BIODIVERSITY SECTOR

ECOLOGICAL IMPLICATIONS OF GMOS

Robust methodologies for ecological risk assessment

Summary report: Best practice 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

1 Introduction

This report provides a brief summary of best practice ecological risk assessment for the unconfined release of genetically modified plants and microorganisms. It is a synopsis of a much larger review of best practice and current practice in ecological risk assessment for genetically modified organisms (GMOs).

2 Summary of best practice

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

Endpoints are an expression of the values that the analyst is trying to protect by undertaking a 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.

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.

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).

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

Ecological hazards may manifest in natural, arable and marginal environments and cut across all levels of biological organisation. There are a number of techniques available to identify hazards in complex ecological systems. 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.

Inductive techniques such as logic trees, hazard and operability analysis, failure modes and effects analysis and hierarchical holographic modelling 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. 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 above—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.

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

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. 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 risk assessments may not err on the side of conservatism, even when they purport to do so.

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

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

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.

In most cases simpler models are the most 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 meaningful descriptions of parameter uncertainty. 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 also exist in four other areas: altered farming practice, physical habitat modification and creation of new crop pests and viruses.

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

Uncertainty occurs throughout the process of constructing and releasing a GMO. Ecological risk assessment for unconfined releases is 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. Uncertainty can be divided into two types: linguistic and epistemic. Linguistic uncertainty occurs as vagueness, ambiguity and under-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.

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 below). 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. 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. 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.

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

There are a large number of techniques (uncertainty caclulii) 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. 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 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.

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

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. 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; 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 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 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.

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, 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.

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

Some of the potential hazards associated with GM crops are highly uncertain and potentially highly damaging. 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. These concepts *inter alia* 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. 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.

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

Risk assessment is an iterative process. It is important that monitoring strategies are implemented to test the predictions of prior risk assessments and provide information that will inform future risk assessments, thereby “closing the regulatory loop”. 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 use 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. Replications should be sufficient to detect changes of a pre-specified magnitude otherwise the strategy may run the risk of being underpowered, and of failing to detect important consequences.

Many experimental field studies are designed around null hypothesis tests. 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.

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. 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. 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.

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.