically associated with GMOs and some of the processes by which they might be realised. This type of checklist is just one of a number of techniques available to identify hazards in complex ecological systems (Table 4). All of the techniques listed in Table 4 have been successfully applied to industrial systems for many decades. Some are new to ecological risk assessment but almost all have been successfully applied to ecological problems at least once (see for example Royal Commission on Environmental Pollution, 1991; Haimes, 1998; Hayes, in review a; b).

Checklists and unstructured brainstorming are deductive approaches that are simple and easy to use. They will usually identify most if not all of the hazards that lie within the operating experience of those involved but do not encourage the participants to extend their expertise further. They do not confirm that all aspects of the system have been questioned, and may therefore give the impression that all the potential hazards have been identified when this is not in fact the case. All of the other techniques listed in Table 4 are inductive. They are designed to encourage a group of 'experts' to collectively interrogate the system and thereby apply their expertise beyond their own experience. These techniques are rigorous and systematic and will usually identify more potential hazards than either of the deductive approaches. They can also play an important heuristic role and are an excellent means to gather insight and possibilities from various stakeholders and interested parties, including non-scientists. Gathering the opinions and values of these groups in a systematic and transparent fashion is an important characteristic of best practice (see section 2.5). These inductive techniques take much longer to complete, however, and usually need to be facilitated by a skilled analyst.

A hazard assessment should initially be conducted without prejudice to the likelihood of events. Subsequent analysis (including common sense) may eliminate hazards that are simply too unlikely. It is also important to recognise that a single hazard can lead to multiple adverse effects. Conversely several hazards can have the same effect. Thus it is not sufficient to simply list all the potential hazards identified by the hazard assessment. The analysis should properly define the event series (or conditions of exposure) that lead from the hazard(s) to the endpoint(s) of the assessment, again emphasising that hazard is a function of the properties of the organisms and circumstances of the introduction. This may require the co-ordinated application of two or more of the techniques listed in Table 4—for example, hierarchical holographic modelling to identify a broad suite of hazards followed a fault tree analysis to identify the event chains associated with the most significant hazards.

A mathematical series is defined as an "aliorelative, transitive and connected" relation, (Russell, 1993). These properties establish a set of rules that uniquely define a series. A risk assessment should attempt to assign similar rules when defining the series of events between hazards and endpoints. Parallels can be drawn here with Koch's Postulates—a set of criteria logically necessary to establish the causal relationship between a specific microorganism and a specific disease. The analogous properties of an ecological event series are defined by the following questions: is the cause established for each step in the chain of events, (connected); does one step in the chain lead exclusively to the next step, (transitive) or to a possible number of alternative steps; are the intermediate steps from hazard to endpoint solely due to the hazard, (aliorelative) or does the endpoint occur naturally, via other events, etc.; and, is it reversible.

Table 3 Possible ecological hazards following the release of GM crops and the processes by which they might be realised

| HAZARD | PROCESSES |
|---|---|
| Toxicity to non-target organisms | Unpredicted expression (location, concentration) of toxin |
| Reduction in ecological fitness | |
| - same species | Decreased: pollen/seed dispersal, viability, density dependence threshold, r, mutualism, competitive ability; higher mortality; gene stacking |
| - related species | As above via gene flow, hybridisation and introgression; gene stacking |
| - unrelated species | As above via horizontal gene flow |
| Loss of genetic diversity | Gene flow and hybridisation particularly within “centres of origin” |
| Habitat modification | |
| - physical (fire, flood, low flows, soil erosion) | Altered interactions between man, environment and species composition |
| - biogeochemical cycles | Altered geochemical and nutrient fixation rates |
| Cascading ecosystem effects | Changes to mutualistic relations with non-target competitors, prey, hosts, symbionts, pathogens, herbivores or pollinators |
| Persistence in existing habitat | Decreased mortality, increased competitive ability, pollen/seed viability; increased resistance to biocides; gene stacking |
| Invasion of new habitats | |
| - same species | Increased: pollen/seed dispersal, density dependence threshold, intrinsic rate of increase, mutualism or competitive ability; lower mortality; gene stacking |
| - weedy relative | As above via gene flow, hybridisation and introgression; gene stacking |
| - unrelated species | As above via horizontal gene flow |
| Increased use of biocides | Selective pressure hastens biocide resistance in bacterial, viral, fungal or insect pests |
| Loss of biodiversity | Competition; added effects (gene, population, species) |
| Creation of new crop pests | Increased host range of viruses, fungi, pathogens or microbes. |
| Creation of new viruses | Partial or full coating of the genetic material of one virus with the coat protein of a different virus. Altered transmissibility, infectivity, latency or potency |
| Altered farming practice | Changes to cultivation of marginal land, seed line dependency, frequency of tillage, pesticide use, and energy and soil conservation |

Table 4 Hazard assessment techniques that can be applied to complex ecological systems

| TECHNIQUE | DESCRIPTION |
|---|---|
| Checklists | Simple and generally used to check compliance with good practice. Comprehensive checklists are often long and cumbersome, and can mislead the user into believing that all aspects that ought to be questioned have been without confirming that this is so. |
| Unstructured brainstorm | Simple brainstorming usually amongst a team of recognised experts. Quick and easy to conduct. Will identify most if not all hazards that lie within the operating experience of those involved. Does not, however, encourage participants to extend their expertise outside their immediate experience and may therefore mislead the user into believing that all aspects which ought to be questioned have been without confirming that this is so. |
| HAZOP analysis | Hazard and Operability Studies (HAZOP) uses guide words (such as 'more of', 'less of', 'reverse flow', etc.) to prompt a small team of experts to apply 'what if' type questions to each component of a system in a systematic manner. Other than providing the opportunity to address scenarios without the normal operating conditions of the system, HAZOP has the further advantages in that it is an open-ended procedure, which is more likely to identify all potential hazard scenarios than a simple checklist. |
| Logic Tree Analysis | Logic Tree Analysis is the construction of logic diagrams containing all conceivable event sequences that could lead to, or develop from, an undesired event. There are two alternative approaches; a top down approach in which the event is specified and all causative chains of events leading to this are identified (fault tree analysis), or a bottom up approach in which a triggering event is identified and the possible outcomes investigated (event tree analysis). Taken together the two approaches are sometimes termed cause-consequence diagrams. |
| Failure Modes and Effects Analysis | Failure Modes and Effects Analysis (FMEA) examines the behaviour and interaction of individual components of a system to enable the consequences of undesired events, upon the safety of the wider operation or process, to be assessed. The process identifies the system components and scores the likelihood of failure, the consequences of failure and the probability of detection based on current controls. |
| Analytical Hierarchy Process | The Analytical Hierarchy Process is a formalised procedure for ranking hazards. It uses the opinion of an 'expert' group of assessors to rank hazards based on their perceived importance. |
| Hierarchical Holographic Modelling | Hierarchical Holographic Modelling (HHM) examines complex systems from a number of different perspectives in order to identify important interactions between the various components and processes of the system. HHM is most effective when the system is "well defined" when the analyst is able to identify and list all the important components and processes in a hierarchical fashion. |

Ideally the analyst is seeking a unique, connected, transitive and aliorelative relation in the chain of events between the hazards identified in the hazard assessment and the endpoint(s) of the assessment. This is not to say that there will be only one possible event series, there may be several, but each series should exhibit these properties. The principal advantages of the inductive hazard assessment techniques listed in Table 4 are that they force the analyst to think very carefully about why the system may fail and to construct the event series linking hazards to endpoints. In this way the analyst is able to construct a qualitative "proto-model" of the hazard scenarios. This proto-model forms an excellent basis for qualitative and quantitative risk assessment.

**Best practice recommendation #2: Construct good qualitative models of all hazard scenarios using structured deductive and inductive hazard assessment techniques**

## *2.3 Likelihood and consequence of events*

Arguably the most vexing question in ecological risk assessment is whether to adopt a qualitative or quantitative approach. Qualitative assessments score the likelihood and consequence of events as negligible, low, medium, high, etc. The overall risk estimate is defined via simple combination rules, usually portrayed in a two dimensional matrix that compares the likelihood of events and their consequences (see for example Standards Australia, 1999).

Qualitative risk assessments are attractive because they are relatively quick and easy to conduct, and they maintain the two dimensions of risk (likelihood and consequence) in the final calculation. They are often recommended as an initial screening activity where the level of risk does not justify the time and effort of a quantitative analysis, or where there are insufficient data for such an analysis (Standards Australia, 1999). This last point is particularly pertinent to ecological risk assessment. Analysts and regulators often argue that a full quantitative treatment is not possible because of the complexity of ecological systems and/or a lack of sufficient data. Retreating to qualitative techniques, however, dispenses with methods that treat uncertainty transparently, in favour of subjective techniques that are less capable of dealing with uncertainty and are less transparent about doing so.

People are poor judges of probabilistic events. Their judgement is adversely affected by the level of control they have over the outcome, their level of understanding, the extent of their personal experience, the apparent dreadfulness of the outcome, who ultimately bears the burden of risk, and the visibility of the hazard. Furthermore when individuals assess risks subjectively they are often influenced by cognitive bias (overconfidence in one's ability to predict), framing effects (judgements of risk are sensitive to the prospect of personal gain or loss, in which losses loom larger than gains), anchoring (the tendency to be influenced by initial estimates) and insensitivity to sample size (Burgman, 2002). One important result of these effects is a tendency to make overly narrow estimations of the probability distributions, driven largely by an unfounded optimism about the uncertainty surrounding our subjective predictions—both naïve and sophisticated subjects tend to be more confident about their predictions than they should be. For this reason qualitative assessments may not err on the side of conservatism even when they purport to do so (Ferson and Long, 1995). Qualitative risk assessments are also much more vulnerable to linguistic uncertainty than quantitative techniques. As a result it may be harder for a qualitative risk assessment to satisfy the scientific principles described above.

**Best practice recommendation #3: Consider the influence of cognitive bias, framing effects, anchoring and sample size on qualitative decisions.**

An alternative approach is to treat one or more components of the system quantitatively and the rest qualitatively. Under these circumstances the analyst must be very careful when combining experimental data, or quantitative expressions of risk, with expert judgement (Pollard, 2001). In practice it is difficult to combine the two approaches in a meaningful way—although possible so long as all qualitative terms are defined as numeric intervals. Another possible solution is to seek a simpler assessment endpoint that is sufficient for regulatory purposes and can be completely described with quantitative methods (see for example Hayes and Hewitt, 2001).

The time and cost of quantifying the likelihood and consequences of GMO releases varies depending on the organism, the hazard and the endpoint concerned. A proportionate, step-wise, response seems appropriate given the limited resources available to most regulatory authorities. Ideally releases should proceed in a step-wise manner from contained laboratory tests to small-scale field trials and then large-scale field trials prior to full commercial use.

Risks that are demonstrably and unequivocally negligible can be screened from the analysis. Demonstrably negligible risks are, for example: a) risks that have been assessed for identical circumstances and found to be negligible; b) risks whose likelihood is vanishingly rare under all exposure conditions; or c) inconsequential risks[3]. Simple screening models that incorporate robust expressions of uncertainty can identify negligible risk and thereby eliminate the need for more detailed assessments (see below).

This review has identified a large number of test systems, measurement methods and models that can be used to investigate, and in some circumstances quantify, biological processes relevant to GMO risk assessment. These are listed by hazard in Table 5. Table 5 is by no means exhaustive but it does indicate the variety of methods that are currently available. These methods and models are the foundation of best practice ecological risk assessment for GMOs.

**Best practice recommendation #4: Consider the full spectrum of ecological models from simple (screening) to detailed ecosystem models.**

The types of models available to the risk analyst vary from simple toxicity extrapolation models, to single species population models, meta-population models, ecosystem models and detailed landscape models. The choice of model is site- and issue-specific and depends on the endpoint concerned, the practicality, reliability and regulatory acceptance of the model, and the quality and quantity of available data. Practicality refers to the degree of development, ease of estimating parameters and resource efficiency of the model. Reliability refers to the biological realism, relevance, flexibility and how the model treats uncertain parameters.

Table 6 scores each of the models discussed in Appendix B against these criteria (*sensu* Pastorok et al., 2002). In virtually all cases the degree of regulatory acceptance is assumed or inferred from similar (or identical) models used in chemical risk assessment because there are very few examples of quantitative ecological risk assessment for GMOs. Despite the rich literature very few of the models discussed here appear to be advocated or recognised by regulatory agencies. The one exception to this

[3] Generally speaking it becomes increasingly more difficult to demonstrate negligible risk from a) to c).

is the Canadian Food Inspection Agency, which identifies several seed replacement and seed dormancy studies within their regulatory directives (see section 3.3).

Table 5 Laboratory, green house and field-methods, protocols and models relevant to risk assessment for GM plants and microorganisms

| HAZARDS AND PROCESSES | REFERENCES FOR METHODS AND MODELS[a] |
|---|---|
| Toxicity to non-target organisms | |
| - soil microorganisms | (3) Wolfenberger & Phifer, 2000; (1) Beyer & Linder, 1995 |
| - soil invertebrates | (42) Jepson *et al*, 1994; (2) Beyer & Linder, 1995 |
| - insects, spiders and mites | (7) Wolfenberger & Phifer, 2000; (3) Beyer & Linder, 1995 |
| - birds | (9) Hoffman, 1995 |
| - mammals | (3) Hoffman, 1995 |
| - amphibians and reptiles | (?b) Hoffman, 1995 |
| Reduction in ecological fitness OR increased persistence OR invasion of new habitats OR loss of genetic diversity | |
| - density dependant threshold | (17) Kjellsson & Simonsen, 1994 |
| - fitness and competitive ability | (24) Kjellsson & Simonsen, 1994; (2) Lenski, 1991; (1) Kim *et al,* 1991; (1) Wolfenbarger and Phifer, 2000 |
| - intrinsic rate of increase (plants) | (1) Parker & Kareiva, 1996; (1) Crawley et al., 1993; (1) Andow, 1994 |
| - pollen dispersal | (12) Kjellsson *et al*, 1997; (2) Kareiva et al, 1994; (6) Giddings et al., 1997a; 1997b; (2) Lavigne *et al*, 1998 |
| - pollen viability | (6) Kjellsson *et al*, 1997 |
| - seed dispersal | (7) Kjellsson & Simonsen, 1994; (1) Pessel *et al*, 2001; (1) Crawley & Brown, 1995 |
| - seed dormancy and replacement | (1) Linder & Schmitt (1994); (2) Kjellsson & Simonsen, 1994; (1) Crawley *et al*, 1993; (1) Rissler and Melon, 1993 |
| - hybridisation and introgression | (40) Kjellsson *et al*, 1997; (5) Wolfenberger and Phifer, 2000 |
| - gene flow (direct measurement) | (1) Hokanson *et al*, 1997 |
| - gene flow (indirect) | (3) Slatkin & Barton, 1989 |
| - horizontal gene flow | (12) Kjellsson *et al*, 1997; (1) Landis et al., 2000; (1) Strauss et al., 1985 |
| - spread | (13) Strauss & Levin, 1991; (5) Manasse & Kareiva, 1991 |
| - establishment | (2) Tomiuk and Loeschcke, 1993; (4) Williamson, 1989 |
| Habitat modification: Physical | |
| Habitat modification: Biogeochemical cycles | |
| - carbon (microorganisms) | (10) Jepson *et al*, 1994 |
| - nitrogen (microorganisms) | (7) Jepson *et al*, 1994 |
| - phosphate (microorganisms) | (1) Jepson *et al*, 1994 |
| - sulphur cycle (microorganisms) | (1) Jepson *et al*, 1994 |

Table 5 continued...

| | |
|---|---|
| Cascading ecosystem effects | |
| - food-web effects | (1) Pimm, 1982 |
| - changes to trophic flow | (1) Li et al., 1999; (1) Ulanowicz, 1992 |
| - changes to mutualistic relations | (1) Wotton (1994) |
| Increased use of biocides | |
| - selective pressure | (10) Endler, 1986 |
| Loss of biodiversity (plants) | (2) Kjellsson *et al*, 1997 |
| Creation of new crop pests | |
| Creation of new viruses | |
| - transcapsidation | (1) Wolfenberger & Phifer, 2000 |
| Altered farming practice | |
| - non-point source pollution | (1) Strauss & Levin, 1991 |
| - weed demography & cropping effects | (26) Colbach & Debaeke, 1998 |

[a]Numbers in parenthesis indicate the approximate number of methods, protocols or models available—precise figures are not always available and some double counting may occur.

[b]The author notes that several protocols are published.

Table 6 Practicality, reliability and acceptance of some models that are relevant to potential ecological risks associated with GM plants and microorganisms

| **Hazardous process** | **Model type** | **Practicality** | **Reliability** | **Acceptance** |
|---|---|---|---|---|
| Toxicity to non-target organisms | HCp toxicity extrapolation | High | High | High |
| Altered fitness or competitive ability | Malthusian-like differential equation | High | Low | High |
| Altered fitness or competitive ability | Extreme value function | Med | High | Medium |
| Altered fitness or competitive ability | Deterministic geometric function | High | Med | High |
| Pollen dispersal | Reliability functions | Med | Med | Med |
| Horizontal gene flow | Malthusian-like differential equation | High | Low | High |
| Horizontal gene flow | Meta-population model | Med | High | Med |
| Spread | Simple density-dependant | High | Med | High |
| Spread | Lotka-Volterra with competition term | Med | Med | Med |
| Establishment | Velhurst-Pearl, Reproductive ratio | High | Low | High |
| Establishment | Individual based | Low | High | Low |
| Cascading ecosystem effects | Food web | Low | Med | Low |
| Cascading ecosystem effects | Trophic flow/pathway analysis | Low | Med | Low |

It is clear that there are a number of quantitative techniques designed to address many of the hazards typically associated with GM plants and microorganisms. The practicality, reliability and regulatory acceptance of these techniques, however, vary markedly. In most cases the simpler models are the more practical—these types of models are well developed, easy to parameterise and quick to run. Their biological reality, however, may be low. They are therefore most useful in the early screening stages of a risk assessment to identify demonstrably low or high-risk scenarios, particularly if combined with meaningly descriptions of parameter uncertainty.

**Best practice recommendation #5: Recognise that even simple models can incorporate uncertainty and be useful in ecological risk assessment**

Models with a high to medium level of reliability, but a concomitantly low to medium level of practicality are better suited to uncertain but potentially high-risk scenarios that warrant additional time and effort. Individual based models and meta-population models, for example, can give valuable insight into the emergent behaviour of an ecological system that is virtually impossible to identify with qualitative methods. Again, however, it is essential that these models incorporate an adequate uncertainty analysis even if this occurs at the cost of precise estimates—it is better to be broadly right than precisely wrong.

Food web models and trophic flow/pathway analysis are very relevant to potential ecosystem level hazards. These techniques have the potential to be biologically realistic but are extremely labour and computer intensive, and to date are not well developed or widely employed in ecological risk assessment. This is an important area for future research. Important knowledge gaps appear to exist in four other areas: altered farming practice, physical habitat modification and creation of new crop pests and viruses. For example, transcapsidation (a rare phenomenon whereby the coat protein of one virus completely or partly encloses the genome of another virus) has been demonstrated in the laboratory but there appears to be no other information on its likelihood and consequences in the field (Wolfenberger and Phifer, 2000). It is known to occur in crops simultaneously infected by several viruses, and may transiently alter the range of the encapsulated virus, thereby effectively creating a new virus (Rissler and Melon, 2000). This scenario, however, does not appear to have received sufficient attention from quantitative ecologists studying the potential impacts of biotechnology.

## *2.4 Uncertainty analysis*

Uncertainty occurs throughout the process of constructing and releasing a GMO (Figure 3). The term “genetic engineering” implies a great deal more precision than actually occurs (Rissler and Mellon, 2000). Scientists know that a number of transgenic techniques work but, with the exception of Agrobacterium-mediated transformation, they do not know what the precise integration mechanisms are (Walden and Wingender, 1995). In most cases gene insertions occur at random, unpredictable loci (OECD, 2000b). The insertion may disrupt other coding or regulatory regions causing insertional mutagenesis. Further sources of uncertainty occur during transcription and protein synthesis leading to a variety of phenotypic or metabolic responses in the organism. For these reasons first generation GMOs are rarely released outside of the laboratory. Usually many generations of the organism need to be selected within the laboratory in order to weed out the pleiotropic “side effects” of the technology, arriving finally at the expected phenotype. Ecological risk assessment for unconfined releases is therefore primarily motivated by three sources of uncertainty: the context (environment) specific performance of the GMO; its interaction with other ecosystem components and processes; and, landscape changes, following commercial scale release, over evolutionary time scales.

Uncertainty analysis is a critical component of ecological risk assessment. It distinguishes risk assessment from impact assessment, promotes transparency and credibility, and improves decision-making—indeed it is the very rationale of risk-based environmental management. Various authors have offered different taxonomies of uncertainty (see for example Faber et al., 1992 or Finkel, 1990). Figure 4 outlines a taxonomy based on Regan et al., (2002) and Morgan and Henrion (1990) in which uncertainty is divided into two types: linguistic and epistemic. Linguistic uncertainty occurs as vagueness, ambiguity and lack of specificity (arising when a statement does not provide sufficient detail), and whenever the analyst fails to specify the context in which a proposition is made. Linguistic uncertainty is particularly prominent in qualitative risk assessment. Terms such as "low risk" for example are routinely used without reference to exposure—with sufficient time or number of "trials", low risk events may be more or less certain.

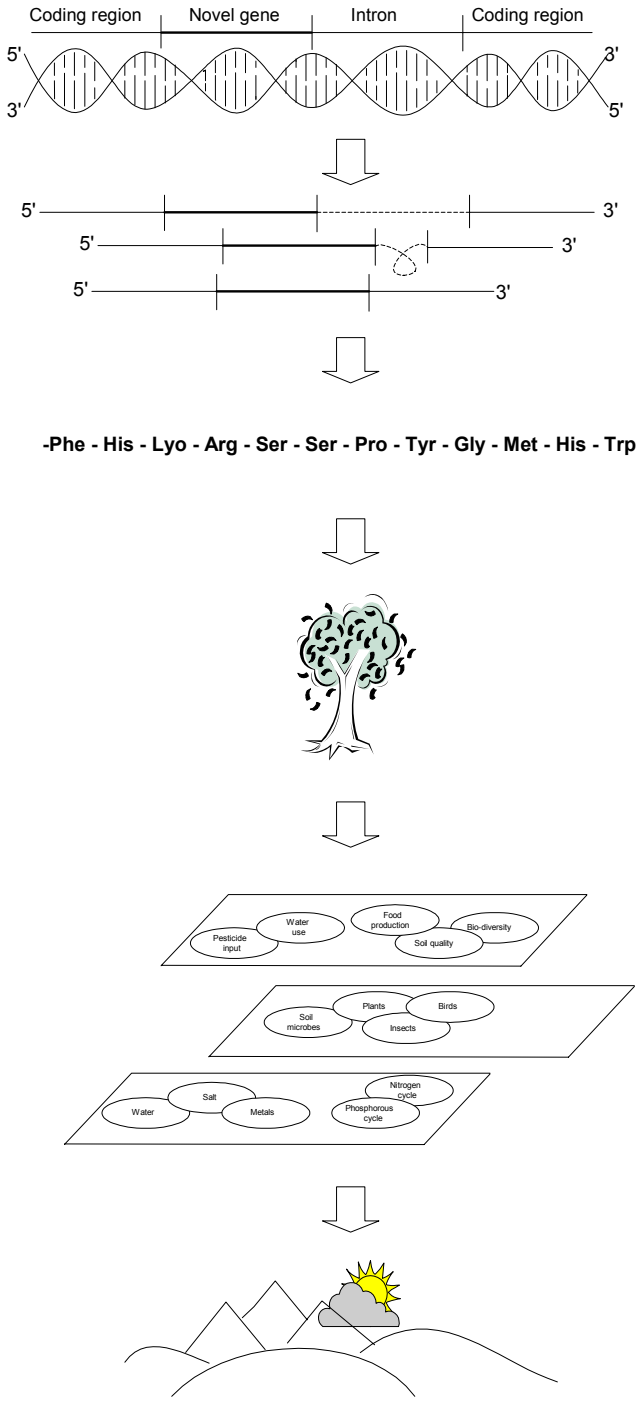
**Best practice recommendation #6: It is essential to include a transparent analysis of uncertainty**

Contextual uncertainty can occur in the spatial and temporal components of the assessment and in its scope, resolution and boundaries. Ecosystem level risk assessments may be contextually under-specified because it is often difficult to precisely define the boundaries and scope of the assessment. It makes no sense, however, to be uncertain about contextual uncertainty—the analyst must select appropriate spatial and temporal boundaries, and time steps within the assessment. It is important, however, that stakeholders and interest groups are involved in this decision and made fully aware of its implications (see section 2.5). The analyst can vary the scope of the assessment to examine how this affects the result, although this is usually quite a time-consuming process. The other forms of linguistic uncertainty can be reduced by carefully defining the assessment's terms of reference and language. Ultimately, however, they can only be eliminated mathematically.

Epistemic uncertainty reflects our limited knowledge of ecological systems. It occurs as measurement error (random—resulting from imperfect measuring devices, and systematic—resulting from bias), natural variation, model error, subjective judgement (a result of data interpretation in which expert opinion determines the value of a variable) and ignorance. Ignorance, model error and measurement error are often collectively known as incertitude because they can be reduced with empirical effort. Random measurement error is minimised by taking additional measurements. Systematic measurement error is minimised by careful experimental design and instrument calibration. Natural variability cannot be reduced with empirical effort but can be described using uncertainty calculi (see below).

Model error occurs in the boundaries, structure and components of a model, in the types and parameters of probability distributions used to represent uncertain empirical quantities, and in the specification of dependencies among randomly varying elements (Ferson and Ginzberg, 1996). Analysts are generally aware, before the fact, that models are caricatures of reality. The error this causes is only apparent after the fact, if at all, and cannot be addressed in a predictive manner. The validity of a model, however, can be tested against data that are significantly different from the calibration conditions, a suite of candidate models that use different assumptions, or by comparing its predictions against observations of reality using a statistical goodness-of-fit test. Reckhow and Chapra (1983) for example list goodness-of-fit tests that can be adapted to virtually any model. Model uncertainty can also be minimised by choosing the simplest assessment endpoint that meets the needs of environmental managers and regulatory authorities. In this context the simplest assessment endpoint is that which can be analysed with the greatest precision (Reckhow, 1994; Hayes and Hewitt, 2001).

Figure 3 Uncertainty occurs at various stages in the construction (gene insertion, transcription and protein synthesis) and release (case specific phenotypic and metabolic characteristics, interactions within the ecosystem and landscape processes) of a GMO. Ecological risk assessment is primarily concerned with uncertainty following the release of a GMO.




**A: Gene** - disruption in either a gene coding region or associated regulatory regions

**B: Transcription** - pattern of transciption is an integrated response to internal and external factors impinging on the cell

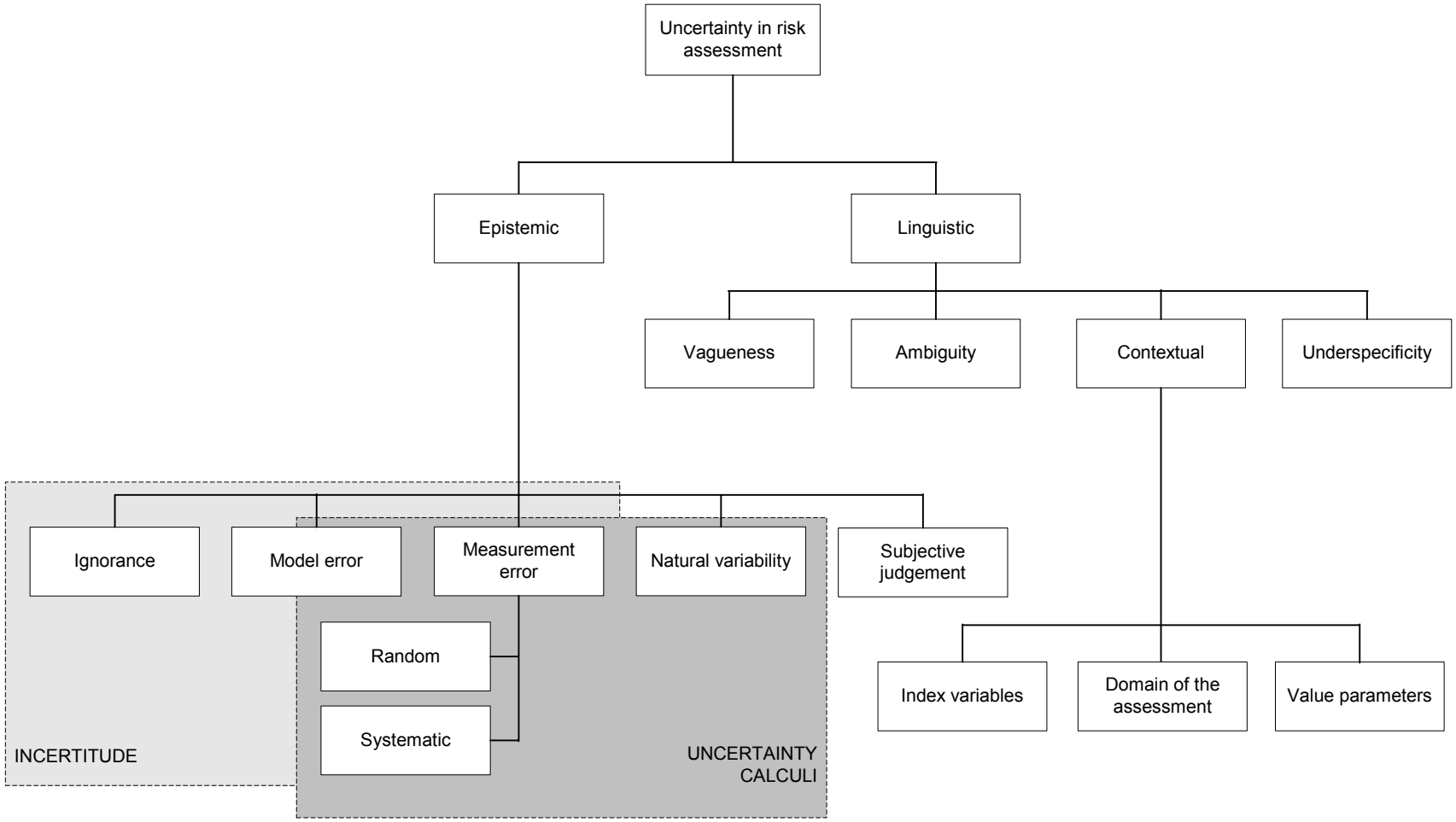
**C: Protein profile** - not all changes in gene transcription will necessarily be reflected in predictable changes in the proteins synthesised

**D: Phenotype** - altered phenotyopic and metabolic characteristics of the organism. Ecological performance is not genotype specific but context (environment) specific

**E: Ecosystem** - new or altered relations between ecosystem objects and processes

**F: Evolutionary landscape** - evolutionary fate of the transgene and new phenotype and unpredicted changes in landscape processes over evolutionary time scales

Figure 4 The types of uncertainty in ecological risk assessment. Incertitude can be reduced with empirical effort, natural variability cannot. Techniques exist that enable risk assessment calculations under conditions of model error, measurement error and natural variability.



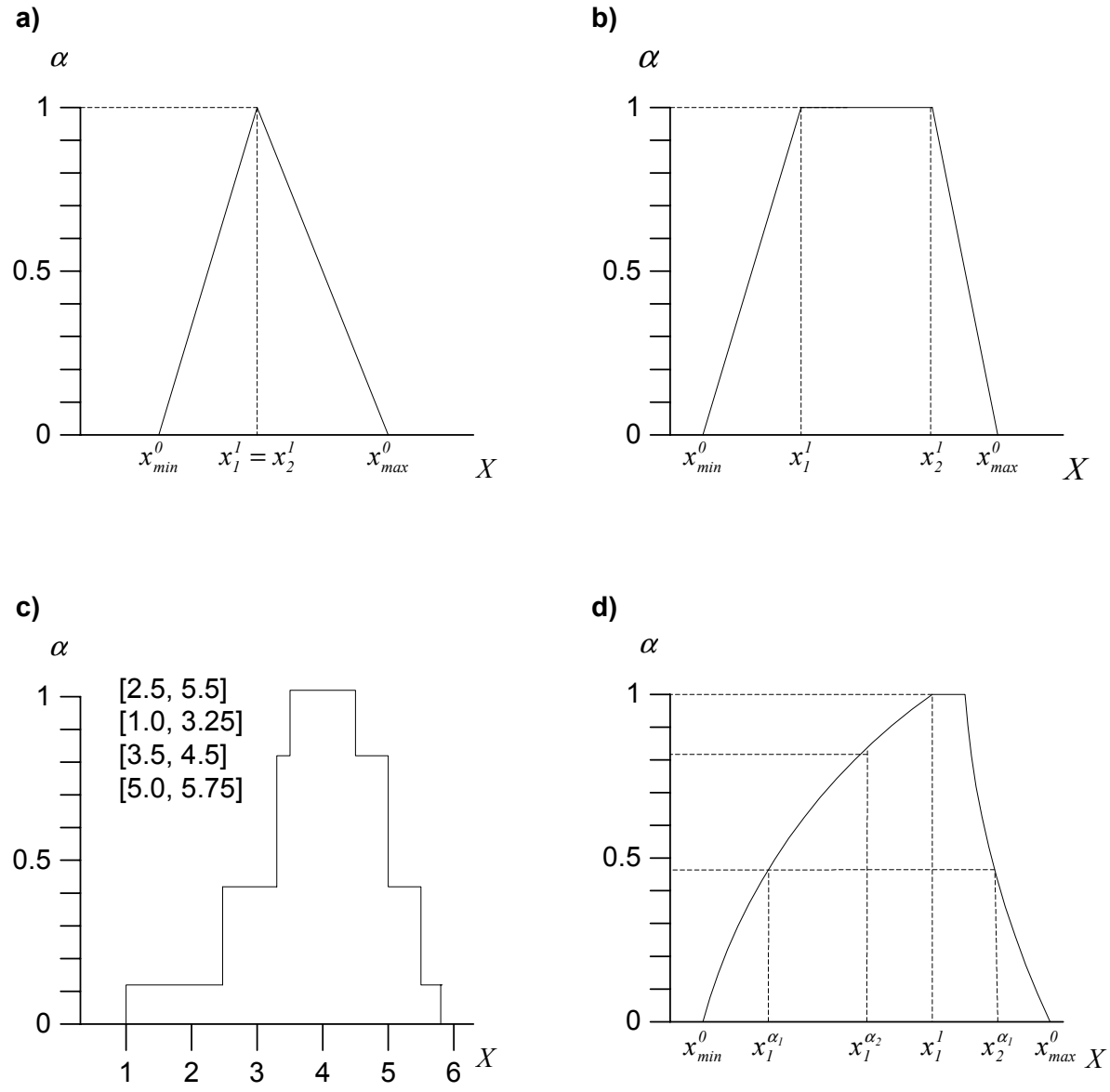
There are a large number of techniques (uncertainty calculi) to address measurement error and natural variability. The most common are: worst case analysis; interval arithmetic; fuzzy arithmetic; Monte Carlo analysis (including second order methods) and probability bounds analysis. Worst-case analysis usually entails simple models with plausibly extreme parameters. It accounts for uncertainty by being conservative and is easy to perform. The level of conservatism, however, is sensitive to the (arbitrary) number of calculations within the risk assessment, can quickly become hyper-conservative, and is not consistent between studies—it is therefore impossible to compare risks between studies (Ferson and Long, 1995).

Interval arithmetic belongs to a family of techniques known as bounding. It is used when only the upper and lower bounds of a continuous variable are real and are known, or can be estimated. All the usual mathematical operations can be easily performed with intervals allowing the analyst to specify the possible range of a risk assessment function or model output. If *X* and *Y* are non-negative real random variables on the interval $[x_1, x_2]$, $[y_1, y_2]$ then sum, substraction multiplication and division are simple and intuitive operations. If X and Y take non-negative values then the process is slightly more complicated (refer to Kaufman and Gupta, 1985, or any of the numerous texts on interval analysis for more details). The approach is rigorous, intuitive and easy to perform. Interval analysis does not, however, provide any information on the likelihood of values within the range, or at its tails, which are often particularly important to risk managers. Furthermore the overall range of multiple calculations with multiple interval variables grows very quickly, often to the extent that the result holds little predictive or discriminatory power.

Fuzzy numbers simultaneously specify the range of an uncertain variable and the plausibility or possibility of intervening values. The level of "presumption" for any number of values on the range $[x_{min}, x_{max}]$ describes the level of possibility of these values between 0 and 1. Fuzzy numbers are formally defined as a fuzzy subset that is convex and normal (Kaufman and Gupta, 1985). Put simply they are intervals with one "peak" at a level of presumption of 1. The simplest fuzzy numbers are triangular or trapezoidal and have the form $[x^0_{min}, x^1_1 = x^1_2, x^0_{max}]$ and $[x^0_{min}, x^1_1, x^1_2, x^0_{max}]$ respectively. More complicated forms, however, can be constructed by "stacking" a series of interval estimates or simply specifying three or more intervening values (and their associated level of presumption) on the interval range (Figure 5). Various software packages are available that are capable of performing all the standard mathematical functions with fuzzy numbers. Fuzzy arithmetic is simple, requires little data and potentially very useful because it simultaneously yields "worst-case" and "best-estimate" results (Ferson, 1994). Fuzzy arithmetic becomes cumbersome, however, with repeated variables and cannot use knowledge of correlations to tighten the risk bounds (Ferson et al., 2001).

Monte Carlo analysis requires the analyst to specify parametric or non-parametric probability distributions for natural variables (and the correlations or dependencies between these variables) within a model. The analysis proceeds by randomly selecting values from each of these distributions, returning the value to the model and completing the calculation. This process is repeated (ideally several thousand times) and the results of the model collated and presented in a probabilistic form. The analyst can capture the uncertainty associated with natural variables with a number of statistical techniques. These techniques include, in roughly increasing order of complexity: sample distribution functions; kernel density estimators; discrete and continuous probability density functions; and, extreme-value distributions. Sample distribution functions are simple and do not make any assumptions about the data. They do not require large amounts of data (20 or more observations) and it is easy to quantify the uncertainty surrounding the distribution using simple second order methods (Appendix C). They do not, however, represent the tails of a distribution well—indeed the analyst must somehow choose the maximum and minimum values of the distribution.

Figure 5 Four examples of a fuzzy number: a) triangular; b) trapezoidal; c) constructed by stacking four interval estimates; and, d) general (convex and normal).



Kernel density estimates have similar advantages; they require few assumptions and can be applied to small (only 22 observations are needed to achieve a good density estimate of a symmetrical, unimodal distribution, Epanechnikov, 1969) and large data sets. The reliability of kernel density estimators, however, is very sensitive to the choice of bandwidth *h*. The best bandwidth is essentially a subjective judgement although "automatic" methods are available (refer to Appendix C).

Parametric approaches use the sample data to estimate the parameters of a theoretical distribution. These approaches assume that the data represent a random sample drawn from the population distribution. They require relatively large amounts of data and in practice involve finding the most likely population distribution, because it is virtually impossible to find a distribution that exactly fits. There are various means to fit a probability distribution (for example method of moments, maximum-entropy, chi squared, Anderson-Darling, or Kolmogrov-Smirnov—see Palisade, 1996 and Ferson et al., 2001 for details). With the exception of the method of moments and maximum entropy, however, all of these techniques are computationally intensive.

The problem is further compounded by the large number of theoretical distributions that could potentially describe the sample data—there are 62 classes of univariate and multivariate distributions for discrete data and 108 classes of univariate distributions for continuous data (Patil et al., 1984a; b). The analyst will therefore require a software package that can perform goodness-of-fit tests quickly, for a variety of distributions, unless there is clear prior evidence to suggest a particular distribution for the parameter in question (rarely true for environmental systems). A parametric distribution may also allocate finite probability estimates to values that were not actually observed, or worse, are nonsensical, and often may not fit the data extremes (the tails of the distribution) well.

Extreme values are often the most interesting to a risk analyst because environmental extremes are often the most ecologically significant events (Gaines and Denny, 1993). It is often better, therefore, to explicitly model sample extremes using an extreme-value (EV) distribution (Appendix C). Large amounts of data (long time-series) are usually required to obtain an EV distribution. Furthermore, the asymptotic theory of extreme values assumes that the data are independent and identically distributed, and therefore stationary. Long term trends (non-stationary) and autocorrelation (non-independent) within a time series are common violations of these assumptions. Trends within the data, however, are easily removed by analysing the residuals from a regression analysis. If the dependency between samples decreases with increasing time intervals the asymptotic distribution of the extremes is the same as in the case of independent and identically distributed samples. The practical constraint is that the sampling interval should be longer than the interval between essentially independent samples (Gaines and Denny, 1993).

From a risk assessment perspective, EV distributions have a number of important advantages over other probability distributions. The most important is that EV distributions of dependant variables are always independent (Gumbel, 1962). This is important because ecological variables are more often than not dependant and Monte Carlo simulation does not provide accurate answers without information about the correlation between variables. Most simulations are performed, however, without any empirical information about correlation because of the considerable cost of the sampling effort needed to estimate it (Ferson and Long, 1995). This is not an issue if the simulation is performed with EV distributions, although the sampling effort needed to generate these distributions may incur similar costs.

The principal advantage of probability bounds analysis is that it allows accurate arithmetic operations on random variables without making any assumptions about the correlation among these variables. Furthermore, it is more efficient and provides more precise results than Monte Carlo analysis, and can

be employed with virtually any distribution of a random variable. Probability bounds analysis constructs discrete upper and lower approximations of the distribution function $F(x)$ by dividing the vertical axis (cumulative probability) into $n$ equal segments. This creates $n$-1 'probability boxes' that simultaneously place interval bounds on the cumulative probability of a value $x$, and on the value of a cumulative probability $F(x)$ (Ferson et al., 2001). The latter are then used to calculate the best possible point-wise bounds for addition, substraction, multiplication and division of any two random variables, based on the quantiles $i/n$ and $(i + 1)/n$ for the upper and lower values of the cumulative probability (see Williamson and Downs, 1990 for details).

Probability bounds analysis is generalisation of interval analysis and probability theory. It gives the same answer as interval analysis when there is little information, and the same answer as Monte Carlo methods when there is abundant data, but importantly makes no assumptions about the dependency between random variables. Furthermore, it is capable of mixing these approaches within the same analysis and is therefore capable of handling information of widely different quality. It cannot, however, incorporate information on correlation between variables (in the rare cases that this is actually available) and cannot therefore use this information to tighten the bounds within the analysis. The approach also entails a number of other minor drawbacks—for example, distributions on an infinite support must be truncated to finite limits, and risk algorithms with multiple occurrences of the same variable need to be re-specified otherwise probability bounds analysis will yield answers with artificially inflated levels of uncertainty (Ferson and Long, 1995).

**Best practice recommendation #7: When information is sparse use probability bounding analysis to express uncertainty**

## 2.5 Social appraisal of risk

New technologies usually present a variety of potential hazards. Genetically modified crops, for example, may have a variety of economic, agricultural, social and ecological impacts (some of which are highlighted in Table 5). Great diversity exists even within these categories, such that the ecological risks of GMOs, for example, cannot be characterised by a single uniform metric—they are multidimensional, usually incommensurable, more or less amenable to quantification (deductively or inductively) and are characterised by different types of uncertainty. This level of complexity is further compounded by the fact that different cultural groups, political constituencies or economic groups typically attach different degrees of significance to different hazards and hence their decisions regarding the acceptability of risk are based on much more than just its absolute estimate and associated uncertainty.

Different perspectives on the significance and acceptability of risk are largely driven by the degree to which exposure to the risk is voluntary[4]; who benefits; the temporal and geographical scope of the risk; the extent to which the impacts are reversible; and, the extent to which risk is known or understood by society and whether or not it has been successfully managed in the past. These characteristics are often used by regulators when deciding how cautious or risk averse they are when evaluating the risk (see for example, Environmental Risk Management Authority, 2000), and should properly be acknowledged within the risk assessment process. The combination of multi-dimensional, incommensurable hazards with different (but equally legitimate) significance attributes precludes any single analytical fix to the problems encountered in the social appraisal of risk. Best practice

[4] The public is thought to accept risks from voluntary activities that are roughly 1000 times as great as it would tolerate from involuntary hazards that provide the same level of benefit.

ecological risk assessment is therefore as much about systematic qualitative evaluation of divergent social values as it is about numerical characterisation of the likelihood and consequences of hazards (Stirling, 1999).

**Best practice recommendation #8: Examine opportunities to promote on-going stakeholder participation in the risk assessment**

The lay person's perception of risk often lacks important pieces of information regarding the likelihood and consequences of hazards, but their conceptualisation of risk is typically much richer than the risk analyst's and reflects legitimate concerns that are often omitted from exclusively 'science-based' risk assessments. Techniques to include stakeholders in a risk assessment should therefore be designed to more fully inform both the stakeholders of the science and the analyst of stakeholder values. The inductive hazard identification techniques described in section 2.2, for example, provide an excellent means to harness the imagination and intuition of 'non-scientific experts' with a variety of different perspectives—such as farmers and landowners in the case of GMO crops. This helps inform the analyst by converting ignorance into tractable uncertainty and at the same time raises awareness within stakeholders groups of the risk assessment process and procedure.

Some of the potential hazards associated with GM crops are very uncertain and potentially highly damaging (for example the creation of new viruses and crop pests). Risk estimates for these hazards are complicated by the high level of ignorance and are likely to be strongly contested by stakeholders and other interested parties. Best practice risk assessment should not shy away from active social contention and healthy dissent—they are important engagement and quality control tools in the social appraisal of risk. A precautionary approach to these types of hazards, however, is warranted. The practical implementation of such an approach invokes a range of sub-ordinate principles and concepts which are summarised in Table 7. These concepts recognise the limitations of science and the legitimacy of values held by different interest groups. They therefore require a strong element of 'social discourse' within a risk assessment (Stirling, 1999).

**Best practice recommendation #9: Adopt a precautionary approach to high consequence, but highly uncertain, hazards**

It is also particularly important in these circumstances to ensure effective collaboration between risk analysts, policy advisors and regulatory agencies because policy failures in these circumstances quickly undermine public confidence in the competence of those formally charged with the governance of new technologies. For example the October 2000 report of the BSE inquiry (http://www.bse.org.uk/) concludes, *inter alia*, that whilst the UK Government introduced measures to guard against the risk that BSE might be a matter of life and death not merely for cattle but also for humans, the possibility of a risk to humans was not communicated to the public or to those whose job it was to implement and enforce the precautionary measures. When on 20 March 1996 the Government announced that BSE had probably been transmitted to humans, the public felt that they had been betrayed. Confidence in government pronouncements about risk was therefore a further casualty of the crisis.

Table 7 Key subordinate principles and concepts associated with precautionary approaches to risk assessment (after Stirling, 1999)

| | |
|---|---|
| **Subordinate principles** | 'Prevention'—a duty to prevent rather than control or treat an impact |
| | 'Polluter pays'—the placing of burdens on all parties responsible for, or benefiting from, damaging activities |
| | 'No regrets'—presumption in favour of options simultaneously satisfying economic, environmental and wider criteria |
| | 'Clean production'—adopt only those investment or technology options which are demonstrably of lowest impact |
| | 'Biocentric ethic'—recognise the intrinsic value of non-human life |
| **Associated concepts** | Acknowledge the limitations of science, humility about knowledge and anticipation of surprise |
| | Recognise the vulnerability of the natural environment |
| | Uphold the rights of those who are adversely affected by the new technology |
| | Take account of the availability of technical alternatives |
| | Consider the complexity of behaviour in real organisations |
| | Pay attention to variability of local and other contextual factors |
| | Assign equal legitimacy to different value judgements |
| | Adopt long-term, holistic and inclusive perspectives to appraisal |

## *2.6 Monitor and review*

Risk assessment is an iterative process. Experiments and observations designed to test the predictions of the assessment should generate information that adds to the body of evidence used to describe uncertainty (Figure 2). It is much harder to monitor for "general" events than for specific events, in a scientifically valid manner. The risk assessment should therefore guide the analyst on what to look for and where and when to look for it. This is only possible, however, if the hazard identification and risk assessment is rigorous and systematic.

Monitoring strategies should include a statement of objectives, precise descriptions of the design of experiments, data that will be collected and the methods of analysis to test for statistical significance and the power of the test procedures. Standard collection, handling and experimental protocols should be used wherever possible to help minimise experimental error and allow comparisons between sites and crops. Field trials, treatments and monitoring strategies should be well replicated within sites and over a wider variety of sites (again, ideally in each biome that the GMO might be released into) to ensure that the GMO is tested in an appropriate range of arable and natural habitats (Crawley, 1990). Again replications should be sufficient to detect changes of a pre-specified magnitude, otherwise the assessment may run the risk of being underpowered, and of failing to detect important consequences.

Many experimental field studies are designed around null hypothesis tests such as those listed in Table 8. The default assumption is that if no problem is observed then none exists such that the burden of proof lies with the monitoring program. In these circumstances reliability depends on statistical power—on the ability of a method to detect real outcomes against a background of natural variability, measurement error and ignorance concerning biological processes. Poorly designed monitoring

programs usually do not have sufficient power to detect actual changes—i.e. they reduce apparent impacts. If this is the case then regulators may be blind to substantial impacts because the tests they apply lack statistical power. The unfortunate corollary is that there is no incentive to improve the monitoring strategy because nothing appears to be amiss.

Table 8 Examples of null hypotheses that could be tested relative to typical GMO hazards (after McIntosh, 1991).

| Hazard | Null Hypothesis |
|---|---|
| Reduction in ecological fitness | The number of engineered organisms decreases in time at the same rate as the number of non-engineered organisms |
| Cascading ecosystem effects | The release of the GMO does not affect competing species |
| Persistence in existing habitat | The GMO is not present<br>The number of engineered organisms does not decrease over time |
| Invasion of new habitats | The GMO is not moving<br>The GMO is moving at the same rate as the non-engineered organism |
| Loss of biodiversity | The abundance of engineered and non-engineered organisms is the same |

The power of a statistical test is defined as one minus the probability of a Type II error (1- ). It depends on: the statistical significance ( ); the square root of the sample size (N) used in the test; the expected effect size (ES); and the inverse of the inherent variability of the data ( ). The power of a test should always be at least 0.80. The 0.80 convention is arbitrary (in the same way that a significance level of 0.05 is arbitrary) but is widely regarded as acceptable (Murphy and Myors, 1998). Cohen (1992) lists the sample sizes needed to achieve this level of power for eight standard statistical tests and three effect sizes. Similar guidance for a much larger variety of statistical tests, however, is readily available (Cohen, 1988).

It is important to recognise that there is a trade off between the power of a statistical test (and the attendant Type II error) and the probability of a Type I error. Conventional statistical standards seek to minimise Type I errors. For example alpha levels in null hypothesis tests are usually set at 0.05 or 0.01. Type I errors cause overestimates of risk, and tend to have an increasingly disproportionate impact on the results of analysis as the events of concern become rare (Kareiva et al., 1994). On the other hand, Type II errors always cause underestimates of risk, and may therefore cause environmental harm. Precautionary approaches to risk assessment seek to minimise Type II errors (Scientists Working Group on Biosafety, 1998). Risk analysts may therefore be well advised to employ a more lenient level of statistical significance (e.g. = 0.1). This also allows a lower sample size for the same level of statistical power (Murphy and Myors, 1998).

In many field studies the analyst must also have to fit (or assume) a probability distribution model to the expected spatial pattern of organisms in order estimate the sample size (or effect size) for the desired level of statistical significance. Table 9 lists the probability distributions most commonly used by ecologists when designing sampling strategies, particularly sequential strategies. Sampling strategies are either ‘fixed’ or ‘sequential’. In the former, the number of samples that will be taken is fixed in advance of the study. In the later, sampling continues until the phenomenon of interest (e.g. population density) has been estimated with the desired level of precision.

Other important techniques that may help avoid the errors commonly associated with null hypothesis testing include confidence interval analysis and statistical process control. Plotting confidence intervals of all test statistics will often illustrate trends within repeated studies that may be hidden by mixed reports of statistical significance. Statistical process control techniques impose strict management requirements on processes that exhibit test statistics (such as the mean error rate) above or below pre-specified control limits, and do not therefore rely on statistical significance to trigger management action.

**Best practice recommendation #10: Consider statistical power, effect size, model based sensitivity analysis, and other remedies for hidden conventional pitfalls in monitoring**

Experimental design protocols should also advise on suitable controls. Controls should ideally consist of null segregated organisms, or untransformed parent organisms, making appropriate allowance for the possibility of somaclonal variation.[5] Wild controls should be selected from all the major biomes that the GMO might be released into (Linder and Schmitt, 1993). The analyst may also need to evaluate the performance of hybrids in order to be thorough. These may therefore have to be created under controlled conditions for the purpose of the test(s). Similarly it may be an effective strategy for soil tests to monitor for original products and secondary metabolites (Morra, 1994).

For reasons of sexual compatibility and hybrid fitness, the entry, survival and establishment of a new genetic entity in a native gene pool is likely to be rare and will probably take place over a period of decades rather than years (OECD, 1993). Monitoring must therefore continue over a similar period of time—i.e. well beyond the few years usually taken to assess the phenotypic performance of the GMO.

[5] GM plants regenerated by tissue culture may display a different phenotype to that of the plant from which the cells originated – this is known as somaclonal variation (Walden and Wingender, 1995).

Table 9 Probability distributions most commonly used to describe spatial pattern when designing sampling plans (after Young and Young, 1998)

| **Distribution** | **Derivation and application** |
|---|---|
| Poisson | Traditionally viewed as the expected distribution if a population of organisms is allowed to distribute randomly over a field or other habitat. Random, as used here, means that every microhabitat has an equal opportunity of being occupied by any organism. For example if plant invades a field and each plant has an equal opportunity to occupy any part of that field, the probability distribution of the number of plants in randomly chosen quadrats is Poisson. The Poisson distribution has one parameter, $\lambda$. |
| Negative binomial | Often referred to an aggregated or clumped probability distribution, implying that the spatial pattern of organisms is also aggregated (although this is not always the case). The negative binomial is widely employed in population dynamics and is thought to be the most common distribution found in insect control studies. The negative binomial has two parameters $k$ and $\mu$. $k$ is referred to as the aggregation parameter—higher values of $k$ are associated with lower aggregation. As $k$ increases the distribution approaches Poisson (random). |
| Geometric | The geometric distribution is a special case of the negative binomial where $k = 1$ and is the limiting case of Bose-Einstein statistics. The geometric distribution has only one parameter $\mu$. The variance is simply ($\mu + \mu^2$). The special relationship of the mean to the variance gives this distribution special properties. The number of samples or cells that have no organisms is always largest regardless of the size of the mean. As the mean surpasses 100 the distribution becomes so flat that the probability of observing any particular value is very close to the probability of observing any other value. This is the most probable distribution of freely moving organisms in a uniform habitat. |
| Binomial | The binomial distribution is commonly described as the probability of $X$ success in $n$ independent Bernoulli trials. A Bernoulli trial is a single test with two possible outcomes, for example, is a weed seed hybrid or not. If we select a single unit from a population and observe the characteristic of interest (or not) then we have a Bernoulli trial. Ecologists, however, are not usually interested in a single unit but rather the proportion of the population exhibiting the characteristic of interest. If $n$ units are randomly selected from the population and the number $X$ exhibiting the characteristic is recorded, the probability distribution of $X$ is binomial. |

# 3 Actual risk assessment—principles, frameworks and methods

Appendix A summarises eight risk assessment frameworks developed by various trans-national and national regulatory authorities to manage the ecological risks associated with GMOs. This section of the paper reviews the scientific principles and structure of these frameworks, together with the hazard, risk and uncertainty assessment, and monitoring methods that they identify or recommend.

## *3.1 Scientific principles and frameworks*

All of the regulatory frameworks reviewed here assess the ecological risks associated with GMOs on a case-by-case basis. In most instances, however, the frameworks allow for derogation or a "differentiated procedure" for organisms that are well known and well characterised. The Canadian regulatory directives, however, only apply to organisms that are neither familiar nor substantially equivalent to plants that are already in use in Canada and considered safe, irrespective of their genetic origins. The technical guidelines developed by the United Nations Environment Programme (UNEP) suggest (in a similar fashion) that the length and extent of the assessment should be based on the analyst(s) familiarity with the organism, and that the assessment can serve for functionally equivalent groups of species, as knowledge and experience accumulates—i.e. as one moves from inductive to deductive risk assessment.

All but two of the frameworks call for the risk to be compared to background levels posed by the equivalent unmodified organism. Half of the assessment frameworks require the assessment to be scientifically sound, accurate and/or transparent. The same proportion invokes the precautionary principle but usually in a weaker form than that originally stated in the Rio declaration—for example, lack of full scientific certainty should not be used to postpone or prevent "appropriate" or "cost effective" measures. Two of the frameworks stipulate that the development, assessment and release of GMOs should take place in a step-wise fashion, and two refer to "recognised" risk assessment techniques. The risk assessment framework developed by the Australian Office of the Gene Technology Regulator (OGTR) and the European Union (EU) Directive, 2001/18/EC, refer to "best practice" and "state of the art" methods.

The structure of most of the frameworks (six of the eight) reviewed here bears a reasonable resemblance to the quantitative risk assessment paradigm. Each has anywhere between three and six steps starting with hazard identification followed by the calculation of risk expressed as a function of the likelihood and consequences of adverse events. All of these frameworks also include the identification of management options. Only three of eight frameworks, however, identify uncertainty analysis as a separate step in the risk assessment process (see section 3.4).

The safety considerations published by the Organisation for Economic Cooperation and Development (OECD) identify the risk assessment approach developed by the US Office of Technology Assessment (OTA). This approach is process-orientated, highlighting the steps in the formation, release, proliferation and establishment of GMOs, leading ultimately to potential human or ecological impacts. This framework clearly derives from ecotoxicology—the first two stages are roughly equivalent to the risk-source characterisation stage of the toxicological risk assessment paradigm, whilst the last stage corresponds to the dose-response assessment stages. This approach is not ideally suited to biological stressors such as GMOs, however, because the intermediates stages of proliferation and establishment differ markedly from the concept of exposure as commonly understood in the ecotoxicological paradigm (Hayes, 1997).

The Canadian regulatory directives specify a different approach again based on assessment of the GM plant relative to its unmodified counterpart, its potential environmental impact and specific species replacement/competition studies. Of all the frameworks review here, this approach least resembles a formal risk assessment—indeed there is little evidence within the documents reviewed here of a coherent risk-based framework. This is perhaps best illustrated by the particular emphasis on species replacement/competition studies, as opposed to a formal, more comprehensive, evaluation of all potential hazards.

## *3.2 Hazard identification*

Of the eight regulatory approaches reviewed here, only one—New Zealand's Environmental Risk Management Authority (ERMA)—identifies a range of deductive and inductive hazard identification techniques, including brainstorming, checklists, logic trees and HAZOP (Hazard and Operability) analysis. It also lists 40 questions designed to help identify the hazards associated with the release of a GMO. The framework encourages applicants to be thorough and systematic, and to consider the widest possible range of hazards regardless of the likelihood of occurrence.

The OECD safety considerations identify fault trees and events trees as a means to quantify probability (rather than identify hazards) but otherwise only provide a de-facto checklist comprising 39 'environmental and agricultural considerations'. Four of the remaining frameworks provide similar checklists. The OGTR provides the most comprehensive list. Part 2 of the framework lists over 100 'prescribed information requirements' covering the genetics of the GMO, its production and release characteristics, its parent organism, its potential interaction with the environment and the health and safety of people, risk management details, and additional information specific to GM plants, microorganisms, aquatic organisms, etc.

Appendix II, Table 4 of the Canadian directive requires the applicant to fill in 72 cells of an 'anticipated impact matrix'. This in effect is a simple hazard analysis, although it is not referred to as such within the document. Annex II of the EC directive lists five (rather broad) potential adverse effects such as toxicity, impacts on population dynamics, altered susceptibility to pathogens and effects on biogeochemistry, but then goes on to list 18 information requirement for higher and non-higher plants. The UK Department for Environment, Food and Rural Affairs (DEFRA, formerly the UK Department of Environment Transport and Regions) guidance simply repeats the potential adverse effects and information requirements listed in the EC directive.

Two of the eight frameworks provide very little guidance on this issue: the UNEP guidelines and the Cartagena Protocol on Biosafety list hazard identification as the first step in the risk assessment process but give no further information on how this might be achieved nor discuss any of the hazard identification techniques listed in section 2.2.

## *3.3 Likelihood and consequences of events*

On the whole, the regulatory frameworks reviewed here provide very little guidance on how to assess the likelihood or consequence of GMO hazards. The Canadian framework, however, is a notable exception. Directive 2000-07 identifies three examples of quantitative replacement and seed dormancy analysis from the scientific literature—specifically Crawley et al., (1993); Linder and Schmitt, (1994) and Rissler and Melon (1993). These references include experimental protocols and simple algorithms to calculate the replacement capacity of a genetic type in a population of plants (see also Appendix B). The directive is unique in this regard, bridging the divide between regulation and science, albeit for a very limited set of possible hazards.

The ERMA and OECD guidelines note that there are various forms of mathematical analysis that may be applied to (at least some) GMO hazards depending on the quality of available data. Between them they identify event trees, fault trees, simulation, extrapolation, epidemiology and toxicology, but do not discuss specific examples or models. The ERMA guidelines contrast quantitative, semi-quantitative and qualitative approaches, relying heavily on the AS/NZS 4360:1999 (Standards Australia, 1999). The OECD guidelines simply recommend a mixed qualitative/quantitative assessment but do no acknowledge the problems that may occur when the analyst attempts to mix qualitative and quantitative expressions of risk. The OGTR risk assessment framework emphasises that some risks (but does not identify which ones) can be analysed quantitatively given sufficient information, and states that the Regulator will conduct a quantitative assessment where the data permits. It notes that applicants will be advised to use early trials to collect quantitative data, and warns that an application may be rejected if a quantitative assessment is critical to analysing a particular hazard and data are deficient. It does not, however, identify where a quantitative assessment is critical or conversely unnecessary.

The remaining frameworks do not discuss quantitative or qualitative approaches to risk calculation but rather allude to the process in quite general terms. The EC directive is typical in this regard—taken together Article 4.3 and Annex II require an accurate, case-by-case estimation of risk to be made, as far as possible, given the state of the art, by combining the likelihood of an adverse effect and the magnitude of the consequences if it occurs. No further guidance on the state of the art, however, is provided[6]. The Cartagena Protocol and DEFRA guidelines simply state that the manner of the GMO release and the characteristics of the receiving environment will be important factors in the risk calculation. The UNEP guidelines refer to forecasting models and international databases that may help in the development of models (implying a quantitative approach?) but give no other details.

### *3.4 Uncertainty analysis*

The ERMA framework contains fairly comprehensive guidance on the types of uncertainty in a risk assessment but is less comprehensive on the techniques of uncertainty calculus. It distinguishes between variability, sampling error and lack of knowledge, but only recommends the use of probability distributions to describe variability—no mention is made of the alternative techniques such as interval analysis, fuzzy arithmetic or probability bounds analysis. The framework does, however, recommend checking information for bias, statistical competence and peer review.

The OGTR framework requires the applicant to address the level of certainty in their risk estimates but gives no further guidance on how this might be achieved, or indicate the types of uncertainty the analyst may face. Chapter 4 of the framework states that the applicant may assume a worst-case scenario if uncertainty is high and implement management strategies on the basis of this assumption, but does not warn of the potential pitfalls of this approach.

Surprisingly, all of the other frameworks reviewed here give very scant regard to the uncertainty of the risk estimate (surprising in light of its critical role in risk assessment). Annex III of the Cartagena Protocol on Biosafety simply mentions uncertainty, noting that it can be addressed by requesting further information, implementing appropriate risk management and/or monitoring the GMO in the environment. The other frameworks do not mention uncertainty at all.

[6] The directive does state, however, that Annex II will be supplemented by guidance notes that will be completed by 17 October 2002.

## 3.5 Social appraisal of risk

The OGTR framework places considerable emphasis on the significance of risk and risk communication with applicants, stakeholders and communities. The significance of risk is described in terms of the number and severity of hazards, the severity, cumulation, extensiveness and scale of impacts (acute and chronic) and the extent to which they are reversible. Communication and consultation takes place throughout the risk assessment process from initial application to actual release. The Regulator is required by law to consider all submissions made in response to the prescribed consultations and where requested.

The ERMA guidelines place similar emphasis on the social appraisal of risk. Communication and consultation take place throughout the risk assessment process. Importantly the guidelines encourage applicants to engage stakeholders and interested parties at an early stage and include them in the risk identification process where possible. The significance of risk is assessed relative to its persistence, geographic spread, reversibility, the extent to which it is known or understood by society, the extent to which exposure is involuntary and the lack of experience in managing the potential impacts.

Article 9 of the EU Directive requires member states to consult the public and other interested parties of any proposal to release a GMO, and allow a reasonable time period to allow these groups to express their opinion. All information exchanged between the member state's regulatory authority and the applicant must also be made available to the public, except confidential material. The environmental risk assessment, however, cannot be kept confidential. The directive does not explicitly address risk significance issues. Similarly, the UNEP guidelines and the Cartagena Protocol (article 23) specifically provide for public awareness and participation in the assessment process, and the dissemination of information via the Biosafety Clearing House. Annex 7 of the UNEP guidelines also list ways of providing information to the public. Neither document, however, goes on to discuss the significance of risk in any detail.

The Canadian directives do not provide for an explicit social appraisal of risk. They do, however, require the applicant to consider the geographic scope and duration of potential impacts in natural and managed ecosystems (Appendix II, Table 4). None of the other frameworks reviewed here address the social appraisal of risk.

## 3.6 Monitor and review

The EC directive and DEFRA guidelines require a detailed case by case monitoring plan that takes account of the risk assessment, the characteristics of the GMO and release environment. Monitoring must also continue for a sufficient time period to detect immediate, direct impacts as well as delayed indirect impacts. Neither of these documents, however, explicitly refers to the potential problems of statistical power nor do they identify techniques to calculate Type I or Type II errors. They do state that the monitoring plan should consider mechanisms for identifying and confirming observed adverse effects. This could be construed as a reference to Type II error but at best it is very veiled reference.

The 1992 OECD guidelines are a little more specific on this issue. They state that scientifically acceptable field research requires careful experimental design, including *inter alia* the formulation of a hypothesis and methods of analysis to test for statistical significance. Again, however, there is no explicit reference to statistical power or any warning of the pitfalls commonly associated with null hypothesis tests based solely on statistical significance.

The Canadian Regulatory Directive Dir2000-07 provides species-specific terms and conditions for 'confined' field trials of GM plants. These conditions include specific isolation distances, post harvest

restrictions and monitoring requirements. Monitoring, however, is only required in order to remove volunteer plants not to gather evidence of effect or test the predictions of a risk assessment. Indeed an assessment of the environmental safety of GM plants (Directive Dir94-08) is only required for commercial 'unconfined' release which is defined as use without reproductive isolation, post harvest restrictions or monitoring by the Canadian Food Inspection Agency (CFIA). Post-release monitoring is only required if the applicant becomes aware of new information regarding risks to the environment. In assessing the environmental safety of GM plants the directives require the applicant to obtain data from experimental designs using sound statistical methods but again no mention is made of statistical power or Type II error.

The UNEP guidelines state that monitoring may vary from very simple observations to an extensive research programme, and should be used to verify whether risk management strategies are effective, but gives no information on how to actually achieve this with any degree of statistical confidence. The Cartagena Protocol recommends that the risk assessment take account of the power and reliability of methods used to detect GMOs but gives no further guidance on these issues, or the related issue of monitoring for the potential impact of a GMO relative to the predictions of the risk assessment. The ERMA and the OGTR simply state that they will monitor and inspect release sites but again give no further information or guidance. It is possible, however, that ERMA will address this issue further in a separate technical guide on ecological risk assessment, which is being prepared.

# 4 Discussion

Table 10 summarises the extent to which the risk assessment frameworks reviewed here either: a) do not mention the components of an ideal risk assessment; b) mention the component but do not give any details; c) discuss the component by dedicating at least one reasonable paragraph of text to it; or d) guide a potential applicant towards "best practice" or "state of the art" techniques as defined in this paper.

Most of the frameworks identify a fairly consistent set of scientific principles for the risk assessment, and recommend broadly similar steps in the risk assessment process. The Canadian regulatory directives, however, adopt a customised approach that is not consistent with the usual principles of risk assessment. Familiarity features prominently in at least three of the frameworks reviewed here, usually as a means to screen out organisms that do not need to be assessed. The concept of familiarity is closely linked to that of "substantial equivalence" and is based on the fact that, to date, most GMOs are developed from crop plants whose biology is well understood (OECD, 2000b). It is, however, a difficult concept to define precisely and is clearly threatened by the next generation of GMOs. It has been soundly criticised because it does not precisely specify how much information is needed to ensure that a GMO is similar enough to a conventional organism to be considered familiar (and thereby safe) and, perhaps more importantly, does not ask "what can go wrong?" with this GMO (van Dommelen, 1998; Royal Society of Canada, 2001). Furthermore, experience with invasive species indicates that the biological characteristics of species are not good predictors of invasion success, and that in many instances the difference between success and failure is determined by just a few genes (Williamson, 1994; 1996). It is therefore inappropriate to base predictions about the ecological risks of GMOs solely on familiarity as defined in terms of the number of altered genes or qualitative properties of the phenotype (Giddings, 1999). The Royal Society of Canada recently recommended that the role substantial equivalence plays in Canada's regulatory regime should be replaced with rigorous scientific assessment of the potential to cause harm at six levels: genome, transcript, protein metabolite, health impacts and environmental impacts. Environmental impacts can be further sub-divided at the levels of the phenotype, ecosystem and evolutionary landscape (Figure 2).

**Table 10 Summary of the risk assessment frameworks reviewed in this study**

| Name | Scientific Principles | Framework | Hazard Assessment | Likelihood & Consequence | Uncertainty analysis | Social appraisal | Monitor & Review |
|---|---|---|---|---|---|---|---|
| Annex III of the Cartagena Protocol on Biosafety | √√ | √√ | _ | _ | _ | √ | _ |
| UNEP International technical guidelines for safety in biotechnology | √ | √√ | _ | _ | X | √ | _ |
| Directive 2001/18/EC on the deliberate release into the environment of GMOs | √√ | √√ | √ | _ | X | √ | √ |
| OECD recombinant DNA safety considerations 1986 and 1992 | √√ | √ | √ | √ | X | X | √ |
| Canadian Food Inspection Agency regulatory directives 1994-08 and 2000-07 | X | na | √ | √√[b] | X | _ | √ |
| DETR guidance on principles of risk assessment and monitoring for the release of GMOs, 1999 | √√ | √√ | √ | _ | X | X | √ |
| OGTR risk analysis framework, November 2001 | √√ | √√ | √ | √ | _ | √√ | _ |
| ERMA Technical Guides ER-TG-01-1 9/99 and ER-TG-03-1 7/00[a] | √√ | √√ | √√ | √ | √ | √√ | _ |

X = is not mentioned

- = is mentioned

√ = is discussed

√√ = is indicative of “best practice” and/or “state of the art” techniques.

a = a separate technical guide on ecological risk assessment is being prepared.

b = for a very limited set of hazards

Hazard assessment usually comprises of simple checklists of varying length. This is clearly the "status quo" for the majority of risk assessment frameworks for GMOs. Only one of the assessment frameworks refers to inductive hazard assessment techniques. Reliance on simple prescriptive checklists sets a dangerous precedent for two reasons: 1) checklists do not force the analyst to think about what can go wrong with the system. Instead they tend to mislead the analyst into believing that all aspects of the system have been questioned without confirming this to be true; and, 2) it is not sufficient to simply list possible hazards. The analyst should properly construct a unique connected, transitive and aliorelative event series linking hazards to endpoints.

Hazards most commonly identified in these frameworks (and the scientific literature) are persistence of the GMO, invasion into new arable or non-arable environments and gene flow to other organisms that increases their persistence or invasive ability. Other hazards that are less commonly cited include impacts on non-target organisms, biocide resistance in insects, altered relations between ecosystem objects and processes, and threats to biodiversity. It is not difficult therefore to generate a checklist, such as that in Table 3, from the literature. Checklists are simple to construct and easy to use but risk analysts will inevitably miss potential hazards if future assessments do not progress beyond these lists, particularly as the next generation of GMOs comes off the drawing board. There is a large gap here between current practice and best practice.

The assessment of likelihood and consequence is characteristically poor in all of the frameworks reviewed here. This study has highlighted a wide variety of quantitative techniques and models for most of the GMO hazards listed in Table 3 and assessed the extent to which these are practical, reliable and accepted by regulators (Tables 5 and 6). Despite the rich literature and wide variety of experimental and quantitative methods there is very little evidence of "best practice" or "state of the art" techniques in any of the regulatory approaches reviewed here. The Canadian framework is the only one that directs the analyst to best practice techniques for replacement and seed dormancy analysis.

Some of the regulatory frameworks recognise that quantitative approaches are possible in certain circumstances, but neither the circumstances (i.e. which hazards) nor available techniques are identified. Canadian Regulatory Directive Dir2000-07 is the only regulatory framework that identifies specific quantitative methods in relation to specific hazards. For the main part, however, it is not clear when and how quantitative techniques are expected of the applicant. This of course does not preclude applicants from applying their own initiative to identify and apply quantitative methods. Very few applications made to the United States Department of Agriculture between 1988 and 1990, however, supported the statements they made in their ecological assessments with experimental data and quantitative methods (Wrubel et al., 1992). Whilst these conclusions are almost ten years old now (a similar study on more recent applications would certainly be instructive) qualitative methods will probably remain the mainstay of ecological risk assessment given the current regulatory frameworks. It is interesting to note, however, a clear presumption in favour of quantitative techniques in the OGTR framework.

There are a number of problems with qualitative risk estimates. In the first instance, they do not tackle uncertainty or they tackle it poorly. Risk descriptions such as "low" only provide information on the location of a variable if they are defined in numerical terms (e.g. low = 0.01), and do not provide information on the potential spread of a variable. Range estimates such as "negligible to medium" give information on spread (if defined numerically) but do not define the most likely location of an uncertain variable. These terms provide no information whatsoever about uncertainty if they are not defined numerically—indeed they are an important source of linguistic uncertainty, particularly between different stakeholders and value groups.

Qualitative assessments do not adequately address multiple risk sources—you cannot perform mathematical operations, such as sum, on qualitative assertions. You can, however, perform these mathematical operations if these terms are numerically defined. Furthermore qualitative estimates do not allow statistical hypothesis testing and power analysis, and are therefore of little use to analysts designing monitoring strategies.

As a result, qualitative risk assessments are easy to challenge and are vulnerable to other political or economic imperatives (Hayes, in review, c). As currently practiced they fail to satisfy at least two important scientific principles—transparency and repeatability. Different analysts using the same risk assessment framework and the same data will not necessarily reach the same (or similar) conclusions. This is clearly evident in the current practice—different nations are reaching different conclusions about GMO risks using similar data and assessment procedures. This is due in part to different environmental conditions in each nation, but also to differences of interpretation (OECD, 2000c), values and the level of precaution adopted within the assessment procedure. It is important to note that similar differences occur in quantitative risk assessments (see for example Stirling, 1999) but in these instances the reasons for the differences should be much more explicit—at the very least the assessment should not suffer from the same degree of linguistic uncertainty.

Quantitative risk assessments have a number of other advantages:

- they allow proper, probabilistic expressions of variability—they can capture information on the most likely location and spread of an uncertain variable;

- they quickly identify what is unknown—by just "doing it" analysts are forced to break down complex systems into their contributing parts, think very hard about what is and isn't known and to be more precise about inferences they draw from data;

- they are well suited to an iterative assessment cycle—calculate risk, collect data, ground-truth predictions, refine models, and re-calculate risk—and provide better insight into how experimental data should be collected; and,

- they can be used to compare alternative management strategies through a risk-benefit analysis.

Quantitative risk estimates, however, are not necessarily "objective". Important subjective judgements are involved in all quantitative risk-assessments—all probability-based inferences rely on a statistical model, but the choice of model is largely subjective. Even the simplest hypothesis test involves fundamentally subjective choices about the design and duration of the experiment (Berger and Berry, 1988). For this reason quantitative risk assessments may also fail on the scientific principle of repeatability. They are less likely to do so, however, because the subjective elements are better defined and much more transparent. The strength of quantitative risk assessment, as in science, lies not in its objectivity but rather in the way it exposes subjective input.

Quantitative models are also the only way to investigate potential hazards that are a function of the magnitude and spatial scale of the release—i.e. the scale-up effects that may arise following full commercial production of the organism. An organism with an advantageous trait will quickly, sometimes exponentially, increase in number particularly in arable systems (Darmency, 1994). These types of organisms will eventually be spotted in commercial operations given sufficient time. By contrast it is impossible to "guess" at the eventual rate of spread of engineered organisms based on field trials of one or two years (Manasse and Kareiva, 1991). Similarly, dominance and invasion by hybrids with slight fitness disadvantages may not be witnessed during a field trial but may still be possible, given sufficient gene flow, under commercial conditions (Gliddon, 1994; Thompson et al., in

prep). These types of events can be predicted with models well before they become apparent during full-scale commercial release of the organism.

It is clear from the analysis presented here, however, that quantitative techniques are not currently available for all of the potential hazards associated with GMOs. High consequence, high uncertainty impacts (such as the creation of new viruses or food-web impacts) are unlikely to be satisfactorily addressed by quantitative techniques in the near future. More rigorous qualitative techniques, however, including a wider social discourse and directed research, are achievable in the near term. Furthermore the degree of practicality, reliability and acceptance of quantitative techniques for less uncertain hazard scenarios varies from model to model. In general terms simple models are the most widely accepted and, when used in conjunction with a rigorous analysis of uncertainty, can provide meaningful answers for risk assessment purposes. Qualitative assessments are often recommended as an initial screen to eliminate low risk events from a potentially lengthy assessment process. This review, however, suggests quite the opposite: simple quantitative techniques should be used where-ever possible to screen high and low risk scenarios—qualitative assessments become most important for highly uncertain but potentially high impact scenarios. To be successful these qualitative assessments must have a strong element of social appraisal including, for example, the use of systematic hazard identification techniques (section 2.2) to capture the imagination and intuition of non-scientific ‘experts’.

The consistently poor treatment of uncertainty in international and national risk assessment frameworks (Table 10) further underscores the importance of this last point. None of the eight frameworks reviewed here provides any evidence of best practice, and only two discuss this issue. This is without doubt the most important failing of current practice because uncertainty analysis is the very rationale of risk assessment. Furthermore the range of uncertainty calculi reviewed here provide a means to bridge qualitative and quantitative assessment, allowing the analyst to avoid the unnecessary polarisation of one or the other. Interval analysis, for example, is capable of translating a qualitative risk estimate into a quantitative one, so long as terms such as low, medium, high, etc are defined on an interval. Probability bounds analysis is able to extend the risk estimate to include measures of central tendency and spread as and when data is made available to the assessment. Best practice risk assessment should therefore strive to progressively convert qualitative assessments into quantitative ones using the iterative cycle of assess, monitor, collect data and re-assess (Figure 2).

All of the frameworks reviewed here discuss or at least mention monitoring but none points to best practice in this area. All of the frameworks could be improved by drawing the analyst’s attention to power calculations for typical monitoring strategies. Monitoring must provide the data and information that completes the iterative risk assessment loop. Field trials should therefore collect information that allows accurate predictions of risk (OECD, 1992). In reality this may not happen for a number of reasons, particularly if biosafety issues are not the main objective of the monitoring programme. For example, 1180 field locations in OECD member nations were sown with GM crops between 1986 and 1992. These trails, however, failed to provide any new information on antibiotic resistance, gene flow or the biosafety implications of geographical location and climate, and in many cases post-trial monitoring was too short to test other potential hazards (OECD, 1993). Similarly despite clear demonstrations that much of the biology of GM canola had been misdescribed, early risk assessments in the United Kingdom were not revisited or revised (Williamson, 1996). Again these conclusions are dated now, but they do point to potential flaws in the regulatory regime that, if persistent, will seriously undermine the risk assessment process.

# 5 Summary and recommendations

The hallmarks of high quality scientific investigation are accuracy, thoroughness, rigorous analysis, experimentation whenever possible, quantification whenever possible and common sense (Scientists Working Group on Biosafety, 1998). The hallmarks of high quality, best practice ecological risk assessment moulds all of these with effective social discourse on the nature of the technology and the limitations of the science.

Current practice in GMO risk assessment appears to be failing on a number of these fronts:

1. hazard identification as currently practiced is largely restricted to prescriptive checklists. Analysts will identify a larger range of potential hazards, and gain a better understanding of the event chains associated with these hazards, if they used inductive hazard identifications techniques and consulted non-scientific experts;

2. uncertainty analysis is the very rationale of risk assessment, and yet this is by far the weakest component of current practice. Well-established statistical techniques exist to describe random measurement error and environmental variability. Model error can be approached by ground-truthing risk assessment predictions and testing alternative model formulations. Techniques also exist that bridge the divide between qualitative and quantitative approaches to risk assessment, and thereby facilitate a progression from one to the other;

3. regulators can assist quantitative risk assessment by helping proponents identify models and analysis techniques relative to specific GMO hazards. Regulators should insist that proponents obtain the necessary data and information in order to achieve best practice and to reduce areas of significant uncertainty. Current field trials only appear to gather information on crop performance. These trials are an ideal opportunity to gather the types of data needed to improve the science of GMO risk assessment;

4. well corroborated quantitative techniques exist for some of the potential hazards associated with GMO field release. There are appear to be important gaps, however, in the following areas: food-web and trophic interactions, the transfer of viral particles to other viruses, increases in the host range of viruses, fungi and other pathogens, altered farm practice and physical habitat changes. National regulatory authorities should encourage data collection and research in these areas;

5. monitoring strategies will need to continue well beyond the usual period needed to assess the efficacy of the phenotype in order to detect potential ecological impacts. It is important that these strategies test the predictions of prior risk assessments and provide information that will inform future risk assessments, thereby “closing the regulatory loop”. Furthermore these strategies must explicitly include an appropriate power analysis to avoid blindness to Type II error; and,

6. regulators should continue to monitor the quality of GMO risk assessments relative to the continual improvements in state of the art techniques, and the quality and quantity of data generated by laboratory tests and field trials.

## Acknowledgements

This report is based on the proceedings of the workshop on best practice risk assessment for genetically modified organisms, held at the CSIRO division of entomology, Canberra, on the 4th and 5th of April 2002 (http://crimp.marine.csiro.au/ERAworkshop.htm). I am indebted to Mark Burgman, Lev Ginzberg, Scott Ferson and Paul Barnes for their contributions to the workshop. I would also like to thank Mark Burgman, Nic Bax, Peter Stoutjesdijk and David MacDonald for their comments on earlier drafts of this report. This study was partially funded by Environment Australia as a measure under the Commonwealth's National Biotechnology Strategy.

## References


Andow, D. A., 1994. Community response to transgenic plant release: Using mathematical theory to predict effects of transgenic plants. *Molecular Ecology*, 3: 65-70.

Asian Development Bank (1990). Environmental Risk Assessment: Dealing with uncertainty in environmental impact assessment, ADB Environment Paper No. 7, Office of the Environment, Asian Development Bank, 182 pp.

Barnthouse, L. W., Suter, G. W., Bartell, S. M., Beauchamp, J. J., Gardner, R. H., Linder, E., O'Neill, R. V. and Rosen, A. E. (1986). Users Manual for Ecological Risk Assessment. NTIS DE86-010063, ORNL-6251, National Technical Information Service (NTIS), Springfield, USA, 207 pp.

Beyer W. N. and Linder G. (1995). Making sense of soil ecotoxicology, pp. 104-116. In: Hoffman D. J., Rattner B. A., Burton G. A. and Cairns J. (Eds), Handbook of Ecotoxicology. Lewis Publishers, Boca Raton, USA,.

Beer, T. and Ziolkowski, F. (1995). Environmental Risk Assessment: An Australian Perspective, Supervising Scientist Report 102, Office of the Supervising Scientist, Canberra, Australia, 125 pp.

Berger J. O., Berry D. A. (1988). Statistical analysis and the illusion of objectivity, *American Scientist*, 76:159-165.

Burgman M. H. (2002). Flaws in subjective assessments of ecological risks and means for correcting them. *Australian Journal of Environmental Management*, 8: 219-226.

Burgman, M. A., Ferson, S. and Akcakaya, H. R. (1993). Risk Assessment in Conservation Biology. Chapman and Hall, London, England.

Cairns J. and Pratt J. R. (1986). Ecological consequence assessment: Effects of bioengineered organisms. In: Fiksel J. and Covello V. T. (Eds) Biotechnology risk assessment: Issues and methods for environmental introductions, Pergamon Press, New York, USA, pp.88-108.

Canadian Food Inspection Agency (2001a). Regulatory Directive Dir94-08: Assessment criteria for determining environmental safety of plants with novel traits, http://inspection.gc.ca/english/plaveg/pbo/dir9408.shtml

Canadian Food Inspection Agency (2001b). Regulatory Directive 2000-07: Guidelines for the environmental release of plants with novel traits within confined field trials in Canada, http://www.inspection.gc.ca/english/plaveg/pbo/dir0007e.shtml

Cardwell, R. D., (1989). An overview of aquatic ecological risk assessment methodologies. Oceans '89: The Global Ocean, Volume 2: Ocean Pollution.

Cohen J. (1988). Statistical Power Analysis for the Behavioural Sciences. Lawrence Erlbaum Associates, Hillsdale, New Jersey, USA, 567 pp.

Cohen J. (1992). A power primer. *Psychological Bulletin*, 112(1): 155-159.

Cox D. R. and Hinkley D. V. (1974). Theoretical Statistics. Chapman and Hall, London, England.

Crawley M. J. (1990). The ecology of genetically engineered organisms: Assessing the environmental risks. In: Mooney, H. A. and Bernardi, G. (Eds), Introduction of Genetically Modified Organisms into the Environment, pp.133-150. John Wiley & Sons, Chichester, England.

Crawley M. J. and Brown S. L. (1995). Seed limitation and the dynamics of feral oilseed rape on the M25 motorway. *Proceedings of the Royal Society of London Series* B, 259: 49-54.

Crawley M. J., Hails R. S., Rees M., Kohn D. and Buxton J. (1993). Ecology of transgenic oilseed rape in natural habitats. *Nature*, 363: 620-623.

Darmency H. (1994). The impact of hybrids between genetically modified crop plants and their related species: introgression and weediness, *Molecular Ecology*, 3:37-40.

Department for Environment, Food and Rural Affairs (DEFRA, formerly the Department of the Environment, Transport and Regions) (1999). Guidance on principles of risk assessment and monitoring for the release of genetically modified organisms, DETR/ACRE guidance note 12, Department of the Environment, Transport and Regions, London, UK .

Endler J. A., (1986). Natural Selection in the Wild. Princeton University Press, Princeton, New Jersey, USA.

Environmental Risk Management Authority (1999). Identifying risks for applications under the Hazardous Substances and New Organisms Act 1996, ER-TG-01-1 9/99, ERMA, New Zealand .

Environmental Risk Management Authority (2000). Preparing information on risks, costs and benefits for applications under the Hazardous Substances and New Organisms Act 1996, ER-TG-03-1 7/00, Environmental Risk Management Authority, New Zealand, 43 pp.

Epanechnikov V. A. (1969). Non-Parametric Estimation of Multivariate Probability Density. *Theory of Applied Probability and its Applications*, 14:153-158.

European Commission (2001). Directive 2001/18/EC of the European Parliament and of the Council of 12 March 2001 on the deliberate release into the environment of genetically modified organisms and repealing Council Directive 90/220/EEC. *Official Journal of the European Communities*, L106(17.04.2001): 1-38.

Faber M., Manstetten R. and Proops J. (1992). Toward an Open Future: Ignorance, Novelty and Evolution, In Costanza R., Norton B. G. and Haskell B. D. (Eds), Ecosystem Health: New Goals for Environmental Management, pp.72-96, Island Press, Washington D. C., USA.

Ferson S. (1994). Software review. *Risk Analysis*, 14(6): 1123.

Ferson S., Cooper J. A. and Myers D. (2001). Beyond Point Estimates—Risk Assessment Using Interval, Fuzzy and Probabilistic Arithmetic. SRA Workshop Booklet, Applied Biomathetics, New York, USA.

Ferson S. and Ginzberg L. R. (1996). Different methods are needed to propagate ignorance and variability. *Reliability Engineering and System Safety*, 54: 133-144.

Ferson S. and Long T. F. (1995). Conservative uncertainty propagation in environmental risk assessment, In: Hughes J. S., Biddinger G. R. and Mones E. (Eds), Environmental Toxicology and

Risk Assessment – Third Volume, pp. 97-110. ASTM STP 1218, American Society for Testing and Material, Philadelphia, USA.

Fiksel J. R. and Covello V. T. (1985). The suitability and applicability of risk assessment methods for environmental application of biotechnology. In: Fiksel J. and Covello V. (Eds) Biotechnology Risk Assessment: Issues and Methods for Environmental Introductions, Pergamon Press, New York, USA.

Finkel A. M. (1990). Confronting Uncertainty in Risk Management: A Guide for Decision Makers. Centre for Risk Management Resources and the Future, Washington DC, USA, 68 pp.

Gaines S. D. and Denny M. W. (1993). The largest, smallest, highest, lowest, longest and shortest: extremes in ecology. *Ecology*, 74(6): 1677-1692.

Gates, C. E., (1979). Line transect and related issues. In: Cormack, R. M., Patil G. P. and Robson, D. S., (Eds), Sampling Biological Populations. International Co-operative Publishing House, Fairland, Maryland, USA, pp. 71 – 154.

Giddings, G. D., (1999). The role of modelling in risk assessment for the release of genetically engineered plants. In: Ammann, K., Jacot, Y., Simonsen, V. and Kjellsson, G. (Eds), Methods for Risk Assessment of Transgenic Plants III—Ecological Risks and Prospects of Transgenic Plants. Birkhauser-Verlag, Basel, Switzerland, pp.31-41.

Giddings G. D., Sackville Hamilton N. R. and Hayward M. D., (1997a). The release of genetically modified grasses. Part 1: Pollen dispersal to traps in *Lolium perenne*. *Theoretical and Applied Genetics*, 94: 1000-1006.

Giddings G. D., Sackville Hamilton N. R. and Hayward M. D., (1997b). The release of genetically modified grasses. Part 2: The influence of wind direction on pollen dispersal. *Theoretical and Applied Genetics*, 94: 1007-1014.

Gliddon C. (1994). The impact of hybrids between genetically modified crop plants and their related species: biological models and theoretical perspectives, *Molecular Ecology*, 3: 41-44.

Gumbel E. J. (1962). Statistical theory of extreme values (main results), In: Sarhan, A. E. and Greenberg, B. G. (Eds), Contributions to Order Statistics, pp.57-93. John Wiley & Sons Inc., New York, USA.

Haimes Y. Y. (1998). Risk Modelling, Assessment and Management. John Wiley & Sons Inc., New York, USA, 726 pp.

Hallerman E., King D. and Kapuscinski A. (1999). A decision support software for safely conducting research with genetically modified fish and shellfish. *Aquaculture*, 173: 309-318.

Hayes K. R. (1997). A Review of Ecological Risk Assessment Methodologies. CSIRO CRIMP Technical Report Number 13, CSIRO Division of Marine Research, Hobart, Australia, 116 pp.

Hayes K. R. (1998), Bayesian Statistical Inference in Ecological Risk Assessment. CSIRO CRIMP Technical Report Number 13, CSIRO Division of Marine Research, Hobart, Australia, 94 pp.

Hayes K. R. (in review, a). Identifying hazards in complex ecological systems—Part 1: Fault tree analysis for ballast water introductions. *Bioinvasions*.

Hayes K. R. (in review, b). Identifying hazards in complex ecological systems - Part 2: Infection Modes and Effects Analysis for small craft. *Bioinvasions*.

Hayes, K. R. (in press). Biosecurity and the role of risk-assessment. In: Ruiz G. M. and Carlton J. T., (Eds) Bioinvasions: Pathways, Vectors, and Management Strategies. Island Press, Washington, D.C., USA.

Hayes K. R. and Hewitt C. L. (2001). Risk Assessment Framework for Ballast Water Introductions - Volume II. CRIMP technical report number 21, CSIRO Division of Marine Research, Hobart, Australia, 197 pp.

Hoffman D. J., (1995). Wildlife Toxicity Testing. In: Hoffman D. J., Rattner B. A., Burton G. A. and Cairns J. (Eds), Handbook of Ecotoxicology, pp. 47-69, Lewis Publishers, Boca Raton, USA.

Hokanson S. C., Hancock J. F. and Grumet R. (1997). Direct comparison of pollen-mediated movement of native and engineered genes. *Euphytica*, 96: 397-403.

Jacocks J. L. and Kneile K. R. (1974). Statistical Prediction of Maximum Time-Variant Inlet Distortion Level. Arnold Engineering Development Centre Technical Report AD/A-004, National Technical Information Service, United States Department of Commerce, Washington D.C., USA, 74 pp.

Jepson, P. C., Croft, B. A. and Pratt, G. E. (1994). Test systems to determine the ecological risks posed by toxin release from Bacillus thuringiensis genes in crop plants. *Molecular Ecology*, 3: 81-89.

Johnson N. L., Kotz S. and Balakrishnan N. (1995). Continuous Univariate Distributions – Volume 2. John Wiley and Sons Inc., New York, USA, 719 pp.

Kaplan S. (1997). The words of risk analysis, *Risk analysis*, 17(4): 407-417.

Kappeli O. and Auberson L. (1998). Planned releases of genetically modified organisms into the environment: the evolution of safety considerations, *Chimia*, 52: 137-142.

Kareiva P., Morris W. and Jacobi C. M. (1994). Studying and managing the risk of cross-fertilisation between transgenic crops and wild relatives. *Molecular Ecology*, 3: 15-21.

Kaufman A, and Gupta M M. (1985). Introduction to Fuzzy Arithmetic. Van Nostrand Reinhold Company, New York, USA, 351 pp.

Kim, J., Ginzburg, L. R. and Dykhuizen, D. E. (1991). Quantifying the risks of invasion by genetically engineered organisms. In Ginzberg, L. R. (Ed), Assessing Ecological Risks of Biotechnology, Butterworth-Heinemann, Boston, USA , pp.193-214.

Kjellsson G., Simonsen V. and Ammann K. (1997). Methods for Risk Assessment of Transgenic Plants II. Pollination Gene-Transfer and Population Impacts, Birkhauser Verlag, Basel, Switzerland, 308 pp.

Kjellsson G. and Simonsen V. (1994). Methods for Risk Assessment of Transgenic Plants I. Competition, Establishment and Ecosystem Effects, Birkhauser Verlag, Basel, Switzerland, 214 pp.

Koojiman, S. A. L. M. (1987). A safety factor for $LC_{50}$ values allowing for differences in sensitivity among species. *Water Research* 21, 269-276.

Landis, W. G., Lenart, L. A. and Spromberg, J. A. (2000). Dynamics of horizontal gene transfer and the ecological risk assessment of genetically engineered organisms. *Human and Ecological Risk Assessment*, 6(5): 875-899.

Lavigne C., Klein E. K., Vallee P., Pierre J., Godelle B. and Renard M. (1998). A pollen dispersal experiment with transgenic oilseed rape. Estimation of the average pollen dispersal of an individual plant within a field. *Theoretical and Applied Genetics*, 96: 886-896.

Li H. W., Rossignol P. A. and Castillo G. (1999). Risk analysis of species introductions: Insights from qualitative modelling. In: Claudi R. and Leach J. H. (Eds), Nonindigenous Freshwater Organisms - Vectors, Biology and Impacts, Lewis Publishers, Boca Raton, USA, pp.431-447.

Linder C. R. and Schmitt J. (1994). Assessing the risks of transgene escape through time and crop - wild hybrid persistence. *Molecular Ecology*, 3: 23-30.

Lenski R. E. (1991). Quantifying fitness and gene stability in microorganisms. In Ginzberg, L. R. (Ed), Assessing Ecological Risks of Biotechnology, Butterworth-Heinemann, Boston, USA, pp.173-192.

Manasse R. and Kareiva P. (1991). Quantifying the spread of recombinant gene and organisms.In: Ginzberg L. R. (Ed), Assessing Ecological Risks of Biotechnology, Butterworth-Heinemann, Boston, USA, pp. 215-231.

McIntosh M. A. (1991). Statistical techniques for field testing genetically engineered micro-organisms. In: Levin M. A. and Strauss H. S. (Eds), Risk Assessment in Genetic Engineering. McGraw-Hill, New York, USA, pp. 219-239.

Morgan M. G. and Henrion M. (1990). Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge University Press, Cambridge, England, 332 pp.

Morra M. J. (1994). Assessing the impact of transgenic plant products on soil organisms, *Molecular Ecology*, 3: 53-55.

Murphy K. R. and Myors B. (1998). Statistical Power Analysis, Lawrence Erlbaum Associates, Mahwah, New Jersey, USA, 120 pp.

National Research Council (1989). Conclusions and recommendations: plants, Field Testing Genetically Modified Organisms: Framework for Decisions, pp. 70-76.

National Research Council (1989). Conclusions and recommendations: micro-organisms, Field Testing Genetically Modified Organisms: Framework for Decisions, pp. 123-131.

Office of the Gene Technology Regulator (2001). Risk assessment framework for licence applications to the office of the gene technology regulator, November 2001, Office of the Gene Technology Regulator, Canberra, Australia, 80 pp.

Organisation for Economic Cooperation and Development (1986). Recombinant DNA safety considerations, Organisation for Economic Cooperation and Development, Paris, France pp.74 pp.

Organisation for Economic Cooperation and Development (1992). Safety considerations for biotechnology, Organisation for Economic Cooperation and Development, Paris, France, 45 pp.

Organisation for Economic Cooperation and Development (1993). Field Releases of Transgenic Plants, 1986-1992: An Analysis. Organisation for Economic Cooperation and Development, Paris, France, 39 pp.

Organisation for Economic Cooperation and Development (2000a). Summary of Data from OECD's Database of Field Trials, http://www.oecd.org/ehs/summary.htm

Organisation for Economic Cooperation and Development (2000b). Report of the Task Force for the Safety of Novel Foods and Feeds, C(2000) 86/ADD1. Organisation for Economic Co-Operation and Development, Paris, France, 72 pp.

Organisation for Economic Co-operation and Development (2000c). Report of the working group on harmonisation of regulatory oversight in biotechnology, C(2000) 86/ADD2, Organisation for Economic Co-operation and Development, Paris, France, 65 pp.

Palisade (1996), Guide to Using @RISK. Palisade Corporation, Newfield, USA, 307 pp.

Parker I. M. and Kareiva P. (1996). Assessing the risks of invasion for genetically engineered plants: Acceptable evidence and reasonable doubt. Biological Conservation, 78: 193-203.

Parkhurst B., Warren-Hicks W., Cardwell R., Volosin J., Etchison T., Butcher J. and Covington S. (1995). Aquatic ecological risk assessment aids decision making. *Water Environment and Technology*, 7(11): 39-43.

Pastorok R. A., Bartell S. M., Ferson S. and Ginzburg L. R. (2002). Ecological Modelling in Risk Assessment: Chemical Effects on Populations, Ecosystems and Landscapes. Lewis Publishers, Boca Raton, USA, 302 pp.

Patil G. P., Boswell M. T., Joshi S. W., Ratnaparkhi M. V. (1984a). Dictionary and Classified Bibliography of Statistical Distributions in Scientific Work - Volume 1: Discrete Models, International Co-Operative Publishing House, Maryland.

Patil G. P., Boswell M. T., Joshi S. W., Ratnaparkhi M. V. (1984b). Dictionary and Classified Bibliography of Statistical Distributions in Scientific Work - Volume 2: Continuos Univariate Models, International Co-Operative Publishing House, Maryland.

Pimm S. (1982). Food Webs. Chapman and Hall, London, England.

Pollard S. J. T. (2001). An overview of the use of risk assessment for environmental regulation in the UK - key drivers and regulatory initiatives. *Risk Decision and Policy*, 6: 33-46.

Pretty J. (2001). The rapid emergence of genetic modification in world agriculture: contested risks and benefits. *Environmental Conservation*, 28(3): 248-262.

Reckhow K. H. (1994). Water quality simulation modelling and uncertainty analysis for risk assessment and decision making. *Ecological Modelling*, 72: 1-20.

Reckhow C. H. and Chapra S. C. (1983). Confirmation of water quality models. *Ecological Modelling*, 20: 113-133.

Regan H. M., Colyvan M. and Burgman M. A. (2002). A taxonomy and treatment of uncertainty of ecology and conservation biology. *Ecological Applications*, 12(2): 618-628.

Rissler J. and Mellon M. (2000). The Ecological Risks of Engineered Crops. The MIT Press, Cambridge, Massachusetts, USA, 168 pp.

Royal Commission on Environmental Pollution (1991). GENHAZ: A system for the Critical Appraisal of Proposals to Release Genetically Modified Organisms into the Environment, HMSO, London, England, 55 pp.

Royal Society of Canada (2001). Elements of Precaution: Recommendations for the Regulation of Food Biotechnology in Canada, RSC.EPR 01 - 1, Royal Society of Canada, Ottowa, Canada pp.245.

Russell, B. (1993). Introduction to Mathematical Philosophy. Routledge, London, England, 208 pp.

Scientists Working Group on Biosafety, (1998). Manual for Assessing Ecological and Human Health Effects of Genetically Engineered Organisms. Part One: Introductory Materials and Supporting Text for Flowcharts. The Edmunds Institute, Washington, USA, 158 pp.

Silverman B. W. (1986) Density Estimation for Statistics and Data Analysis. Chapman and Hall, London.

Silverman B. W. (1978). Choosing the window width when estimating a density. *Biometrika*, 65(1): 1-11.

Slatkin M. and Barton N. H. (1989). A comparison of three indirect methods for estimating average levels of gene flow, Evolution, 43(7):1349-1368.

Standards Australia, (1999). Risk Management. AS/NZS 4360: 1999, Standards Association of Australia, Strathfield, Australia, 44 pp.

Stirling A. (1999). On Science and Precaution in the Management of Technological Risk. Institute for Prospective Technological Studies, Seville, Spain, 56 pp.

Strauss H. S. and Levin M. A. (1991). Use of fate and transport (dispersal) models in microbial risk assessment, pp.240-271in Levin M. A. and Strauss H. S. (Eds), Risk Assessment in Genetic Engineering, McGraw-Hill, New York, USA.

Strauss H. S. (1991). Lessons from chemical risk assessment. In: Levin M. A. and Strauss H. S. (Eds), Risk Assessment in Genetic Engineering, McGraw-Hill, New York, USA.

Strauss H., Hattis D., Page G. S., Harrison K., Vogel S. and Caldart C. C. (1985). Direct Release of Genetically Engineered Microorganisms: A Preliminary Framework for Risk Evaluation under TSCA. CTPID 85-3, Center for Technology, Policy and Industrial Development, Massachusetts Institute of Technology, Massachusetts, USA.

Suter G. W. (1993). Ecological Risk Assessment. Lewis Publishers, Michigan, USA.

Suter,G. W. (1990). Endpoints for regional ecological risk assessment. *Environmental Management*, 14(1): 9-23.

The Royal Society (1983). Risk Assessment: Report of a Royal Society Study Group. The Royal Society, London, England, 75 pp.

Thompson C. J., Thompson B. J. P. and Burgman M. A. (in prep). Risks from competitively inferior immigrant populations: Implications of mass for species conservation.

Tomiuk J. and Loeschcke V. (1993). Conditions for the establishment and persistence of populations of transgenic organisms. In: Wohrmann F. and Tomiuk J. (Eds), Transgenic Organisms: Risk Assessment of Deliberate Release, Birkhauser Verlag, Berlin, Germany, pp.117-133.

Trevors J. T., Kuikman P. and Watson B. (1994). Transgenic plants and biogeochemical cycles. *Molecular Ecology*, 3: 57-64.

Ulanowicz R. E. (1992). Ecosystem health and trophic flow networks. In: Costanza R. et al., (Eds), Ecosystem Health: New Goals for Environmental Management. Island Press, Washington DC, USA, pp.190-206.

United Nations Environment Programme (1995). International technical guidelines for safety in biotechnology, United Nations Environment Programme, Nairobi, Kenya.

van Dommelen A. (1998). Useful models for biotechnology hazard identification: What is this thing called familiarity? In: Wheale P., von Schomberg R. and Glasner P. (Eds), The social Management of Genetic Engineering. Ashgate, Aldershot, England, pp.219-236.

van Straalen N. M. and Denneman C. A. J. (1989). Ecotoxicological evaluation of soil quality criteria, Ecotoxicology and Environmental Safety, 18: 241-251.

Vose D. (2000). Risk analysis: A Quantitative Guide. John Wiley and Sons Ltd., Chichester, England.

Walden R. and Wingender R. (1995). Gene-transfer and plant-regeneration techniques, Tibtech, 13: 324-331.

Williamson M. (1996). Can the risks from transgenic crop plants be estimated? Trends in Biotechnology, 14: 449-450.

Williamson M. (1994). Community response to transgenic plant release: predictions from British experience of invasive plants and feral crop plants. *Moelcular Ecology*, 3:75-79.

Williamson M., (1989). Mathematical models of invasion. In: Drake J. A. (Ed), Biological Inavsions: A Global Perspective, SCOPE 37. John Wiley & Sons Ltd., New York, USA, pp. 329-350.

Williamson R. C. and Downs T. (1990). Probabilistic arithmetic I. Numerical methods for calculating convolutions and dependency bounds. International Journal of Approximate Reasoning, 4: 89-158.

Wootton J. T. (1994). Predicting direct and indirect effects: An integrated approach using experiments and path analysis. *Ecology*, 75(1): 151-165.

Wolfenbarger L. L. and Phifer P. R. (2000). The ecological risks and benefits of genetically engineered plants. *Science*, 290: 2088-2093.

Wrubel R. P., Krimsky S. and Wetzler R. E. (1992). Field testing transgenic plants: An analysis of the US Department of Agriculture's environmental assessments. *BioScience*, 42(4): 280-289.

Young L. J. and Young J. H. (1998). Statistical Ecology. Kluwer Academic Publishers, Boston, USA, 565 pp.

## *Appendix A1 Cartagena Protocol on Biosafety*

The Parties to the Convention of Biological Diversity adopted the Cartagena Protocol on the 29th January 2000. Annex III of the protocol describes the general principles, methodology and points to consider when conducting a risk assessment for GMOs intended for direct use as food or feed, or for processing (Articles 11 and 15).

### *Scientific principles and framework*

The risk assessment is to be carried out in a scientifically sound manner, in accordance with Annex III and taking into account recognised risk assessment techniques, expertise of, and guidelines developed by, relevant international organisations. The risk assessment is to be carried out on a case-by-case basis and should be compared to the background risk posed by the equivalent non-modified recipients or parental organisms. The protocol specifies a slightly weaker version of the precautionary principle than that adopted in Agenda 21 by the United Nations Conference on Environment and Development: lack of full scientific certainty should not prevent a party to the protocol from taking an "appropriate" decision.

The protocol defines a six-step risk assessment procedure. The first step is hazard identification followed by an assessment of the likelihood and consequences of adverse effects. The fourth step is risk calculation followed by a recommendation as to whether or not the risks are acceptable or manageable. Finally, where there is uncertainty about the level of risk this should be addressed by obtaining more information, management and/or monitoring.

### *Hazard identification*

The protocol does not recommend or detail any specific hazard identification techniques. It simply states that the analyst(s) identify adverse effects on human health and biological diversity associated with novel genotypic and phenotypic characteristics of the organism concerned.

### *Likelihood and consequence assessment*

The protocol does not recommend or identify any specific techniques to identify or quantify the likelihood or consequences of adverse effects. It does, however, state that the likelihood assessment should be based on the level and kind of exposure in the likely potential receiving environment. It recommends that the receiving environment be described in terms of its location, climate, ecological characteristics, biological diversity and centres of origin.

### *Uncertainty and significance analysis*

The protocol does not recommend or identify any techniques to analyse uncertainty, other than obtaining more information or managing/monitoring the problem. The significance of risk is only expressed in terms of how acceptable or manageable they are.

### *Monitor and review*

The protocol does not discuss monitoring techniques other than as a means to deal with uncertainty in the level of risk (see above). It does, however, recommend that the risk assessment take account of the specificity, sensitivity and reliability of methods used to detect and identify the GMO.

### *Appendix A2 UNEP International Technical Guidelines for Safety in Biotechnology*

The United Nations Environment Programme launched its International Technical Guidelines for Safety in Biotechnology in December 1995, four years before the Cartagena Protocol was adopted. The guidelines were developed to provide a common framework for biotechnology safety assessment without prejudice, but as a complement, to the protocol.

#### *Scientific principles and framework*

Familiarity features prominently in the UNEP guidelines. They suggest, for example, that the extent and length of the risk assessment should depend on the analyst(s) familiarity with the organism concerned. Initially risk assessments should be conducted on a case-by-case basis, but as knowledge and experience evolve, the assessment may serve for a functionally equivalent group of organisms. The assessment should be carried out in a scientifically sound manner. The guidelines anticipate that, in most cases, the ecological risks be low from well-known crop plants (that have been modified by altering or adding only a few genes) introduced into arable environments. Risks that are identified should be compared to the background risks associated with non-modified organisms.

The guidelines go on to state that the risk assessment can range from a routine ad hoc judgement by the analyst to adherence to a formalised procedure. It specifies a three-step risk assessment procedure starting with hazard identification. If hazards are identified the assessment proceeds by calculating risk as the combined effect of the consequences and likelihood of the hazard being realised. Finally management strategies, commensurate with the level of risk, should be designed and implemented.

#### *Hazard identification*

The guidelines do not identify or recommend any hazard identification techniques.

#### *Likelihood and consequence assessment*

The guidelines do not identify or discuss any specific techniques to assess the likelihood or consequences of hazards identified. They do, however, state that forecasting models could be developed in the future that may help the assessment, and that international databases help in the development of models—implying that quantitative approaches might be appropriate? More explicitly they only suggest that full regard be given to experience with the organism elsewhere, relevant literature and consultation with available experts and public authorities.

#### *Uncertainty and significance analysis*

The guidelines do not refer to uncertainty within the risk assessment process, or the significance of the risk estimates.

#### *Monitor and review*

The guidelines state that monitoring (that may vary from a very simple observation to an extensive research programme) may be used to verify the assumptions of the risk assessment, and should be used to evaluate the efficacy of risk management measures. No further information is given.

### *Appendix A3 EC Directive 2001/18/EC on the Deliberate Release into the Environment of GMOs*

This directive repeals the earlier Council Directive 90/220. It aims, *inter alia*, to provide a common European-wide methodology for ecological risk assessment and common objectives for monitoring GMO releases to the environment. It requires an ecological risk assessment, in accordance with Annex III, prior to any deliberate release of GMOs. Guidance notes on the risk assessment procedure and monitoring plans will be completed by 17th October 2002.

#### *Scientific principles and framework*

In accordance with the precautionary principle, the potential direct, indirect, immediate, delayed and cumulative effects of GMOs are to be accurately assessed, on case by case basis. Releases are to be carried out in a stepwise fashion and must be field-tested in ecosystems that could be affected by their use. A differentiated procedure is permitted for GMOs that are well known and well characterised.

The directive specifies a six-stage risk assessment process starting with hazard identification. An evaluation of potential consequences and likelihood of adverse effects, and an estimation of the risk follow this. The estimation of risk is to be made as far as possible given the "state of the art". The fifth step identifies management options, followed by an evaluation of the overall risk taking management into account.

#### *Hazard identification*

The directive does not identify or recommend any inductive hazard assessment techniques. It notes that potential adverse effects will vary from case to case and lists five generic hazards such as toxicity, impacts on population dynamics, altered susceptibility to pathogens and effects on biogeochemistry.

#### *Likelihood and consequence assessment*

The directive does not identify or discuss any specific techniques to assess the likelihood or consequences or adverse effects. It simply notes that the environment into which the GMO is released and the manner of the release are major factors.

#### *Uncertainty and significance analysis*

The directive does not refer to uncertainty within the risk assessment process, or the significance of the risk estimates.

#### *Monitor and review*

The directive details the objectives, principles and design requirements of a monitoring plan. The objective of the plan is to confirm the assumptions made in the risk assessment and to identify the occurrence of adverse effects that were not anticipated in the assessment. The plan is to incorporate general surveillance for unanticipated effects as well as specific monitoring for those effects identified in the assessment. The latter must be continued for a sufficient period of time to identify delayed and indirect effects. The plan must be implemented in a systematic manner and consider mechanisms for identifying and confirming any observed effects.

### *Appendix A4 OECD Safety Considerations for Biotechnology, 1986 and 1992*

The 1986 report was the first attempt to set international safety guidelines for industrial, agricultural and environmental applications of biotechnology. It presents scientific principles that could underlie risk management for the release of GMOs into the environment. The 1992 report follows on from this and *inter alia* defines "Good Development Principles" for the design of safe, small-scale field trials of GM plants and microorganisms.

#### *Scientific principles and framework*

Proposals to release GMOs should be considered on a case-by-case basis. This is not meant to imply, however, that every case will require review since various classes of proposal may be excluded. The development and assessment of GMOs should take place in a step-wise fashion moving from the laboratory to the greenhouse, to small-scale field trials and then large-scale field trials. Each step in the process should generate information to predict the safety of the next step. Safety concerns should focus on whether GMOs pose an "incremental risk" above and beyond the background risks of conventional agriculture.

The 1986 report identifies the descriptive framework developed by the US Office of Technology Assessment (OTA). The framework considers five stages in the development and release of GMOs: formation, release, proliferation, establishment and effect. The first two stages correspond to risk-source characterisation, the last to the traditional dose-response stage of an ecotoxicological risk assessment.

#### *Hazard identification*

The reports do not identify or recommend any inductive hazard assessment techniques. The 1986 report, however, does identify fault trees and event trees as a means to quantify probability (see below). Part E of Appendix D provides a de facto checklist of environmental hazards.

#### *Likelihood and consequence assessment*

Fault trees, event trees and simulation can be used to quantify the probability and the magnitude of consequences in the first two stages of the assessment framework. The last stage can be analysed by adapting conventional epidemiological or toxicological methods, although ecological consequence assessment is less well developed than its human counterpart. The 1986 report notes that stages 3 and 4 of the framework are difficult to analyse using existing risk assessment methods. It therefore suggests that qualitative risk assessment can be used to compare the propensities for survival, establishment and genetic stability under different environmental conditions.

#### *Uncertainty and significance analysis*

The reports do not explicitly discuss uncertainty within the risk assessment process or the significance of the risk estimates.

#### *Monitor and review*

The 1992 report states that scientifically acceptable and environmentally sound field research requires: formulation of an hypothesis and statement of objectives; development of specific methodologies to introduce, monitor and mitigate the organisms; a precise description of the design of experiments, including planting density and treatment pattern; and a description of specific data to be collected, and of methods for analysis to test for statistical significance.

### *Appendix A5 Canadian Regulatory Directives 94-08 and 2000-07*

Environmental releases of GM plants in Canada are regulated by the Canadian Food Inspection Agency (CFIA) under powers granted by Part V of the Seed Act and the Canadian Environmental Protection Act. Regulatory directive 2000-07 contains guidelines for the environmental release of GM plants within confined field trials. Directive 94-08 describes the information used by the CFIA to identify potential adverse environmental impacts associated with the unconfined release of GM plants.

#### *Scientific principles and framework*

The CFIA conducts a case-by-case, environmental safety assessment for all plants with novel traits (PNTs) prior to authorising confined field trials and unconfined releases. PNTs are defined as plants derived from recombinant DNA technology or conventional breeding techniques that are neither familiar nor substantially equivalent to plants that are in use and generally considered as safe in Canada. The environmental safety assessment consists of five components. The first two require a description of the plants, its modification and novel traits. The third interaction assessment compares the biological characteristics of the modified plants with its unmodified "counterpart", including a post-harvest, residual effect analysis on any three of five indicator species (forage grass, legumes, annual cereal, corn or oilseed). The fourth stage requires an environmental impact assessment for natural and arable ecosystems addressing the degree of change, geographic scope, duration and relative impact on plants, animals, microbes, substance presence/persistence, sustainability, agronomic practice, resource conservation, other concerns and overall environmental quality. Finally, the guidelines require species replacement/competition and seed dormancy studies if there is reason to believe that the behaviour of the plant has been altered in unpredictable ways.

#### *Hazard identification*

The guidelines do not identify or recommend any formal hazard identification techniques. The environmental impact and interaction tables provide de facto checklists.

#### *Likelihood and consequence assessment*

Applicants considering commercialising GM plants are encouraged to include experiments, during the confined field trials, designed to meet the regulatory requirements of directive 94-08. Data provided by the applicant to support the interaction, environmental impact assessment and species replacement/competition and seed dormancy studies must be generated using statistically valid experimental designs and protocols. The guidelines identify three examples of replacement and seed dormancy analysis in the scientific literature: Crawley et al, 1993; Linder and Schmitt, 1994; and, Rissler and Melon, 1993.

#### *Uncertainty and significance analysis*

The guidelines do not refer to uncertainty within the environmental safety assessment. The significance of environmental impacts addressed in terms of duration, geographical scope and relative impact of the GM plant.

#### *Monitor and review*

Directive 2000-07 specifies species-specific terms and conditions for confined field trials, including the frequency of monitoring during the trial and the period of post-harvest restriction. The applicant must monitor unconfined releases sites if they become aware of any new information, relevant to the release, regarding risks to the environment or human health.

### *Appendix A6 Principles of Risk Assessment and Monitoring for the Release of GMOs*

The UK Department for Environment, Food and Rural Affairs (DEFRA, formerly the Department of Environment Transport and Regions) guidance on the principles of risk assessment and monitoring for the release of GMOs were developed to assist applicants during the revision of EC directive 90/220, which was subsequently repealed by directive 2001/18 (refer to Appendix A3). The guidelines were issued in 1999 without prejudice to the revisions of directive 90/220 that were taking place at that time. Their status relative to the new directive 2001/18, however, is unclear.

#### *Scientific principles and framework*

The guidelines state that, in accordance with the precautionary principle, the risk assessment should be transparent, scientifically sound, carried out on a case-by-case basis and re-examined if new information becomes available. The precautionary principle, however, is not defined in the guidelines. The risk assessment is to consider the direct, indirect, immediate and delayed effects of the GMO release—each of these terms are defined. The assessment should also be compared to the risks presented by the use of the unmodified organism in corresponding situations.

The guidelines specify a six-step risk assessment procedure. The first four steps consisting of hazard identification, consequence and likelihood assessment and risk calculation. The latter should be made, as far as possible, given the current 'state of the art'. The fifth step requires the analyst(s) to identify risks that need to be managed and how best to manage them. The analyst is then required to re-calculate the overall risk of releasing the GMO taking into account any proposed management strategies.

#### *Hazard identification*

The guidelines do not identify or recommend any inductive hazard assessment techniques. They do, however, provide a comprehensive checklist of potential GMO hazards. They also stress that potential adverse effects are not to be discounted on the basis that it is unlikely to occur.

#### *Likelihood and consequence assessment*

The guidelines do not identify or discuss any specific techniques to calculate the consequence or likelihood of adverse effects following the release of the GMO. They state that the characteristics of the environment into which the GMO is released, and the manner of the release, will be major factors in the consequences and likelihood of adverse effects.

#### *Uncertainty and significance analysis*

The guidelines do not refer to uncertainty in the risk assessment process, or the significance of the risk estimates.

#### *Monitor and review*

The applicant is required to submit a monitoring plan as part of the application to market a GMO, designed to confirm the assumptions made in the risk assessments and identify the occurrence of any unanticipated effects. Monitoring must be systematic and continue long enough to identify delayed and indirect effects, and give consideration to mechanisms for identifying and confirming any observed effects—implying a statistical approach to the probability of Type II error?

### *Appendix A7 OGTR Risk Assessment Framework, November 2001*

Section 50 of the Gene Technology Act 2000 requires the Office of the Gene Technology Regulator (OGTR) to prepare a risk assessment prior to issuing a licence for a dealing involving intentional release of a GMO into the environment. The OGTR risk assessment framework is a guide for applicants as to how the Regulator will undertake the risk assessment. The framework was finalised in November 2001 following public consultation. It will, however, be reviewed again after it has been in operation for approximately 12 months.

#### *Scientific principles and framework*

The risk assessment is to be scientific and transparent to applicants and the broader community alike. It will be conducted on a case-by-case basis and will consider short and long-term risks. The Regulator will use 'best practice' risk assessment methodologies when conducting the assessment. The risk will be compared to the risks posed by the unmodified organisms, and they will be re-examined in the light of new information. The precautionary principle is invoked in a medium strength form: lack of full scientific certainty should not be used as a reason for postponing 'cost-effective' measures to prevent environmental degradation. The Regulator will, however, assess the significance of incomplete or absent information, and if uncertainty about the environmental impact remains a licence will not be granted. Similarly, if a risk cannot be managed a licence will not be granted.

The risk assessment framework has three steps. The first hazard identification step identifies the type and sources of hazard, and the level certainty in the hazard identification process. The second risk assessment step addresses the probability of harm and consequences following exposure to the hazard, the level of certainty in the risk estimate and the significance of the risks. It also identifies management options and the acceptability of risk and management. The final step develops and implements the risk management plan including monitoring during and after the release where necessary.

#### *Hazard identification*

The framework provides a checklist of hazards that will be identified for specific taxonomic groups of organisms (e.g. plants, vertebrates, aquatic organisms, microorganisms living in animals, etc.). The document stresses that the checklist should not be seen as a comprehensive list of every possible risk. It encourages the applicant to adopt a comprehensive approach to identify the full range of hazards. It does not, however, identify or recommend any specific inductive or systematic hazard identification techniques.

#### *Likelihood and consequence assessment*

The framework does not identify or recommend any specific methods to calculate the likelihood and consequences of hazards. It does, however, refer to emerging international standards such as the OECD consensus documents. It also emphasises that some risks can be analysed quantitatively given sufficient information but warns wide confidence intervals will reduce the utility of the estimate. The regulator will a) undertake a quantitative analysis if the necessary information is available; and, b) specifically require the applicant to obtain quantitative data where this is necessary for best practice risk assessment, or where it is needed enable risk analysis in areas of scientific uncertainty. The Regulator will conduct a qualitative assessment, based on expert opinion, public consultation and analogous situations, if a quantitative approach is not possible.

#### *Uncertainty and significance analysis*

The framework requires the applicant to address the level of certainty in their likelihood and consequence assessments. If uncertainty is high the applicant may skip the actual risk calculation, assume a worst case, and

implement management strategies on this basis. The framework states that the significance of risk should be gauged relative to the number and severity of hazards, the magnitude, geographical extent, duration and frequency of impact, cumulative impacts and reversibility.

### *Monitor and review*

The OGTR will monitor and inspect releases but the framework does not state how this will be achieved or how long it will continue.

### *Appendix A8 ERMA Technical Guides ER-TG-01-1 9/99 and ER-TG-03-1 7/00*

Section 25 of the Hazardous Substances and New Organisms Act 1996 prohibits the import, development, field testing or release of new organisms, including *inter alia* GMOs, without prior approval of the Environmental Risk Management Authority (ERMA). These technical guides discuss techniques for identifying risks and preparing information on risks, costs and benefits for applications under section 25 of the Act. A separate technical guide on ecological risk assessment is currently being prepared by ERMA.

#### *Scientific principles and framework*

Section 7 of the HNSO Act requires ERMA to take into account the need for caution in managing adverse environmental effects when there is scientific and technical uncertainty about those effects. ERMA will use recognised risk identification, assessment, evaluation and management techniques when evaluating applications under Part V of the Act. Information provided by applicants must be "necessary and sufficient" for decision-making. The assessment should be conducted on a case-by-case basis and the results compared to the baseline—i.e. what would happen if the application were refused.

The guidelines adhere to the Australian and New Zealand risk management standard (AS/NZS 4360: 1999). The risk assessment consists of five steps: establish the context; hazard identification; calculate risk by combining estimates of likelihood and consequence; and treat risks. Monitoring and reviewing occurs at each step, together with consultation and communication with interested parties.

#### *Hazard identification*

The guidelines encourage applicants to demonstrate that they have conscientiously considered the widest possible range of obvious and non-obvious risks. The hazard identification must examine all possibilities of harm regardless of the likelihood of occurrence. The analysis must be thorough and systematic and may include stakeholders and interested parties. The guidelines provide a comprehensive list of hazard identification techniques, including: informal brainstorming, analogy to known cases and failure analysis, the Delphi technique, checklists, fault and event trees and HAZOP analysis.

#### *Likelihood and consequence assessment*

Neither of the guidelines discusses specific methods to estimate the likelihood or consequences of hazards. They do, however, contrast qualitative, semi-quantitative and quantitative approaches. Qualitative measures of likelihood, consequence and overall risk based on AS/NZS 4360: 1999 are provided as examples. The guidelines note that quantitative approaches may include various forms of statistical analysis, fault and event tree analysis, and extrapolation. The quality and validity of these approaches depends on the availability of data, and on the accuracy and completeness of the numerical values and methods (e.g. experiments, models) used to derive the data.

#### *Uncertainty and significance analysis*

The guidelines distinguish between variability and uncertainty. They suggest that probability or frequency distributions can be used to analyse variability. Two sources of uncertainty are identified: sampling error and lack of knowledge (about the consequences or likelihood of risk). They recommend checking information for bias, statistical competence and peer review, and obtaining further information where appropriate. The guidelines encourage applicants to consider the significance of adverse environmental effects, specifically their distribution over time and space, and whether they are acute, chronic or irreversible.

*Monitor and review*

Monitoring and reviewing are explicitly highlighted in the risk assessment framework but these components are not discussed further in the guidelines.

## *Appendix B Quantitative methods and models*

### *Toxicity*

The toxicity of GM plants and microbes can be measured with a variety of well-established ecotoxicological methods. The toxic effects of a substance are a function of the substance's concentration, the duration of exposure, the severity of the effect and the proportion of a population or community responding (Suter, 1993). Risk analysts usually collapse this problem along one or more dimensions describing, for example, sigmoid curves for the percentage of species responding against increasing concentration of the toxicant (Cardwell, 1989; Parkhurst *et al*, 1995).

State of the art soil toxicity tests have been developed for 42 soil invertebrates representing all the major functional groups—primary producers, herbivores, predators, parasites, decomposers and saprophytes (Jepson et al., 1994). Soil toxicology, however, is still many years behind its aquatic counterpart. The analyst may find that basic toxicological data is missing for some of these species. Van Straalen and Denneman (1989) therefore provide a toxicological model that does not require large data sets but can still account for the different sensitivity of species. Their approach assumes that the distribution of No Observed Effects Concentration (NOEC) values, corrected for different soil conditions, within a community of soil organisms can be described by a log-logistic distribution. For a species selected at random, the probability that its ln (NOEC) falls between $x_1$ and $x_2$ is given by

$$\int_{x_1}^{x_2} f(x) \ ,$$

where

$$f(x) = \frac{\exp\left(\frac{\mu - x}{\beta}\right)}{\beta\left[1 + \exp\left(\frac{\mu - x}{\beta}\right)\right]^2} \ . \qquad [3]$$

The hazardous concentration for *p*% of the species (HC*p*) is defined as a value of *x* such that the probability of selecting a species with a NOEC smaller than HC*p* is equal to $\delta_1$, where $\delta_1$ is an arbitrary small number, such as 0.05. Thus

$$\int_{-\infty}^{\ln HCp} f(x)dx = \delta_1 \ ,$$

where *f(x)* is given by equation [3]. After integration and manipulation this can be written as

$$HCp = \exp\left[\mu - \beta \ln\left\{\frac{1-\delta_1}{\delta_1}\right\}\right] \ . \qquad [4]$$

The parameters $\mu$ and $\beta$ are estimated from a measured series of NOEC values derived from chronic toxicity experiments, for example, using growth or reproductive endpoints. If *m* species are tested with mean ln (NOEC) denoted by $x_m$ and standard deviation $s_m$ then

$$\hat{\mu} = x_m = \quad \text{and} \quad \hat{\beta} = \frac{s_m \sqrt{3}}{\pi} \ .$$

One source of error in the model, which leads to an overestimate of *HCp*, is the limited number of test species *m*. The authors correct for this by introducing a correction factor $\delta_m$ that increases the standard deviation of the distribution. Koojiman (1987) tabulates the correction factor for *m* and the probability $\delta_2$ of overestimating *HCp*. $\delta_m$ decreases with increasing m; for $m \to \infty$, $\delta_m \to \pi/\sqrt{3}$. Equation [4] therefore becomes

$$\hat{HCp} = \exp\left[ x_m - \frac{3 s_m d_m}{\pi^2} \ln\left\{ \frac{1-\delta_1}{\delta_1} \right\} \right] \ . \qquad [5]$$

The model may be used inversely to estimate the risk associated with a concentration *c*. Equating the left-hand side of equation [5] with *c* and re-writing the equation yields

$$q = 100\left( 1 - \left[ 1 + \exp\left\{ \frac{\pi^2 (x_m - \ln c)}{3 s_m d_m} \right\} \right]^{-1} \right) \ , \qquad [6]$$

where $q = 100(1 - \delta_1)$ is the percentage of species unharmed when the environmental concentration equals *c*.

This model has a high level of practicality—it has been tested and validated numerous times, the parameters are easily estimated given relatively basic ecotoxicological data and they are not resource intensive. The model reliability is high—the toxicity endpoints (growth, survival or reproduction) are relevant to ecological risk assessment, it is biologically realistic, reasonably flexible and it allows for uncertainty in the mean and variance of the NOEC. The model has been reviewed by national agencies and is well accepted for conventional ecotoxicological risk assessment (Pastorok et al., 2002). This review did not identify any examples of its use within a GMO context. There does not, however, appear to be any reason why the model would not perform equally well.

| Hazardous process | Model type | Practicality | Reliability | Acceptance |
|---|---|---|---|---|
| Toxicity to non-target organisms | HCp toxicity extrapolation | High | High | High |

### *Fitness and competitive ability*

Fitness describes the difference in reproductive success of a genotype relative to another in a particular environment. Differences in fitness create selection for the fitter genotype. Selection is usually so slow that its effects on the dynamics of ecological systems are not noticeable over the time spans that concern ecologists. In microorganisms, however, it is much faster because they have very short generation times, grow exponentially and exist in extremely large populations. For the same reasons it is virtually impossible to measure growth rate in microorganisms, hence their competitive ability is usually measured in terms of fitness and selection (Lenski, 1991).

Genetically modified microorganisms (GMMOs) that are fitter than wild type or parental strains have the potential to displace these strains and competitively dominate the environment into which they are released. The joint effects of segregation[7] and selection in GMMOs are described by the following differential equation

[7] Segregation is the loss of an engineered gene due to the infidelity of replication or transmission, including mutation.

$$\frac{dp}{dt} = -sp(1-p) - \mu p \ , \qquad [7]$$

where $p$ is the frequency of an engineered genotype, $q = 1 - p$ is the frequency of the parental genotype, $\mu$ is the rate of segregation and $s$ is the selection coefficient (Lenski, 1991). Integrating equation [7] gives

$$p_t = \frac{p_0(\mu + s)}{[\mu + s(1-p_0)]\exp\{(\mu + s)t\} + sp_0} \ , \qquad [8]$$

where $p_t$ is the frequency of the engineered genotype after time $t$, and $p_0$ is the initial frequency. Direct measurements of the parameters $\mu$ and $s$ can be obtained from chemostat or microcosm experiments using non-linear regression.

This model is similar to a simple Malthusian population growth model with two parameters $s$ and $p$ as opposed to the usual Malthusian growth parameter $r$. Its practicality is high—ordinary differential equations are well developed and the two parameters are fairly easy to estimate. Its reliability, however, is low primarily because it ignores many biologically important processes (density dependence, demographic and environmental stochasticity, dispersal etc.) and it is deterministic. Its regulatory acceptance is assumed to be high by association with the Malthusian model, which is well accepted.

| Hazardous process | Model type | Practicality | Reliability | Acceptance |
|---|---|---|---|---|
| Altered fitness or competitive ability | Malthusian-like differential equation | High | Low | High |

If $s > 0$ then selection favours the parental genotype; if $s < 0$ selection favours the engineered strain. Selectively inferior GMMOs, however, are not inherently safe. Chance mutations may increase their fitness relative to wild type strains and their mutants—a process known as periodic selection. This may cause a temporary or permanent displacement of wild-type micro-flora. Kim et al. (1991) quantify the risks associated with periodic selection by deriving an extreme-value (EV) distribution for the "maximum fitness" of mutants derived from engineered strains during the "mutational exposure" (i.e. the time taken for the population of engineered strains to go extinct). A similar distribution is derived for wild type mutants during the same period. In each case, the underlying distribution of mutation fitness is assumed to be exponential with variance parameters γ' and γ for engineered and wild type strains respectively. The risk of periodic selection is then given by

$$\text{Risk} = \int_0^\infty \{1 - \exp[-\exp(-y')]\}\alpha_n \exp\{-y[-\exp(-y)]\}dx \ , \qquad [9]$$

where $y'$ and $y$ represent the parameters of the EV distribution of engineered and wild type strains. The wild type parameters are given by

$$y = \frac{1}{\gamma}(x - \mu_n) \ , \qquad [10]$$

$$\mu_n = \gamma \log N_s^{WT} + wf \ , \qquad [11]$$

$$N_s^{WT} = N_e^{WT} \cdot P_{adv}^{WT} \ , \qquad [12]$$

where $P_{adv}^{WT}$ is the mutational rate to advantageous mutants, $N_e^{WT}$ is the total number of wild type cells exposed to mutation during the "mutational exposure", $wf$ is the fitness of the wild type strain and $x$ is the fitness of the mutants. The equivalent GMMO parameters are given by

$$y' = \frac{1}{\gamma'}\left(x - \mu_n^{'}\right) , \qquad [13]$$

$$\mu_n^{'} = \gamma' \log N_s^{GMMO} + wf - \Delta f \ , \qquad [14]$$

$$N_s^{GMMO} = N_e^{GMMO}\left(P_{adv}^{WT} + \rho\Delta f\right) , \qquad [15]$$

where $\Delta f$ is the absolute value of the difference in fitness between the wild type and engineered strains and ρ represents the slope of the change in the rate of advantageous mutants which is assumed to decline linearly with increasing fitness. Again each of the unknowns in these equations can be measured or derived from chemostat or microcosm experiments.

This model uses well developed extreme value theory, however, its practicality is only medium because of the number of parameters that need to be estimated, and the relatively large amounts of data that are needed to achieve this. Model reliability is high because it is relevant to an important ecological risk, is biologically realistic and can be expressed in probabilistic terms. Regulatory acceptance is largely unknown but thought to be medium because extreme value theory is not widely practised in ecology.

| Hazardous process | Model type | Practicality | Reliability | Acceptance |
|---|---|---|---|---|
| Altered fitness or competitive ability | Extreme value function | Med | High | Med |

Selection and growth in higher organisms are much slower than in microorganisms. Fitness and competitive ability are therefore usually expressed in terms of fecundity and/or growth rate. Parker and Kareiva (1996), for example, assess invasion risk by comparing the finite rate of increase of GM plants with the parent lines from which they were derived. The finite rate of increase *r* is expressed as

$$r = P(gs) \cdot F_{adult} \cdot S \ , \qquad [16]$$

where *P(gs)* is the probability of germination and survivorship to adult, $F_{adult}$ is the total number of flowers per adult plant, and *S* is the number of seeds per fruit for the particular plant in question. Again each of these parameters can be determined by greenhouse experiments. A similar approach was adopted by Crawley et al. (1993) to assess the invasion risk of GM canola sown in spring. They expressed the finite rate of increase as

$$r = (1 - d_1 - g) + g(1 - d_2)\overline{F} \ , \qquad [17]$$

where $d_1$ is the proportion of seeds that die in one full year, *g* is the proportion of seeds germinating in the first spring, $d_2$ is the proportion of seeds that die over winter, and $\overline{F}$ is the mean number of seeds produced per seed that germinates. In this example the parameters of the model were estimated from field trials held in 12 different habitats at three sites in the United Kingdom. Values of $r > 1$ imply that the GM plant will increase in abundance under the given set of environmental conditions. The dormancy and viability of seeds in the seed bank, *d*, can be easily measured with *in situ* experiments (Linder and Schmitt, 1994).

These models are very practical—they are simple to construct and easy to parameterise. The models are relevant but deterministic and therefore biologically not very realistic—the parameter estimates are likely to be very site specific and may not be predictive of conditions in other sites. Regulatory acceptance, however, is assumed to be high because the models are simple and it is relatively easy to collect the necessary site-specific data.

| Hazardous process | Model type | Practicality | Reliability | Acceptance |
|---|---|---|---|---|
| Altered fitness or competitive ability | Deterministic geometric function | High | Med | High |

*Pollen dispersal*

Pollen dispersal is an important process in all biotechnology hazards associated with gene flow, and is critical to risk management strategies designed to contain GM plants particularly during field trials. Classic spore and pollen dispersal models describe the amount of pollen *p* deposited at a distance *x* from a point source as

$$p = \frac{a\exp(-bx)}{x}\ , \qquad [21]$$

where *a* represents the total amount of pollen shed into the pollen cloud and *b* represents the proportion of pollen deposited per unit distance (Giddings *et al*, 1997a). This model can be re-specified as a reliability function *R(x)* which gives the probability that a pollen grain travels at least *x* distance units away from its original parent plant (Kareiva et al., 1994). The simplest reliability function is given by

$$R(x) = \exp(-bx)\ , \qquad [22]$$

implying that pollen is carried away from its source and deposited at a constant rate. In this model the mean distance travelled is *1/b* and the variance $1/b^2$. Equation [21] may be modified for the effects of turbulence, which decreases the rate of deposition whilst increasing the dilution of the pollen cloud

$$p = \frac{a\exp[-bx(1-m)]}{x^{(1+m)}}\ , \qquad [23]$$

where *m* is the turbulence parameter. More complicated models such as this can be respecified in terms of a Weibull reliability function

$$R(x) = \exp(-bx^c)\ . \qquad [24]$$

In recent experimental trials, however, equations [21] and [23] did not provide accurate representations of actual pollen deposition, largely because they do not allow for the wind direction or strength. These models were improved by assuming that pollen deposition varies with compass direction , in accordance with a normal distribution about a mean wind direction $\mu_w$ and standard deviation $s_w$ (Giddings et al., 1997b).

Plants *x* m away from a transgenic pollen source will receive pollen from a variety of other sources. Thus *R(x)* does not represent the expected fraction of seed from a plant *x* m that will test positive for the transgene. This fraction is the product of the reliability function and the probability that a pollen grain transported *x* m will fertilise an ovule in the plant on which it lands. It is assumed that this probability is linearly proportional to the reliability function, with the constant of proportionality *c* estimated from out-crossing data.

This model is implemented as follows: $S_k$ seeds are collected for sample station *k* (a known distance from a single source of transgenic pollen) with a total of *N* different sample stations. The number of seeds that test positive for the transgene $M_k$, is a binomial random variable where the probability of "success" is given by *cR(x)*. The site-specific parameters of the model are those that maximise the log-likelihood function

$$\log L = \sum_{k=1}^{N} M_k \log[cR(x)] + (S_k - M_k)\log[1 - cR(x)] \ . \qquad [25]$$

It is important to note that if gene silencing (a phenomenon witnessed in many plant systems) occurs then the transgene test must screen for the gene presence not just the gene product, otherwise this model will underestimate the rate of gene flow (Hokanson et al., 1997).

These practicality and reliability of these models is medium. They are relatively well developed and resource efficient but only become flexible and biologically realistic when the model incorporates additional wind direction and strength parameters. Increasingly large amounts of data are required to fit these additional parameters, and whilst the data collection is relatively straightforward, the model implementation is quite labour intensive. Furthermore the model results are known to be very sensitive to the experimental design (Lavigne et al., 1998). Regulatory acceptance is therefore assumed to be medium.

| Hazardous process | Model type | Practicality | Reliability | Acceptance |
|---|---|---|---|---|
| Pollen dispersal | Reliability functions | Med | Med | Med |

### *Horizontal gene flow*

The area a GMO occupies and its rate of spread underlie many of the potential risks associated with biotechnology. Transgenes may spread with the movement and dispersal of the recipient organism, or independently of their host, a process known as “horizontal gene flow”.

Experiments with different plasmids in *E. coli* K12 have shown that horizontal gene flow via plasmid transfer can be described by a simple mass action model

$$\frac{dn^*}{dt} = k(n)(n_+) \ , \qquad [26]$$

where $n_+$ and $n$ are concentrations of potential donor and recipient bacteria (with and without the plasmid respectively) and $n^*$ are “new” plasmid-containing bacteria. The value of the rate constant $k$ ranged from $10^{-13}$ to $10^{-9}$ ml/cell-hour depending on experimental conditions, for bacteria growing exponentially, dropping by two orders of magnitude at equilibrium (Strauss et al., 1985). This model is essentially a Malthusian differential equation. As noted above this type of model is practical and well accepted but not very reliable because it ignores important biological processes.

| Hazardous process | Model type | Practicality | Reliability | Acceptance |
|---|---|---|---|---|
| Horizontal gene flow | Malthusian like differential equation | High | Low | High |

The model described above is a single population model that does not take into account the transfer of plasmids to other bacterial populations. Landis et al., (2000) describe a more complicated multi-population approach that models the flow of plasmid-borne genes via conjugation[8] between 2, 3 or 4 bacterial populations subject to segregation, selection and density dependant growth in a resource limited environment. Here the frequency of conjugation is treated as a Poisson random variable. The existence of several bacterial populations that can serve as suitable hosts for plasmid reproduction is analogous to a meta-population model. The meta-populations are

[8] There are three ways in which genetic material is exchanged between bacteria: transformation, transduction and conjugation. Conjugation involves the simple union of two individuals or filaments.

the populations of the GM plasmid, and the bacterial hosts represent the various environmental patches. Migration rates between hosts are determined the compatibility of the host populations to exchange genetic information and the Euclidean distance between them. Another additional element is that the "environmental patches" in this model—i.e. the host populations—grow and shrink depending on their growth rate, the carrying capacity of the environment and genetic fitness of the host, as altered by the GM plasmid.

Meta-population models of this type are generally quite practical and highly reliable. Indeed Pastorok et al., (2002) rank these types of models as some of the most relevant and tractable models for ecological risk assessment. These types of models are also well accepted within terrestrial systems within a conservation biology context. They need to be designed around specific habitats and populations and are not therefore very flexible. Acceptance within a GMO context may therefore be reduced and is assumed to be medium.

| Hazardous process | Model type | Practicality | Reliability | Acceptance |
|---|---|---|---|---|
| Horizontal gene flow | Meta-population model | Med | High | Med |

### *Spread*

The spread of an annual plant whose per capita seed production is density dependent can be described as

$$N_{t+1}(x) = r\int \upsilon(x-\xi)N_t(\xi)\exp[-sN_t(\xi)]d\xi \ , \qquad [27]$$

where $N_t(x)$ is the population density at position $x$ in a one dimensional environment in year $t$, $r$ is the maximum reproduction rate (reached as the population density tends to zero), and $s$ describes density-dependant reduction in per capita reproduction (Manasse and Kareiva 1991). The function $\upsilon(x-\xi)$ is the probability that a seed from a plant in position "$x$-$\xi$" will disperse to position $x$. This function can be obtained from field experiments such as those described by Kareiva et al., (1994) above. The other important parameter $r$ is also easily derived using equations such as [16].

This type of model is a relatively simple description of spread along an axis. Models of this type have a high level of practicality—they are well developed and resource efficient but may be difficult to parameterise. The model includes important biological processes such as density dependence and probabilistic dispersion, although the latter is inevitably much simplified in the model—particularly because the environment through which the organism is spreading is unlikely to be homogenous.

There are at least 13 other media-specific models for air, soil, groundwater, etc., which describe how GM microbes disperse in the environment. Some of these models are used and endorsed by regulatory authorities such as the USEPA for chemicals, faecal coliforms, and in one example, for *Pseudomonas fluorescens* genetically modified to metabolise lactose (Strauss and Levin, 1991). Regulatory acceptance is therefore assumed to be high.

| Hazardous process | Model type | Practicality | Reliability | Acceptance |
|---|---|---|---|---|
| Spread | Density-dependant | High | Med | High |

The growth and spread of a GMO that competes with a closely related species, for example a weedy relative of the parent crop, can be analysed by amending the Lotka-Volterra equations to include a dispersal term

$$\frac{\partial N_1(x,t)}{\partial t} = r_1 N_1\left[1 - \frac{N_1}{K_1} - \frac{a_{12}N_2}{K_1}\right] + D_1\frac{\partial^2 N_1}{\partial x^2} \qquad [28a]$$

$$\frac{\partial N_2(x,t)}{\partial t} = r_2 N_2\left[1 - \frac{N_2}{K_2} - \frac{a_{21}N_1}{K_2}\right] + D_2\frac{\partial^2 N_2}{\partial x^2} \quad . \qquad [28b]$$

Here $N_i(x, t)$ is the density of species *i* at position *x* and time *t*. $D_i$, $K_i$ and $r_i$ represent the diffusion coefficient, carrying capacity and intrinsic rate of increase respectively for species *i*, $a_{ij}$ is the competitive effect of species *i* on species *j* (Manasse and Kareiva, 1991).

Lotka-Volterra type models are well developed. Estimating the parameters of the model for two or more species, however, can be quite difficult. The inclusion of the competitive effect coefficient in this model improves the biological realism. The model is highly relevant to GMO risk assessment but does not explicitly incorporate uncertainty—although this could presumably be include within a Monte Carlo shell at the expense of resource efficiency. Overall reliability is therefore rated medium. Regulatory acceptance is rated as medium because Lotka-Volterra models do not appear to be used by any regulatory agencies, despite their high credibility in academic circles (Pastorok et al., 2002).

| Hazardous process | Model type | Practicality | Reliability | Acceptance |
|---|---|---|---|---|
| Spread | Lotka-Volterra with competition term | Med | Med | Med |

### *Establishment*

Establishment occurs when a founding population increases in size or persists during a specific period of time. It is usually modelled as a function of r, K and $N_o$—the initial size of the population. The Velhulst-Pearl logistic equation describes the population at time t as

$$N_t = \frac{K}{1 + [(K - N_o)/N_o]\exp(-rt)} \quad . \qquad [29]$$

As long as r is non-negative and K is positive then the model precludes extinction of the population, i.e. establishment. Alternative approaches are based on the reproductive ratio $R_o$, given by

$$R_0 = \int l(x)\cdot m(x)dx \quad , \qquad [30]$$

where l(x) is the number of survivors to age x and m(x) is reproductive rate (of females) of age x (Williamson, 1989). In a simple stochastic birth and death process with constant per capita birth and death rates, the population survives with probability

$$1 - \left(\frac{1}{R_0}\right)^N \quad . \qquad [31]$$

Both of the models described above are simple, well accepted and have few parameters that are relatively easy to estimate. They therefore score high on practicality. They score low on reliability, however, because they are not biologically realistic or flexible and do not explicitly allow for uncertainty. Despite these shortcomings these models are prominently described in ecological texts and are thought to be well-accepted by regulatory authorities.

| Hazardous process | Model type | Practicality | Reliability | Acceptance |
|---|---|---|---|---|
| Establishment | Velhurst-Pearl, Reproductive ratio | High | Low | High |

Tomiuk and Loeschcke (1993) describe a more complicated individual-based model that allows per capita birth and death rates *b(i)* and *d(i)* to vary as

$$b(i) = [i \cdot V_i + r_i]/2 \qquad [32a]$$

$$d(i) = [i \cdot V_i - r_i]/2 \qquad [32b]$$

where $r_i$ is the individual growth rate with variance $V_i$ when the population consists of $i$ individuals. The persistence (extinction) time $T$ of a population with size $N$ is then given by

$$T = \sum_{i=1}^{N} \left\{ \frac{2}{i(i \cdot V_i - r_i)} + \left[ \sum_{m=i+1}^{K} \frac{2}{m(V_m - r_m)} \prod_{k=1}^{m-1} \frac{(k \cdot V_k + r_k)}{(k \cdot V_k - r_k)} \right] \right\} , \qquad [33]$$

where $K$ is the upper limit of resource availability beyond which the population size decreases.

The critical parameters in these models—$r$, $K$, $R_0$ and $V$—can all be measured in laboratory and field trials. Again, however, all of these parameters are species- and environment- specific. They may be determined by empirical observation but this undermines the predictive utility of the risk assessment. Techniques do exist, however, to vary these values by assuming they lie within a probability distribution generated empirically or from expert probability encoding methods (see for example Burgman et al., 1993).

Individual models such as these are significantly different from the other population state models (individual members of a population are regarded as identical) described so far. Individual models tend to be highly reliable because they are biologically realistic, can easily incorporate uncertainty and are very relevant. They are not, however, very flexible because they are usually designed in great detail around specific populations in specific habitats. These models score low against practicality because they take long time to set up and run, are data intensive and have a low to medium rating for parameter estimation (Pastorok et al., 2002). For these reasons they are also assumed to score low for regulatory acceptance in a GMO context.

| Hazardous process | Model type | Practicality | Reliability | Acceptance |
|---|---|---|---|---|
| Establishment | Individual based | Low | High | Low |

### *Food web analysis*

Food web and trophic flow analysis are important components of modern community theory. They may also provide mathematical insight into some of the potential effects of GM plants and microorganisms that persist or spread into new habitats. Early food-web studies were based on the Lotka-Volterra equations described above. Pimm (1982) has taken this relationship and modelled it within a food web context, for a system of $n$ species, generating an equation of the form

$$X_i^* = X_i \left( d_i + \sum_{j=1}^{n} a_{ij} X_j \right) , \qquad [34]$$

where $X_j$ and $X_i$ represent the density of two different species, and $X_i^*$ is the rate of change of species *i* in the absence of immigration and emigration. The sign and magnitude of the $a_{ij}$ term will depend on whether $X_i$ is the predator and $X_j$ the prey or vice versa. If the two species do not interact then $a_{ij} = a_{ji} = 0$. Furthermore, a number of different relations within a food web can be described by changing the combination of signs for $a_{ij}$ and $a_{ji}$, for example predator—prey (+, -), competitive, (-, -) and mutualistic, (+, +). These equations can be solved deterministically by measuring "typical" biological relationships and applying the appropriate signs and values, or stochastically by choosing the sign and magnitude of the parameters from probability distributions.

The basic equations that underline this approach are relatively well developed (see Lotka-Volterra models above) but their application in this context is not. Furthermore, these types of multi-species models are inevitably data intensive and difficult to parameterise—for these reasons they score low against practicality. Their reliability is medium because whilst they are highly relevant to important GMO hazards, they may not be biologically realistic and do not handle uncertainty well. This type of approach is also relatively new and unlikely to be readily accepted by regulatory authorities.

| Hazardous process | Model type | Practicality | Reliability | Acceptance |
|---|---|---|---|---|
| Cascading ecosystem effects | Food web | Low | Med | Low |

### *Trophic flow analysis*

There are at least three other approaches that can be used to examine ecosystem effects following the introduction of a new or altered species. All use digraphs (directed graphs) that describe an ecosystem in terms of its nodes (populations, species or functional groups) and the type and strength of interaction between these nodes. The simplest qualitative approach only captures the type of interaction between nodes—i.e. positive, negative or zero. This is converted into community matrix that allows quantitative estimates of the stability of the system (Li et al., 1999). The second approach, known as trophic flow network analysis, measures the flow of a commensurable commodity (for example carbon) in defining the activity (total throughput) and organisational status of the system. If $T_{ij}$ is the transfer from nodes *i* to *j* (*i*, *j* = 1, 2, 3...*n*) of some commodity, then the total system throughput is given by

$$T = \sum_{i=0}^{n} \sum_{j=1}^{n+1} T_{ij} \quad , \qquad [35]$$

assuming exogenous inputs are derived from a hypothetical zero node, and exogenous outputs flow to a hypothetical *n + 1* node. The system ascendancy *A* is defined as the product of *T* and the "average mutual information" *I* given by

$$I = \sum_{i=0}^{n} \sum_{j=1}^{n+1} \left( \frac{T_{ij}}{T} \right) \log \left[ \frac{T_{ij} \cdot T}{\left( \sum_{k=1}^{n+1} T_{ik} \right) \left( \sum_{m=0}^{n} T_{mj} \right)} \right] \qquad [36]$$

where *m* and *k* represent all nodes in the system other than *i* and *j*. The system's ascendancy, or rather the ratio of *I* to *T*, gives some insight into the health of the system. For example, high I relative to *T* usually indicates narrow trophic specialisation (low redundancy) suggesting that the system may be vulnerable to changes in its nutrient or energy flows (Ulanowicz, 1992).

The last and most sophisticated of these approaches is known as pathway analysis. It assesses the covariation among species within a "community interaction web" quantifying which of the possible hypotheses describing the relationships between the web's nodes is most likely to be correct, and the relative strengths of direct and indirect interactions between these nodes. The analysis applies multiple regression between nodes of experimentally manipulated communities in a manner determined by the causal hypotheses that are thought to be operating within the web. The regression coefficients indicate how much the dependent node changes with any given change in the different causal nodes. When standardised by the ratio of the standard deviations of the independent and dependant nodes, the regression coefficients also indicate the strength of the association between the nodes and sign of the interaction. The analyst can identify the most likely causal hypothesis by comparing the variance-covariance (correlation) matrix predicted by each hypothesis with the observed correlation matrix given by the experimental data—see Wootton, (1994) for an applied example of pathway analysis.

Again these types of models are not highly practical—they are not well developed, are resource intensive (due in part to the high experimental effort needed to construct and analyse the underlying digraphs) and are not easy to parameterise. They are highly relevant but are very sensitive to the choice of nodes—the model predictions will deviate significantly from reality if important species or processes are not included within the digraph. They are not therefore very flexible. To date none of these approaches appear to have been applied to GMO risk assessment and their regulatory acceptance is thought to be low.

| **Hazardous process** | **Model type** | **Practicality** | **Reliability** | **Acceptance** |
|---|---|---|---|---|
| Cascading ecosystem effects | Trophic flow/pathway analysis | Low | Med | Low |

*Appendix C Distribution functions for Monte Carlo Analysis*

### *Sample distribution functions*

For any discrete or continuous random sample $x_1$, $x_2$, ... $x_n$, with unknown distribution function F(x), the sample distribution function $\hat{F}_n(x)$ is the relative number of $x_i$ that are smaller than or equal to x. Thus

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(x_i \le x) \ , \qquad [37]$$

where the indicator function $I(x_i \le x) = 1$ if xi ≤ x and 0 otherwise. The analyst must subjectively determine the maximum and minimum values of the distribution function. For continuous variables these values are usually outside the observed range of the data (for example plus or minus two standard deviations). Notice that the sample distribution function $\hat{F}_n(x)$ is the frequency in *n* trials of the event that the variable X is less than or equal to x. As n increases $\hat{F}_n(x)$ is expected to approach the probability of the event {X ≤ x}, namely F(x) = Pr(X ≤ x). Thus, in this sense, the sample distribution function is an estimate of the underlying distribution function F(x). The analyst can draw probability estimates directly from the sample distribution function by summing the number of observations ≤ x, and dividing by n. The second order analysis proceeds as follows: the probability of the first order statistic (P1) is given by:

$$P_1 = Beta(1,n) \ , \qquad [38]$$

whilst the remaining order statistics (Pi; i = 2,3….n) are distributed

$$P_{i+1} = 1 - [U(0,1)]^{\frac{1}{n-i}} \cdot (1 - P_i) \ . \qquad [39]$$

Equations [38] and [39] allow the analyst to simulate a series of distribution functions for the same data set and thereby define confidence intervals for the probability of an event {X ≤ x), (Vose, 2000).

### *Kernel density functions*

An alternative non-parametric way to represent data is through a kernel density estimate, defined as

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \ , \qquad [40]$$

where the kernel K(t) is some function which satisfies the condition

$$\int_{-\infty}^{\infty} K(t)dt = 1 \ , \qquad [41]$$

and h is the bandwidth (Silverman, 1986). Very small changes in the bandwidth can cause dramatic changes in the density estimate. Furthermore the optimal bandwidth for a kernel density cannot be calculated precisely without *a priori* knowledge of the distribution function f(x). Silverman (1978) recommends plotting a test function based on the second derivative of the kernel function

$$\hat{f}_n''(x) = \frac{1}{nh^3}\sum_{i=1}^{n} K''(t) \ , \qquad [42]$$

for various values of the bandwidth h. The optimum bandwidth leads to a test graph that has fluctuations that are quite marked but do not obscure any systematic variation. This approach, however, is quite subjective. Silverman (1986) recognises this problem and suggests an "automatic" bandwidth

$$\mathrm{h} = 0.9\mathrm{An}^{-\frac{1}{5}} \ , \qquad [43]$$

where

$$A = min(\sigma, \text{interquartile range} / 1.34) \ , \qquad [44]$$

that provides a good fit for most unimodal and moderately bimodal probability densities. This calculation is trivial and easy to automate.

### *Extreme value distributions*

There are three families of extreme value (EV) distributions, known as Type I, II and III, but Type I is the most common. The family of type I EV distributions, with location parameter a and shape parameter b, are given by

$$G(x) = \exp\left\{-\exp\left[-\left(\frac{x-a}{b}\right)\right]\right\} \ , \qquad [45]$$

with density function

$$g(x) = \frac{1}{b}\exp\left[-\left(\frac{x-a}{b}\right)\right].\exp\left\{-\exp\left[-\left(\frac{x-a}{b}\right)\right]\right\} \ . \qquad [46]$$

All three families, however, can be re-stated in a single generalised EV distribution with three parameters, α, β and ε

$$G(x) = \exp-\left[\frac{\alpha-\beta x}{\alpha-\beta\varepsilon}\right]^{\frac{1}{\beta}} \ . \qquad [47]$$

When applied to temperature data, for example, the ratio α/β is the maximum (or minimum) temperature achievable, ε is the most frequently occurring extreme (mode), and α represents approximately the rate of increase (decrease) of the temperature extremes with the natural logarithm of time (Jacocks and Kneile, 1974).

The term 'extreme value' is attached to these distributions because they can be obtained as the limiting distributions, as n approaches ∞, of the greatest value among n independent random variables, each having the same continuous distribution (Johnson et al., 1995). Although these distributions are called extreme value, it is important to note that they need not necessarily represent distributions of all kinds of extreme values, for example, the extremes from small samples. They can also be used without recourse to an extreme model in the same way as any other probability distribution.