SKI Report 02:21
SSI report 2002:12

Research

Formulation and Presentation of Risk Assessments to Address Risk Targets for Radioactive Waste Disposal

R. D. Wilmot

October 2002

ISSN 1104-1374
ISSN 0282-4434
ISRN SKI-R-02/21-SE

SKI/SSI perspective

Background

SSI has recently issued regulations that impose a risk criterion for radioactive waste disposal. SKI has issued corresponding regulations on long-term safety of geological disposal, including aspects and guidance on safety assessment methodology (such as time frames). Based on these regulations, SSI and SKI need to develop an attuned view of what is expected from the applicant, in terms of risk assessment in support of a license application.

Relevance for SKI & SSI

This report represents the first step in the regulators' work toward such a development, by providing qualitative descriptions of various approaches to risk assessment by reference to assessments in other countries. Moreover, the report identifies a number of issues within the area of risk assessment methodology and presentation that may require some future activities (e.g., model and code development) by SSI and SKI.

Results

The objectives of this project have been fulfilled. Specifically, the objectives were to evaluate the approach to risk assessment for radioactive waste disposal, and how to define and present the results of such risk assessments in safety cases to meet risk targets. Moreover, by reference to risk assessments in other countries, the strengths and weaknesses of the different approaches to risk assessment have been illustrated.

Future work

Future activities will be to take this work forward and provide illustrative examples of simplified risk calculations to define a methodology that will be used for illustrating the concepts, approaches and issues of probabilistic calculations and risk assessments.

Project information

SKI project manager: Bo Strömberg
Project Identification Number: 14.9-010580/01114
SSI project manager: Björn Dverstorp
Project Identification Number: 60/2760/01 och SSI P 1288.01

SKI Report 02:21
SSI report 2002:12

Research

Formulation and Presentation of Risk Assessments to Address Risk Targets for Radioactive Waste Disposal

R. D. Wilmot

Galson Sciences Ltd,
5 Grosvenor House,
Melton Road,
Oakham,
Ruthland LE15 6AX
United Kingdom

October 2002

SKI Project Number 01114
SSI Project Number 1288.01

Executive Summary

The Swedish regulators have been active in the field of performance assessment of radioactive waste disposal facilities for many years and have developed sophisticated approaches to the development of scenarios and other aspects of assessments. These assessments have generally used dose as the assessment end-point. Regulations recently established in Sweden [SSI FS 1998:1] have introduced a risk criterion for radioactive waste disposal: the annual risk of harmful effects after closure of a disposal facility should not exceed 10^{-6} for a representative individual in the group exposed to the greatest risk.

This report evaluates different approaches to the definition and use of probabilities in the context of risk assessments, and examines the presentation of the results of risk assessments in safety cases to meet risk targets. The report illustrates the strengths and weaknesses of different possible approaches to risk assessment by reference to assessments in other countries, and provides suggestions for future activity and development in this area by the Swedish regulators.

The review of experience in other countries has led to a number of key observations relevant to the conduct of regulatory work on risk assessments and preparations for review. These highlight the importance of developing a protocol for conducting calculations, and linking such a protocol to the requirements of risk assessment calculations and to existing code and model capabilities.

There are a number of decisions and assumptions required in developing a risk assessment methodology that could potentially affect the calculated results. These assumptions are independent of the analysis of performance, and relate to issues such as the expectation value of risk, risk dilution, the definition of probability density functions and achieving convergence. A review of a proponent's risk assessment should address these issues in determining the appropriateness and validity of the results presented. Supporting calculations to explore these issues quantitatively could provide additional support for conducting such a review. Regulatory guidance on these issues would be a further means of supporting the review process.

In addition to a review of approaches to the calculation of risk, the report also examines alternative measures that have been proposed for assessing long-term performance of a disposal system. Such alternative performance measures include environmental concentrations, radionuclide fluxes and radiotoxicity. Such measures have been adopted in some regulatory regimes, but their use is not sufficiently widespread to draw definitive conclusions as to their usefulness. Alternative performance measures may be of value in developing an understanding of system performance, but stakeholders may find their use as regulatory criteria less easy to understand than measures of dose or risk. Additional work on developing a methodology for formulating and quantifying alternative performance measures is therefore suggested, together with consultation on the benefits and disadvantages associated with the adoption of such measures.

Sammanfattning

De svenska myndigheterna SKI och SSI har aktivt arbetat med säkerhetsanalyser för slutförvar av radioaktivt avfall och har utvecklat sofistikerade metoder för att beskriva olika scenarior och andra aspekter av analysen. Hittills har man i dessa analyser ofta kvantifierat säkerheten i termer av dos för jämförelse med ett doskriterium. SSI:s föreskrifter om slutligt omhändertagande av använt kärnbränsle och kärnavfall (SSI FS 1998:1) har dock introducerat ett riskkriterium. Dessa föreskrifter säger att den årliga risken för skadliga effekter efter det att förvaret förslutits inte får överstiga 10^{-6} för en representativ individ i den grupp som är utsatt för den största risken. Detta innebär att myndigheterna kräver att både konsekvenser (doser) och sannolikheten för att exponeras för en dos måste ingå i en säkerhetsanalys.

Den här studien utvärderar olika sätt att definiera och använda sannolikheter i riskanalyser. Vidare diskuteras hur resultaten från riskanalyser presenteras och används för att visa att uppsatta riskkriterier inte överskrids. Rapporten belyser också styrkor och svagheter hos olika metoder för att karakterisera risk utifrån en genomgång av riskanalyser som genomförts i andra länder.

Den internationella sammanställningen har använts för att identifiera ett antal områden där ytterligare studier kan vara motiverade, som stöd för framtida myndighetsgranskningar. Förslagen inkluderar framtagandet av ett ramverk för de beräkningar som görs till stöd för myndigheternas granskningsverksamhet. Ett första steg skulle kunna vara att definiera vilka krav som behöver ställas på de modeller och beräkningsverktyg som används för riskberäkningar, samt att göra en bedömning av dessa krav med avseende på befintligt kapacitet hos myndigheterna.

Ett annat viktigt område där ytterligare arbete föreslås, är att ta fram ett vägledande dokument som beskriver myndigheternas förväntningar på SKB:s kommande säkerhetsanalyser. Ämnesområden inom vilka en sådan vägledning skulle kunna tas fram är användandet av iterativa analyser, probabilistisk teknik i riskberäkningar och konditionerade riskberäkningar, samt metoder för att demonstrera konvergens.

Förutom översikten av riskberäkningsmetoder, undersöker studien olika alternativa mått som föreslagits för att utvärdera den långsiktiga säkerheten av slutförvar (t.ex. koncentrationer i miljön, radionuklidflöden och radiotoxicitet). Sådana alternativa säkerhetsindikatorer har använts i vissa länder, men inte i tillräcklig omfattning för att det ska vara möjligt att dra några slutsatser beträffande deras användbarhet. Alternativa säkerhetsindikatorer på en anläggnings funktion kan vara värdefulla då man ska utveckla en förståelse för hela systemets funktion, men att använda dem som säkerhetskriterium kan av olika berörda grupper anses vara svårare än dos- eller riskmått. Därför föreslås att metoder för att formulera och kvantifiera alternativa säkerhetsindikatorer tas fram, samt att för- och nackdelar med användandet av dessa utreds.

Contents

Formulation and Presentation of Risk Assessments to Address Risk Targets for Radioactive Waste Disposal

1 Introduction

The responsibility for regulation of radioactive waste management and disposal in Sweden is shared between the Swedish Nuclear Power Inspectorate (SKI) and the Swedish Institute for Radiation Protection (SSI). Recently introduced Swedish regulations [SSI FS 1998:1] impose a risk criterion for radioactive waste disposal: the annual risk of harmful effects after closure of a disposal facility should not exceed 10^{-6} for a representative individual in the group exposed to the greatest risk. The regulation and the accompanying guidance indicate that the regulatory authorities require a consideration of both consequences (doses) and the probability of receiving a dose to be considered in assessments.

During the preparation of this report, SKI has also published regulations concerning the disposal of nuclear material and nuclear waste [SKI FS 2002:1]. These are accompanied by guidance that describes recommended approaches to safety assessment. The purpose of the safety assessment is to show, *inter alia*, that the risks associated with selected scenarios are acceptable in terms of the SSI regulation. The recommendations therefore include a discussion of the selection of scenarios, classification of uncertainties, and the assignment of probabilities.

The Swedish proponent for radioactive waste disposal, SKB, issued a safety case for spent nuclear fuel disposal, SR 97, that attempted to address this risk target (SKB 1999). However, SR 97 was completed only shortly after the issuing of the regulations, and did not contain a fully-developed methodology for calculating risk. The approach of SKB to assessing risks in SR 97 was evaluated by Galson Sciences Limited (GSL) in a review commissioned by SKI (Wilmot and Crawford 2000). Several areas where further work and documentation was needed were identified.

The Swedish regulators have been active in the field of performance assessment[1] for many years and have developed sophisticated approaches to the development of scenarios and other aspects of assessments (see, for example, SKI (1997) and Stenhouse *et al.* (2001)). These assessments have generally used dose as the assessment end-point. The recent introduction of a risk criterion has, therefore, required an examination of the implications of a change in end-point on the type of calculations conducted and the structure of the assessment.

This report evaluates approaches to risk assessment for radioactive waste disposal, and examines the definition and presentation of the results of such risk assessments in safety cases to meet risk targets. The objectives of the report are to illustrate the strengths and weaknesses of different possible approaches to risk assessment by

[1] The term performance assessment is used in a generic sense in this report to cover all approaches to assessing the long-term behaviour of a facility. The term risk assessment is used in a more specific sense to cover assessments that use risk as a measure of performance.

reference to assessments in other countries, and to provide suggestions for future activity and development in this area by the Swedish regulators.

Following this Introduction, Section 2 of the report discusses the concept of risk and its use in regulations and assessments of disposal systems, drawing lessons as appropriate from the regulations and assessments reviewed. Section 3 of the report summarises the main conclusions from the review and presents suggestions for further work on preparing for SKI's and SSI's regulatory reviews of SKB's forthcoming proposals. Two Appendices present summaries of the documents reviewed.

2 Approaches to Risk Assessment

2.1 Introduction

Performance assessments provide the principal means of investigating, quantifying and explaining long-term safety of a selected disposal concept and site for both the appropriate authorities and the public (OECD/NEA 1991). Assessments of long-term safety rely on both qualitative judgements and quantitative modelling. An important aim of these assessments is an evaluation of performance against a regulatory measure, such as dose, risk or cumulative release of radionuclides.

The regulatory measures in force not only determine the performance measures that are calculated, but also influence the overall way in which the assessment is conducted. In particular, there has, historically, been a distinction between assessments that use probability to represent uncertainty (probabilistic assessments) where the regulatory measure is risk, and assessments that use other approaches to account for uncertainty (deterministic assessments) where the regulatory measure is dose.

In addition to Sweden, three countries have established risk-based regulations or guidance relating to the performance of disposal facilities for radioactive waste and / or explicitly require the use of probabilistic techniques in assessments:

- United Kingdom (Environment Agency *et al.* 1997).
- United States of America (EPA 1993, 1996, 2001; NRC 2001).
- Canada (AECB 1985, 1987).

The relevant parts of the regulations and regulatory guidance in these countries are summarised in Appendix A.

One assessment (the Compliance Certification Application (CCA) for the Waste Isolation Pilot Plant (WIPP)) has been used to demonstrate compliance of a disposal facility with a probabilistic performance measure, and several other assessments of designs or concepts have been undertaken with risk or other probabilistic measures as an end-point. The following assessments are summarised in Appendix B:

- United Kingdom: Her Majesty's Inspectorate of Pollution Dry Run 3 (Summerling 1992).
- United Kingdom: Nirex 97 (Norris *et al.* 1997).
- United States: Compliance Certification Application for the Waste Isolation Pilot Plant (DOE 1996a).
- United States: Total System Performance Assessment for the Yucca Mountain Viability Assessment (DOE 1998).
- Canada: Postclosure Assessment of a Reference System (Goodwin *et al.* 1994).

In the following sections, the similarities and differences between the assessments and regulatory regimes are discussed and used to support the derivation of suggestions for future activities by the Swedish regulators.

2.2 Assessment Structure

All of the performance assessments examined comprise a similar set of activities, even if there is a difference in the terminology applied to the stages in different programmes. The key steps are:

(i) Definition of the disposal system and the features of concern.

(ii) Broad identification of the possible future evolution of the selected disposal system (scenario development), and the consideration of the likelihood of occurrence of alternative scenarios.

(iii) Development and application of appropriate conceptual, mathematical and numerical models and codes, together with associated parameter values, for simulating evolution of the disposal system.

(iv) Evaluation of potential radiological consequences and associated risks.

(v) Uncertainty and sensitivity analyses.

(vi) Review of all components of the assessment.

(vii) Comparison of the results with appropriate criteria.

Although uncertainty analysis is highlighted as a separate stage in this structure, the acknowledgement and treatment of uncertainties are important components of scenario development (Stage (ii)), conceptual model development and parameter value definition (Stage (iii)). It is the way in which uncertainties are treated in these stages that is the key difference between the different types of assessment.

2.3 Classification of Uncertainty

Before analysing the different ways in which uncertainties are treated, it is useful to examine the different types of uncertainty that have been recognised in assessments. Three inter-related and overlapping categories are commonly recognised:

- Uncertainty in the future evolution of the disposal system (often referred to as scenario uncertainty).

- Uncertainty in the models used to represent this evolution.

- Uncertainty in the parameter values used in the modelling programme to evaluate the potential consequences of scenarios.

All of these uncertainties contribute to uncertainty in the estimated performance of the disposal system.

Scenario uncertainty: Over the timescales relevant to an assessment of geological disposal, both the natural environment and the engineered features will change due to natural processes, interaction of the natural environment with the disposal facility and wastes, and human actions (unrelated to the disposal). There is uncertainty over the exact nature of such changes, resulting in uncertainty as to the future state of the disposal system.

Model uncertainty: Quantitative PAs are conducted using a suite of models that describe the possible evolutions of the various components of the disposal system. This suite of models includes conceptual models (sets of assumptions that describe system or sub-system behaviour), and mathematical models (formal mathematical descriptions of the conceptual models). In cases where the mathematical models cannot be solved analytically, computer models or codes are required to allow numerical solutions. Simplifications and assumptions are almost always introduced in the development of conceptual models of the real world, in the development of representative mathematical models, and in the numerical solution of the mathematical equations. In addition, conceptual models of one disposal subsystem may be developed at different levels of detail for different purposes in an assessment. For example, highly detailed research models may be constructed to evaluate specific processes, based on a theoretical framework, supported by laboratory and field studies. These research models and their associated databases may be simplified to form computationally tractable assessment models. Simplifications and assumptions introduced at all these stages introduce model uncertainty.

Parameter uncertainty: This uncertainty may be associated with measurement error, spatial variability, or insufficiency of data to parameterise the system. A substantial effort is required to obtain and interpret sufficient data to adequately characterise a site; even so, it will not be possible to develop a complete understanding of the geological environment.

A key reason for the inability to fully characterise a site is that many geological properties vary at a scale that is less than the region of interest (typically kilometres) but greater than that of measurements in boreholes or outcrops (typically less than a metre). If this spatial variability can be characterised, for example if there is a uniform trend, then parameter values can be interpolated at points between measurement locations. Parameter uncertainty from this type of variability will be relatively low. However, in many cases the pattern of spatial variability is uncertain so that there are large uncertainties in parameter values at all points other than measurement locations. Statistical descriptions of spatial variability can be useful in modelling overall system behaviour in such cases, but do not necessarily reduce parameter uncertainty at specific locations.

The recently published SKI guidance [SKIFS 2002:1] identifies each of these types of uncertainty as being relevant to the safety assessment, and also acknowledges that the distinction between the three types of uncertainty is not always clear-cut. For example, uncertainty as to *when* a particular event occurs (e.g., glaciation, fault movement) could be classified according to the above definitions as either scenario uncertainty or parameter uncertainty. The choice between how something is classified may depend not only on the nature of the uncertainty but also on the way in which the assessment calculations are conducted, and hence on the purpose of the calculations and on any guidance or requirements specified in regulations.

There is another classification of uncertainties that cuts across these categories:

- *Epistemic uncertainty* is associated with data from site characterisation and laboratory experiments. Uncertainties may be large, and the experiments or characterisation programmes necessary to reduce them may be difficult and expensive to conduct. However, in theory, the acquisition of more data will reduce this type of uncertainty. This type of uncertainty has also been termed subjective uncertainty (DOE 1996a).

- *Aleatory uncertainty* is associated with events, such as future human activities, for which there is not, and cannot be, observational data. No amount of additional study can provide additional quantitative information about this type of uncertainty. This type of uncertainty has also been termed stochastic uncertainty (DOE 1996a).

As with scenario, parameter and conceptual model uncertainties, this classification is somewhat subjective, and the assignment of a particular uncertainty to a particular classification may depend on the context and purpose of the overall assessment as well as on the nature of the uncertainty. The characterisation of epistemic uncertainties should be based on data from experiments or site measurements, but some degree of expert judgement will be required before these data can be used in assessment calculations (e.g., to select appropriate data ranges, or exclude anomalous data values). In contrast, the inclusion of aleatory uncertainties in assessment calculations always requires expert judgement to define data values, probability density functions (pdfs) or other information, because there are no measured values on which to base them. An example of uncertainties that could be classified as either epistemic or aleatory are those related to future seismic events. These cannot be directly observed, but additional analysis of past records could reduce uncertainties about future seismic activity.

2.4 Treatment of Uncertainty

All assessments of the post-closure performance of radioactive waste disposal systems need to account for the uncertainties inherent in the long-term behaviour of complex natural systems. Some regulations and regulatory guidance are prescriptive about the approaches to be used, but others provide only general guidance. The recommendations accompanying the recent SKI regulations [SKIFS 2002:1], for example, require that uncertainties are examined both in the selection of calculation cases and in the evaluation of results. However, apart from proposing that both probabilistic and deterministic approaches should be used to complement each other, no additional detailed approaches are described.

The following sections provide a background discussion of the principles involved in the approaches to the treatment of uncertainty reviewed in this report.

2.4.1 Use of probability

There tends to be a distinction between deterministic and probabilistic assessments that mirrors the use of dose and risk as regulatory measures, although this is not a necessary distinction. The results of a probabilistic calculation can, for example, be expressed in terms of dose, either as a distribution or as a single, expectation, value. It is also possible to convert the results of a deterministic dose assessment to a risk by using the dose-risk conversion factor recommended by the International Commission for Radiation Protection (ICRP) for expressing the uncertainty in the response to receipt of a dose. However, the definition of risk made explicit in a number of regulations indicates that other uncertainties should also be considered in a risk assessment.

The definition of risk in the SSI regulations [SSI FS 1998:1] regulations is:

"… the probability of the harmful effects (fatal and non-fatal cancers as well as hereditary damage) as a result of an outflow from the repository, taking into account the probability of the individual receiving a dose as well as the probability of harmful effects arising as a result of the dose." The SKI regulations require that the safety assessment should demonstrate that the design and construction of a facility will allow it to meet the risk criterion in the SSI regulations [SSI FS 1998:1].

The above definition, and analogous definitions in other regulations and guidance, requires that the uncertainties regarding the receipt of a dose are assessed as well as the uncertainties associated with how a given dose affects an individual.

The term "probability" is applied to both of these uncertainties, but there is a difference between them which is also reflected in two different concepts of probability:

Frequentist. This approach to probability is identified with the long-run frequency of an event or process. In other words, how often an event takes place over a long period, or the number of times a particular outcome arises if a process is repeated a large number of times. Epidemiological data, such as the dose-risk factor, is an example of this frequency approach - a large number of individuals are known to have been exposed to a dose, but only a proportion have died or developed cancer. This data can be directly extrapolated to the probability of an exposed individual dying or developing cancer. Because the frequency can be measured or assessed, this type of probability can be regarded as an objective property of the system.

Subjectivist. Under this approach, the concept of probability expresses the "degree-of-belief" of an observer. A key element of this approach is that the degree-of-belief is dependent upon the information available about the system or value being considered. If more information becomes available, then the probability distribution is likely to change. If, for example, a new measurement is outside the range considered likely, then a greater degree of uncertainty will need to be incorporated. Conversely, if many new measurements have similar values, then the degree of uncertainty may be reduced. Probability in this sense is, therefore, not an objective property of the system under study, but is subjective or contingent upon available information.

There are some parallels between these two approaches to probability and the classification of uncertainties as aleatory (stochastic) and epistemic (subjective) described above. However, the classifications of uncertainty and probability are not distinct and, despite the similar terminology, there is not a direct equivalence between the classifications. In terms of data that are normally expressed as frequencies, for example, rates of human intrusion into a repository are aleatory whereas epidemiological data are epistemic. The difference lies in the extent to which measured data can be extrapolated. If there is no basis for the extrapolation, then the only way in which frequencies can be determined is by an *a posteriori* analysis - e.g., how many times was the repository actually intruded into over 10,000 years. In the case of epidemiological data, extrapolation is justified and so a parameter that is expressed as a frequency would nevertheless be classified as an epistemic or subjective uncertainty.

The extent to which probabilities that express "degrees-of-belief" can be justified varies between parameters, and according to whether the probability represents stochastic or subjective uncertainties. The distinction between different types of uncertainty was an important aspect of the WIPP CCA probabilistic assessment, but in general there is little to be gained directly from this classification of uncertainties and probabilities. What is important is transparent and traceable documentation that sets out the data, assumptions and judgements on which models and parameter values are based.

The regulations reviewed in this report require or recommend the use of probabilistic techniques for the treatment of uncertainty, either explicitly or implicitly through the definition of risk. All of the assessments reviewed, therefore, use probabilistic techniques, but other means of treating uncertainty are also used, although not always explicitly acknowledged. Similarly, some but not all of the regulations and regulatory guidance reviewed acknowledge that a variety of means can be used to address uncertainty in a risk assessment. The approaches used are discussed in the following sections that describe the treatment of scenarios, alternative models and parameter values.

2.4.2 Treatment of scenarios

Although not all of the regulations and assessments reviewed include a definition of a scenario, all of them recognise that assessments require a broad description of the disposal system and its evolution as the basis for developing assessment models. There are differences in the way in which assessment models treat the evolution of the disposal system, and two principal approaches, the **scenario** and the **simulation** approaches, have been identified. In practical terms, however, the distinction between these approaches is not clear-cut, and much of the debate about the differences arises from the way in which one particular aspect of the system, climate change, has been treated. In other respects, all assessments are based on sub-sets of the universe of all features, events and processes (FEPs), and these sub-sets fulfil the generally accepted definition of a scenario (OECD/NEA 1992):

> *A scenario specifies one possible set of events and processes, and provides a broad- brush description of their characteristics and sequencing.*

The selection of which FEPs to exclude from an assessment is based on a variety of screening criteria, including low consequence to disposal system performance, low probability of occurrence and exclusion based on regulatory requirements. The remaining FEPs are then divided into one or more consistent sets, or scenarios, for analysis. A common division is between the set of "normal evolution" FEPs and those involving disruption of the disposal system. A further sub-division is that between naturally-occurring disruptive events and disruptive events caused by future human actions.

The next stages of the assessment process, model development and parameterisation, can be conducted for all of the scenarios identified. These lead to conditional consequences for the individual scenarios, but do not necessarily address the uncertainty associated with the occurrence of the scenarios. This uncertainty can be treated in a number of ways:

Simulation. In the WIPP CCA, a number of "disturbed" performance scenarios were identified, depending on whether the disruptive event was drilling or mining. These broad scenarios were further sub-divided, depending on the timing and sequence of the disruptive events. These sub-scenarios were simulated in the assessment calculations by sampling the time of occurrence of the disruptive events. A similar approach was used in Dry Run 3, which used simulation techniques to generate sequences of climate states instead of using separate scenarios for particular climate conditions. In these simulation approaches, the probability of each set of conditions is not explicitly defined, but is implicitly defined by the number of simulations conducted. In other words, if 100 simulations are carried out, then the probability of each set of conditions or sequence of events is 0.01.

The simulation approach addresses one aspect of scenario uncertainty, but still provides only a conditional consequence if the simulated system is not an exhaustive description of the overall disposal system and its possible evolution.

Scenario probability. In this approach, independent calculations are performed of the consequences of each identified scenario. The probability of each scenario is also assessed, and used to develop a probability-weighted measure of system performance. The difficulty of this approach lies in the definition of scenario probabilities.

Scenarios should be exclusive and exhaustive. In other words, there should be no overlap between scenarios, and there should be no events or situations that are not included within a scenario. If these conditions are met, then the sum of scenario probabilities will be one. This in turn means that the probability of one scenario can be determined by subtraction. For example, if the probability of occurrence of the disruptive events which define the "disturbed" performance scenarios are determined, then these assumptions allow the probability of the "undisturbed" scenario to be defined.

In practise, because the probability of the "disturbed" performance scenarios is low, some assessments maintain the probability of the "undisturbed" scenario at one, instead of reducing it by the probability of the disturbed scenarios. This is likely to have only a small effect on any overall value of risk that is calculated, but it is an assumption that should be explicitly acknowledged. A probability sum of greater than one also makes it more difficult to assess whether the conditions of exhaustiveness and exclusivity have been met.

Although the probability of the "disturbed" performance scenarios is low, their greater consequences mean that they may have a significant effect on the overall calculated risk. It is important, therefore, that the uncertainties relating to the estimated probability are examined. The SSI:s commentaries on the regulations [SSI FS 1998:1] note that the probabilities and consequences should be estimated for a sufficiently exhaustive set of scenarios, so as to provide a comprehensive illustration of risk. It is also said that scenarios resulting in doses exceeding 1mSv/y a separate should be presented separately. The SKI recommendations concordantly propose that the probabilities that the scenarios and calculation cases will actually occur should be estimated as far as possible in order to calculate risk. The recommendations also propose that both probabilistic and deterministic approaches should be used in an assessment. **Worst-case scenario**. In this approach, the consequences of all scenarios are calculated, but no attempt is made to quantify scenario probability. Instead, the conditional consequences of each scenario are compared to the regulatory criterion. If the regulatory criterion is risk, then the scenario probability is assumed to be one, giving a conditional risk. If the conditional risk is less than the regulatory criterion for all scenarios, then the probability of different scenarios may not need to be determined.

If all disruptive events, including those related to future human actions, are considered in a risk assessment, then there is almost certainly a "worst-case" scenario that can be envisaged whose conditional consequence will exceed the regulatory criterion (e.g., prolonged handling of excavated materials, large-scale excavations). Such extreme events will be of very low probability, but their effect on calculated risk can only be lessened if this probability is evaluated. Strictly speaking, therefore, this "worst-case" approach is only useful in demonstrating *numerical* compliance if there are constraints on the severity of the disruptive events considered. In practise, because regulatory decision-making includes factors other than simply comparison of calculated consequences with a criterion, qualitative arguments concerning low-probability events may be acceptable, and this approach can still be of use.

The most effective means of excluding severe disruptions from an assessment is via regulatory exclusion, although different stakeholders may have different views on the types of event that should be considered. The Dry Run 3 regulatory assessment excluded human intrusion from the system calculations, although scoping calculations were undertaken. These were based on historical drilling rates, which were judged to be pessimistic. Further work on justifying intrusion rates was noted as a key issue, but no further guidance on the scope of intrusions to be considered has been provided by the UK regulators. Moreover, published guidance in the UK (Environment Agency *et al*. 1997) does indicate that deterministic calculations may be an appropriate means of addressing some uncertainties.

The clearest regulatory statement regarding the treatment of future human actions is the recent Swedish regulation [SSI FS 1998:1], which specifically states that the principal risk assessment should not consider disruption from future human actions. This exclusion reduces the extent of speculation and the associated arbitrary assumptions and parameter values concerning societal evolution and human activities. It also provides an explicit acknowledgement, lacking in other regulations and guidance, that conditional risk calculations are an acceptable basis for demonstrating compliance.

The use of conditional risk calculations is also provided for by the recently published SKI regulations and recommendations on safety assessment [SKIFS 2002:1]. This guidance describes three types of scenario: the **main** scenario, which includes the expected evolution of the disposal system; **less probable** scenarios, which include alternative sequences of events to the main scenario and also the effects of additional events; and **residual** scenarios, which evaluate specific events and conditions to illustrate the function of individual barriers. The residual scenarios should include the direct effects of human intrusion, and the consequences of an unclosed repository. The guidance notes that residual scenarios should be evaluated independently of the probabilities of the events, which means that this group of scenarios will not be included in the overall calculation of risk, but will be evaluated as "what-if" scenarios.

2.4.3 Treatment of model uncertainty

There are two approaches to incorporating model uncertainty into risk assessments. These have been termed lumping and splitting, and differ essentially in terms of whether the alternatives are integrated into a "meta-model" or assessed separately. Some assessments take both approaches depending on the exact nature of the alternative models.

Two examples of the use of a "meta-model" come from the WIPP CCA. Several two-phase relative permeability models can be used to describe two-phase flow (gas and water) through anhydrite. Analysis of experimental data from tests on cores showed that either a modified Brooks-Corey model or the van Genuchten-Parker model could be used to describe the data. The PA model requires data to be extrapolated beyond the range covered by the experimental results and these two different models give different results for this extrapolation. PA calculations therefore sampled between the different models. Similarly, there are alternative models for the microbial degradation of plastics and rubbers in the waste in the repository which lead to different amounts of gas and hence different pressure within these repository. There is no experimental data to support one of these models as being more appropriate, and so the PA calculations sampled between them.

When alternative models are integrated within a single assessment model, then there is need of a mechanism for selecting between the models for each simulation. This is generally done through use of an index parameter that can take one of two values (or more if there are more than two alternative models). A probability is assigned to each value so that sampling selects appropriately from the alternatives.

The disadvantages of the lumping approach are related to whether the alternatives are exclusive and exhaustive. It is easier to demonstrate that alternatives are exclusive (i.e., that there is no overlap between the alternatives), than that they are exhaustive (i.e., that further alternatives do not exist). In either case, if these conditions are not met, then the probabilities of the alternatives will not sum to one, and an index parameter cannot be properly defined. Even if these conditions are met, or the probability of further alternatives is very low or otherwise neglected, then establishing a pdf for the index parameter remains problematic.

Assigning a degree-of-belief to a set of alternative models requires the use of judgement by experts familiar with the models. There may be cases in which there is broad agreement on an appropriate degree-of belief (e.g. an alternative model included because it leads to larger consequences, but which has a universally-held low degree-of-belief). In general, however, each alternative model will have its proponents who will have a high degree-of-belief in its applicability and a correspondingly low degree-of-belief in the other models. In the majority of cases, therefore, there will not be agreement on the relative merits of the alternatives, and the result is that alternatives are assigned arbitrary, equal probabilities. This ensures that the alternatives are used in the assessment, but does not necessarily lead to a greater level of system understanding or confidence. Indeed, this approach can lead to the phenomenon of "risk dilution", whereby the calculated risk is reduced because the spread of uncertainty considered is inappropriately wide.

The principal advantage of lumping is that only one set of assessment results is produced. If alternative models exist for more than one topic, then the number of separate analyses required to explore all the possible combinations of models may become significant.

The advantage of splitting is that it more readily allows the effect of the alternative models to be assessed. If other parameters are treated probabilistically, then two separate analyses, run with otherwise exactly the same inputs and sampled parameter values, will be easier to interpret than a single analysis that samples between alternatives. In the latter case, it is unlikely that there will be directly equivalent simulations, although broad trends and differences will be apparent if there are sufficient simulations.

The principal disadvantage of splitting corresponds to the main advantage of lumping; that is, there will be large numbers of analyses to interpret if there are alternative models in several topic areas.

Lumping is feasible when the alternative models can be readily implemented in an assessment code, for example as an alternative equation or as different coefficients. In cases where there are greater differences between the alternatives, for example, alternative mathematical or computational models or different conceptual models for major system components, the alternatives are less easily implemented in a single "meta-model". In these cases, a system code that allows different sub-models to be linked together is required. Providing the outputs are compatible, alternative models can be implemented as sub-models and linked as appropriate. However, unless the system code allows for sampling from different sub-models, this approach only allows the alternatives to be analysed independently (splitting).

2.4.4 Treatment of parameter uncertainty

The majority of the uncertainties that must be addressed in risk assessments do not satisfy the criteria for being expressed as frequencies, and therefore, if they are expressed as probabilities, they must be expressed as "degrees-of-belief". This means that there must be an element of judgement applied in determining pdfs as the available evidence must be interpreted in terms of a number of factors, including:

- The purpose of the assessment.

- The form of the mathematical model, and any biases or approximations in the model.

- Spatial variability of the measured property.

- Differences between the experimental situation and the modelled environment.

As noted in Section 2.3, for some parameters the treatment of parameter uncertainty is linked to the treatment of spatial variability. If the spatial variability can be described using a simple trend surface (i.e., a uniform change across the region of interest) then a deterministic relationship can be used to calculate parameter values. However, in the majority of cases, spatial variability is more complex and more sophisticated techniques are required to describe the variability and to allow interpolation and extrapolation. These techniques are generally termed geostatistics, and a wide variety of techniques has evolved based either on generalised assumptions about spatial heterogeneity or on distributions that are conditioned by observations at a number of points (Ababou *et al.* 1992; Zimmerman and Gallegos 1993).

It is important to stress that geostatistical descriptions of spatial variability, whether generic or conditioned, are not phenomenological and thus they cannot be used to make predictions about the way the system might evolve. They do, however, provide a means of generating different data-sets that are consistent with observed data and that account for uncertainties in the unobserved parts of the system. In concert with groundwater flow and transport models these data-sets can be used to evaluate uncertainties in the behaviour of the system (e.g., fluxes). A good example of this approach is the generation of transmissivity fields for an aquifer above the WIPP site.

Because they are not phenomenological, geostatistical descriptions, or any other statistical descriptions, cannot be evaluated in the same way as physically-based models. The results of a single experiment or observation can be enough to falsify a physical model, but experiments cannot be devised with the aim of falsifying geostatistical descriptions. If sufficient extra data are gathered, it may be possible to demonstrate that a particular description performs less well than an alternative, but in the context of radioactive waste disposal the integrity of the site may be compromised if invasive techniques are needed to collect the data (Mackay 1993).

The use of geostatistical techniques does not obviate the requirement to consider alternative conceptual models. A geostatistical description of a particular set of features is a means of accounting for uncertainty in those features. Such a description does not, however, account for the uncertainty associated with using that set, rather than a conceptually different set, to describe the system (Hodgkinson 1992). An analogy would be using a normal distribution to account for uncertainty in the widths of channels within fractures. Sampling from such a distribution does not account for uncertainty as to whether channels or capillary bundles are appropriate descriptions of fracture flow.

Previous work for the Swedish regulators (Wilmot and Galson 2000; Wilmot *et al.* 2000) has examined the role of judgements in performance assessments and the use of expert elicitation. Each of the different types of judgement will be used in the derivation of parameter pdfs for a probabilistic assessment, depending on the type of parameter involved and the stage in the assessment programme that the judgement is being made.

A key area for which parameter values must be elicited is that of aleatory uncertainties. These are uncertainties that cannot be reduced through further site characterisation or experiments, and include topics such as future human intrusion. Although human intrusion need not be considered as part of the principal risk assessment under the Swedish regulations, it is a topic that must be addressed in a safety case and so there may still need to be some expert elicitation. Expert elicitation may also be used for parameters characterised by epistemic uncertainty, but for which the necessary site characterisation or experiments require too great a level of resources.

Parameter values that are not determined through elicitation nevertheless also require some judgements to derive pdfs from the available data. A useful classification of parameters that helps to determine the type and extent of judgements required has three principal categories:

- Prescribed (e.g., represents an international standard).
- Generic.
- Site-specific.

Prescribed data are generally constant values and require no further judgements. Generic and site-specific data can be classified as well characterised or poorly characterised, and this classification will affect the extent of judgement required in deriving pdfs from the available data. In the case of well-characterised data, the form and indices of the pdf will be easily determined[2]. More extensive judgements are required to derive pdfs from poorly characterised data, with decisions required on the applicability of data values, the type of distribution that best characterises the data uncertainty and the indices of the selected distribution.

If a probabilistic assessment were to be conducted once only, a large number of parameters would need to be specified in the form of pdfs because there would be uncertainty as to the main influences on risk. Each of these pdfs would require justification and documentation that allowed traceability back to raw data. However, if the assessment was conducted iteratively, the early iterations could use generalised pdfs (e.g., uniform or triangular distributions) with more limited documentation. These early iterations would develop knowledge of the disposal system and identify the key parameters that govern the calculation of risk. In later iterations, the data derivation and documentation effort could be focused on these key parameters. Two approaches can be adopted for parameters to which models are less sensitive:

- Best estimate values can be adopted for these parameters. This may simplify the calculations and the number of simulations required to demonstrate convergence of the results, but justification for the best estimate values will be required.

[2] For example, a triangular distribution requires three indices (minimum, mode and maximum), and a normal distribution requires two indices (mean and standard deviation).

- Pdfs accounting for uncertainty in these parameters can be retained. No further justification is required, and any changes in model sensitivities brought about by model or data changes are not inadvertently neglected.

The best example of this iterative approach to performance assessment is the series of assessments undertaken for the WIPP site in the US. Early iterations helped to develop site and disposal system understanding and to focus research effort into areas that were significant in terms of reducing uncertainty and increasing confidence. For example, PAs prior to 1989 showed that brine inflow into the repository was a key issue, prompting further geophysical investigations. The introduction of 2-D flow and transport models into the 1990 PA showed that the nature of groundwater flow in the aquifer overlying the repository was a key issue, leading to further site characterisation. Using geostatistics to describe hydrogeological patterns in the 1991 PA prompted the development of a regional groundwater model, the results of which were incorporated in the 1996 PA. Similarly, the introduction of gas effects in the 1991 PA demonstrated the importance of these processes to performance, and led to the inclusion of a detailed gas generation model in the 1996 PA. As an example of the iteration between PA and other studies, see DOE 1990.

All of these iterations were published and comments sought from a wide range of stakeholders. The first formal submission to the EPA was the Draft Compliance Certification Application in 1995, and this was followed by the Compliance Certification Application (CCA) in 1996. In response to comments and as system understanding developed, the documentation of the assessment developed so as to provide additional information on the assumptions made and the basis for these and the selected parameter values.

2.5 Models and Codes for Uncertainty Analysis

At a fundamental level, the calculations required for calculating risk in performance assessments do not differ greatly from those required for other end-points. In practice, however, there are differences relating to the number of calculations required, the complexity and robustness of the models, and the treatment of uncertainty, which mean that models and codes appropriate for calculations of dose may not be entirely suitable for calculations of risk. The different requirements of risk calculations relate in large part to the way in which uncertainties are accounted for:

- **Parameter uncertainty**. If probabilistic techniques are used to account for parameter uncertainty, a method of sampling from pdfs is required. Similarly, a method is required for combining results from different samples into a results pdf. Sampling could be undertaken independently to generate sample datasets for input to an unmodified assessment code. Results from each of the separate code runs could be combined using a standalone data analysis tool. However, for efficiency, probabilistic models usually incorporate an integrated control module that undertakes sampling and integration of results. Configuration management tools keep a record of the model set-up for each case and allow results to be reproduced and traced back to particular sets of input values.

- **Scenario uncertainty**. The approaches used to account for scenario uncertainty depend on the way in which scenarios are defined. If scenarios are regarded as similar futures, then a similar approach to that used for parameter uncertainty can be used, with sampling of, for example, the timing and magnitude of scenario-defining events, such as faulting, canister failure or intrusion. If scenarios are defined on a broader scale, however, then it may be more appropriate to conduct separate calculations of scenario consequences using models and codes optimised for particular scenarios. In this case, additional processing of the results to generate a probability-weighted consequence (risk) will also be required. For example, a model accounting for 1-D flow and transport may be used for calculations of a fault-disrupted facility, whereas a full 3-D model may be used for calculations of the expected evolution of the disposal system.

- **Model uncertainty**. Performance assessment codes may be designed around a control structure that allows alternative models of different parts of the system to be easily linked into the overall model. These codes allow different model configurations to be defined at run-time, and may allow different simulations to use different component models. In the absence of such an integrated control structure, the extent to which alternative models can be incorporated in a single code is more limited, and may be restricted to alternative equations for calculating a particular parameter. A control variable to select between such alternatives can be selected or sampled from a pdf. An alternative approach to model uncertainty is similar to the approach used for scenario uncertainty, i.e., to use independent models to determine conditional consequences and then to probability-weight these consequences in a calculation of risk.

Risk assessments use a range of different models, including conceptual, mathematical and computational models, all of which need a demonstration that they are fit-for-purpose. The term validation has been applied to this demonstration, but formal validation is rarely possible in the context of models that address large spatial and temporal scales (see, for example, Wingefors *et al.* 1999; Wilmot and Galson 1994). In this context, the demonstration that the models are fit-for-purpose is part of the overall confidence building process, and a number of approaches can be used.

For site characterisation, every effort should be made to integrate the modelling process and the site characterisation process so that model results are used to predict conditions ahead of characterisation. Such predictions can be either of directly measurable parameters, or of derived parameters that depend on other assumptions. For example, hydrogeological models can be used to predict the groundwater head in a borehole prior to drilling, and to predict the results of pump tests and other hydraulic conditions before they are measured.

Successful prediction of site characterisation data provides confidence in the model concerned, and also provides a basis for determining when sufficient site characterisation has been completed. However, there is no absolute measure for successful prediction, since the context of the prediction and the uncertainties in the models must be taken into account in determining whether a particular measurement negates a model or not. Similarly, the frequency with which models are updated to account for additional data is dependent on the model context, its use in other parts of the programme and the scale of associated uncertainties. It is important that the basis for models is well documented. This can be done through periodic data freezes (typically annual), each of which is followed by assessment and, if required, modification of the models. Alternatively, an integrated data structure that permits data to be traced in both directions allows models that would be affected by new data to be identified and allows the basis for models to be readily determined.

There is a perception that probabilistic codes must incorporate simpler models of system behaviour than the equivalent codes used for deterministic calculations. There is no *a priori* reason for this to be the case, although there are reasons why it is true in practice. One reason put forward is the computing resources required for probabilistic calculations, which may be several orders of magnitude greater if there are large numbers of pdfs to be sampled. Sampling techniques such as Latin Hypercube Sampling (LHS) can be used to optimise the effectiveness of sampling and reduce the number of samples required to explore all parts of parameter space and achieve convergence. If the code takes several hours for an individual run, then the computing burden of a large probabilistic case could be very large (1000 simulations each taking 4 hours would require more than 166 days of computing time). Steps can be taken to reduce the elapsed time by using several computers in parallel, and this would also be more robust and reduce the effects of machine failure.

A second reason for models used in probabilistic calculations being simpler than the corresponding model for deterministic calculations relates to the degree of uncertainty that is to be addressed by the model. A detailed 3-D flow and transport model will, typically, be set-up to correspond to a particular site conceptualisation, and may also be calibrated against measured heads or other parameters. Some uncertainties can be explored with such a model (e.g., sorption coefficients and other transport parameters), but the extent to which boundary conditions or parameters governing flow can be varied without invalidating the model may be limited. Similarly, changes to the modelled domain to accommodate, for example, new faults or erosion, may invalidate any calibration. A model that allows the full range of uncertainties to be explored, along with changes in boundary conditions and model domain, needs to be robust; this generally means a less complex model.

One feature of assessment programmes that use probabilistic techniques is a tendency to develop site- or concept-specific models and codes. This is in contrast to programmes conducting deterministic calculations where there is a greater tendency to use or adapt existing commercial or public-domain codes. Deterministic calculations can be conducted for individual sub-systems, with the output from one model or set of models being used as input to another model. This means that assessment programmes can use generally available codes, with any specific requirements of a particular programme being met using pre- and post-processing techniques, or by the development of bespoke models for particular sub-systems.

Probabilistic calculations conducted as a series of linked calculations using independent codes may require a series of pre- and post-processors to ensure that probabilistic results are correctly passed between codes, and also to ensure compatibility of sampled data between different parts of the overall system (see, for example, DOE 1996b). Although these problems can be overcome, probabilistic calculations are more readily conducted using an integrated system model designed to use a single sampling protocol so that compatible data are used in different sub-models. These integrated models may use established models and routines, but it is the integration within a single control system that is key to traceability and configuration management.

2.6 Regulatory Criteria and Guidance

The previous sections have discussed the overall structure of performance assessments, the different types of uncertainty that must be taken into account and the approaches that can be used to account for these uncertainties. In this section, the discussion focuses on what performance assessments are actually required to calculate. These requirements are generally specified as regulatory criteria, or are described in supporting regulatory guidance. The three main issues discussed are:

- The time-scales over which disposal system performance must be assessed.
- How risks are calculated and presented.
- Alternatives to risk as a measure of performance.

2.6.1 Time-scales for assessments

The regulations and regulatory guidance reviewed in this report either prescribe a relatively short (10,000 years) period for which disposal system performance must be demonstrated, or are open-ended about the time-scale. In the latter case, the time-scale selected in performance assessments is generally one million years for calculations of dose and risk. This is a significant difference, and it is useful to examine the reasons for the difference and to examine whether there is another approach.

Three key arguments are put forward in support of restricting the time-scale of an assessment. First, it is argued that radioactive decay will reduce the inventory over long time-scales. Second, it is argued that all the events and processes expected to affect the disposal system, and thus the peak dose, will have occurred by 10,000 years. Third, it is argued that the increasing level of uncertainty makes the results of long-term calculations unsuitable for comparisons with numerical criteria. The second argument is, however, only applicable to disposal concepts which do not take credit for the long-term effectiveness of waste containers and where the site will be unaffected by glaciation, when the peak dose is likely to occur later.

The key argument for assessing performance over a long time-scale is that the same standard of protection should be applied to future generations as to the present. It is therefore the peak dose that should be compared with any criterion rather than the dose at a particular time. If there is effective containment by engineered barriers until a significant proportion of the inventory has decayed, or if processes such as glaciation will affect the site in several thousand years time, then the peak dose is unlikely to occur within the first 10,000 or even 100,000 years.

The key argument against a quantitative assessment over long time-scales is the increasing level of uncertainty about the state of the disposal system and the processes acting on it.

In discussing these uncertainties, it is useful to divide the overall disposal system into the conventional sub-systems of near-field, far-field and biosphere. These sub-systems have different characteristics in terms of present-day knowledge and long-term uncertainty:

Near-field. The characteristics of the near-field are relatively well-constrained at site closure, because the facility will have been well-characterised during construction and monitored during the operational phase. Depending on the type of waste and containment system, there will be a period of relative stability in near-field conditions, once re-saturation has occurred. In the long-term, however, uncertainties will increase as physical and chemical barriers degrade.

Far-field. Site characterisation and monitoring during site selection, construction and operation will reduce uncertainties in the far-field, but the spatial extent of the region involved and the necessary use of remote sensing rather than direct observation mean that there will be an irreducible degree of uncertainty concerning the far-field. This uncertainty will increase with time as external events such as glaciation and seismic activity change the boundary conditions on the far-field. In a geologically stable region, however, the extent of change and the associated increase in uncertainty will generally be less than for other sub-systems.

Biosphere. There are two aspects of the biosphere that can be considered in terms of characterisation and evolution. The physical characteristics of the biosphere (e.g., topography, soil types, climate) can be well characterised in terms of present-day conditions, as can the human activities and lifestyles (including agricultural practices and food consumption).

- The evolution of the physical characteristics of the biosphere can be modelled. In the case of environments where the rate of change is low, simple models, or even an assumption of no change, may be adequate in addressing the uncertainties. In more dynamic systems, however, the complexities of the interactions between the many processes in the biosphere will lead to an increase in uncertainty with time. With the time-scales involved in assessments of radioactive waste disposal systems, even relatively slow processes such as land uplift may be classified as dynamic in this sense.

- In the case of human activities, there are no feasible models for the evolution of the social structures that underpin human activities.

This brief comparison of the knowledge and uncertainties associated with the three principal sub-systems shows that the biosphere, in both its senses, is probably the greatest source of uncertainty in terms of system evolution. This is particularly the case for disposal facilities in Northern Europe, where there is an expectation that, after a period of global warming induced by human activities, the natural climate evolution will lead to the growth of continental-scale ice-sheets. The presence of ice up to 3 km thick above a disposal facility will have an effect on the groundwater system, but it will have an even more profound effect on the biosphere, changing the physical landscape through erosion and / or deposition of glacial sediments, and displacing human activities. Once the ice has retreated, there will be very large uncertainties as to the form of the physical landscape, and the pattern of human re-settlement and subsequent activities will be conjectural.

This increasing uncertainty with time can be taken into account either through use of different conceptual models for different time-scales (detailed for early times, simplified for the far future), or by adopting different performance measures.

The recently introduced Swedish regulation [SSI FS 1998:1] includes an example of how different levels of assumptions can be applied to different time-scales of an assessment:

> “For the first thousand years following repository closure, the assessment of the repository's protective capability shall be based on quantitative analyses of the impact on human health and the environment.
>
> For the period after the first thousand years following repository closure, the assessment of the repository's protective capability shall be based on various possible sequences for the development of the repository’s properties, its environment and the biosphere.”

Although providing a detailed requirement on the assessment of the first thousand years of repository performance, this regulation does not indicate the overall time-scale for which performance should be considered, or how uncertainties over much longer time-scales should be treated.

The recommendations accompanying the SKI regulation [SKI FS 2002:1] suggest that the time-scale of an assessment should be related to the hazard posed by the inventory in comparison to naturally occurring radionuclides. However, the difficulties of conducting meaningful analyses suggest that detailed assessments are not required for times beyond one million years.

The guidance issued by the Environment Agencies in the UK recognises that there is a limit to the period over which it is reasonable to consider quantitative performance against numerical targets, due to the increasing level of uncertainty over long time-scales. The responsibility for determining and justifying the time-scales considered remains, however, with the proponent. The time-scales will vary according to the type of wastes and the design of the disposal facility concerned.

Models of the evolution of the near-field and far-field can probably be justified for periods of tens of thousands of years but, as the Swedish regulations recognise, one thousand years is a more reasonable limit for models of the detailed biosphere evolution.

Beyond the period over which the biosphere can be modelled, the uncertainties in any biosphere models will be so great that illustrative calculations based on a number of assumptions or scenarios will provide greater levels of confidence than calculations with very large uncertainties about the future evolution of a single biosphere. By evaluating the radiological consequences of a range of plausible future biosphere conditions, it may be possible to show the robustness of the assessment, and also to identify key uncertainties that may warrant further evaluation. These biosphere scenarios can be based on analogues from different regions at the present-day, or maintenance of present-day conditions at the site.

An approach using a series of climate analogues was adopted in the Dry Run 3 assessment, and also in Nirex 97. Such an approach does allow for a single presentation of system performance for the entire assessed period, rather than having to integrate results from different calculations for different time-frames. However, these analyses do not fully account for all the uncertainties involved, but include only the conditional uncertainties associated with the selected conceptual model of climate and biosphere evolution. For example, there is generally no accounting for the transition between climate states or the associated uncertainties, such as the release of accumulated radionuclides from sediments. If conceptual model uncertainty concerning the state of biosphere was also accounted for, there would be greater uncertainties in the calculations of dose and risk and these calculations would be more difficult to interpret.

The second possible approach to addressing the increasing uncertainties associated with the biosphere is to adopt different performance measures for long time-frames. The recommendations accompanying the SKI regulation [SKI FS 2002:1] highlight the potential use of performance measures other than dose, such as fluxes and environmental concentrations, to illustrate barrier safety functions in assessments covering long periods. The commentaries on SSI's regulations [SSI FS 1998:1] states that all intermediate results leading up to dose and risk estimates should be reported. The issues associated with this approach are discussed in Section 2.6.3.

2.6.2 Calculation of risk

In some of the regulations and regulatory guidance reviewed in this report, and in some of the presentations of the results of risk calculations, the terms *mean risk* or *expectation value of risk* are used. This implies that there is a distribution of risk results from which a single measure is derived. This leads to an apparent anomaly: if risk takes into account all of the uncertainties, why do risk calculations yield a distribution and not a single value?

This apparent anomaly is, in part, the result of using probabilistic techniques to calculate risk and in part the result of the end-point being used. The following discussion addresses these and related issues concerning the calculation, presentation and assessment of risk, including:

- Conditional risks.
- Role of the dose - risk factor.
- Time-dependency in risk calculations.
- Maximum risk.
- Skewness of risk distributions.

Conditional risks

The definition of risk is probability multiplied by consequence. The way in which consequence is measured differs from hazard to hazard, but if the hazard is a unique event, then the consequence is one and the risk is simply the probability of the event:

- For example, the risk of flooding (water exceeding a particular level) in an area is equivalent to the probability of flooding in that area.

However, if the consequence can take different values, then there will be a probability associated with each value, and hence there will be a distribution of risk values.

- For example, the number of deaths from a toxic gas release may depend upon the wind direction at the time of release. A calculation of risk based on the probability of the release and the number of deaths if the wind is, say, in the North, would then be different to the risk if the wind was in the South.

The risks calculated for each wind direction in this example are conditional risks, and each calculation can be regarded as being the risk associated with a different scenario. Combining these conditional risks into an overall measure of risk requires information about the relative probabilities of the different scenarios. If each scenario is equally likely, then the overall risk is simply the average of the conditional risks. However, in this example, because there is generally a prevailing wind direction, the scenario probabilities will not be equal, and the conditional risks must be probability-weighted to determine the mean risk.

In the case of a performance assessment using deterministic calculations, conditional risks can be calculated for each scenario. The probability used to calculate conditional risk is the probability that an individual dies or develops cancer as the result of receiving the calculated dose (i.e., the dose - risk factor). If probabilities can also be assigned to each scenario (i.e., the scenarios are exclusive and exhaustive, and the probabilities can be quantified), then a mean or expectation value of risk[3] can be calculated from these conditional risks. If there is only a single scenario identified (normal or expected evolution), then the scenario probability is one and the overall risk is equivalent to the conditional risk.

In the case of probabilistic calculations, however, conditional risks are calculated not only for each scenario, but also for each sampled combination of parameter values. In the case of a single, normal evolution scenario, for example, if the probabilistic calculations use 500 simulations (i.e., 500 sets of parameter values), then the result will be a set of 500 conditional risks. Because each set of parameter values has the same probability or degree-of-belief, the mean or expectation value of risk is simply the average of this set of conditional risks without any probability weighting. If more than one scenario is assessed, then the overall risk would be calculated from the expectation values for each scenario in the same way as for deterministic calculations.

Role of the dose - risk factor

Because the dose-risk conversion factor is a constant value used in all of the calculations, the same overall result will be obtained if an expectation value of dose is calculated from a set of conditional doses, and the dose-risk factor applied to this. The concept of a distribution of doses reflecting the existence of uncertainties may be

[3] The expectation value is in general equivalent to the arithmetic mean. The formal definition of the expectation value differs depending on whether it is applied to a discrete or continuous distribution (see Wilmot 2001).

more readily understood than the equivalent distribution of risks, because risk may be assumed to be a measure of uncertainty and not to have an associated uncertainty itself.

Although the dose-risk factor is a constant and can be regarded simply as a multiplier at the end of the analysis, it is important that assessment calculations are carried through to risk and not only expressed as dose. This is because, in regulatory regimes that define a risk limit or target, it is generally the notion of what is an acceptable (or tolerable) risk that determines the limit, and not what is the risk equivalent of an acceptable dose. This may also be the case where the regulations specify a dose limit. For example, in the US, the EPA regulations for Yucca Mountain (40 CFR 197) include a dose standard, but in the commentary the EPA states that:

> "The dose standard is risk-based, in that the EPA judged that a risk of about 8.5 fatal cancers per million members of the population per year is acceptable. The dose standard is derived from this risk using a dose-risk conversion factor of 0.075 per Sv."

One reason for using the risk limit or target is that such a comparison reduces the potential for confusion over what is regarded as a "safe" dose. In the UK regulatory guidance, for example, the limit for the annual individual dose to a member of the critical group during the operational phase of a facility (source-related dose constraint) is set at 0.3 mSv/y. In the post-operational phase, this guidance establishes a risk target (annual individual risk to a member of a potentially exposed group) of 10^{-6}. Using the ICRP recommended value of 0.06 per Sv[4], this risk is equivalent to an annual dose of 0.017 mSv. The reason for this difference in dose is that there are greater levels of uncertainty regarding the receipt of doses by individuals in the post-operational phase. The regulators believe that these uncertainties should be taken into account in setting a risk target, rather than that they have set a more rigorous standard of safety for the future.

Time-dependency in risk calculations

The simple examples of risk calculations used above do not involve a consideration of system evolution. Clearly, however, the way in which the risks from a disposal facility vary with time is a key element of a performance assessment and regulatory decision-making. In a deterministic calculation, the normal way of displaying time-varying performance is as a curve showing how dose varies with time. On the assumption that the dose-risk conversion factor is constant, simply changing the numbers on the axis will convert this to a plot of risk against time.

In the case of a probabilistic calculation of system evolution, each of the conditional risk calculations arising from a separate sample of parameter values can be plotted as a risk - time curve. This assemblage of curves can be summarised as a single, mean risk against time curve. The mean risk is determined by considering a virtual "cross-section" through the assemblage of conditional risk curves at a particular time and

[4] The Swedish regulation specifies a factor of 0.073 per Sv, giving a dose equivalent to the 10^{-6} risk limit of 0.014 mSv/yr. The reason for the difference appears to be that the Swedish regulation is based on the risk of cancer or serious hereditary defect, whereas the UK guidance is based on the risk of fatal cancer or hereditary defect.

calculating the expectation value of the resulting distribution. Repeating this procedure at a number of times yields a mean risk curve. The set of parameter values that correspond to the mean risk at one particular time will not necessarily be the same set of values that gives the mean risk at any other time. In other words, the mean risk – time curve does not represent in any sense a typical evolution of the system, but is a mathematical construct that allows an easy assessment of system behaviour against the regulatory criterion.

Maximum risk

The description above is based on the assumption that the mean or expectation value of risk is an appropriate value for comparison with a regulatory criterion. There are some circumstances in which this might not be the case.

Each set of parameter values used in a probabilistic calculation represents a *possible* evolution of the disposal system. One set of parameter values may represent the *actual* evolution, but the premise of the probabilistic approach is that there is no means of knowing which set this is. If there is an uncertain event that gives rise to a high dose in some possible futures, then the expectation value is an appropriate measure because it takes account of that uncertainty. However, if the high dose event takes place in all or most possible futures, but at a different time in each future, the expectation value is not appropriate. Under these circumstances, at each time the expectation value is calculated, the dose from the future that has the event at that time will be masked by the doses from the remaining futures. This is a form of risk dilution, which can also occur through an inappropriate broadening of parameter uncertainty ranges.

For example, consider a case where one canister fails and gives a short-lived peak dose of 1 mSv / yr. If this failure occurs at a random time, then in an assessment covering 100,000 years, if doses are calculated at say 100-year intervals and with 1,000 simulations, then each simulation will show the peak dose at a different time-step. If the doses at other times are negligible, then the mean dose at each time-step becomes 1 • Sv / yr, effectively giving a thousand-fold risk dilution. Such an extreme example is unlikely to occur in any real assessment, but it demonstrates the need for the results from all simulations to be illustrated.

The problems associated with use of mean risk in the specific example above could be overcome by use of the maximum conditional risk at each time for comparison with the regulatory criterion. Using the above example of canister failure, risk calculated on this basis would be high throughout the whole of the 100,000 year assessment period. This in itself would not exaggerate system performance, but would misrepresent performance by suggesting releases over a much longer time-scale than is indicated by the underlying assumptions.

The maximum conditional risk at any time may simply represent a particular set of parameter values rather than the occurrence of a discrete event. In such a case, there is no basis for specifying this particular set rather than any other set for comparison with regulatory criteria. Such a requirement would be akin to requiring worst-case or arbitrarily conservative assumptions or parameters in a deterministic assessment and is contrary to adopting a realistic approach to assessment.

The problem illustrated by the example of a high dose event in every future is not really a basis for determining regulatory criteria or guidance. It does demonstrate that some uncertainties, particularly those concerning the timing of events, may be more appropriately assessed using a series of deterministic calculations. It also shows the benefit of requiring that all conditional risk calculations are presented so that the full range of possible system behaviour can be examined. Finally, it reinforces the fact that numerical compliance with a criterion should not be the sole basis for determining acceptability of a design or facility.

Skewness of risk distributions

In any probabilistic risk calculation, some simulations may exceed the regulatory criterion even though the expectation value of risk is less than the criterion. In fact, if the distribution of conditional risks at any particular time was uniform about the mean (e.g., a normal distribution), then virtually half of the conditional risks could exceed the criterion. In practise, however, the distributions are typically highly skewed towards the high consequence end, so that the mean of the distribution may be close to the 95th or even the 99th percentile. In these circumstances, few if any conditional risks will exceed the criterion if the expectation value is compliant, but conversely, the expectation value can be criticised as a descriptor of this type of distribution (see Wilmot 2001).

Supplementary criteria that address the issues of highly skewed distributions and conditional risks exceeding the criterion are useful, and regulations in the USA provide such criteria. In essence, these additional criteria limit the proportion of conditional risk calculations that may exceed the criterion. There is a strong case for requiring all conditional risk results to be shown as part of the presentation of assessment results, because this will allow the full range of possible system behaviours to be examined. Such a presentation could also show that some conditional risk calculations exceed the regulatory criterion and prompt questions as to the meaning and significance of these results. Supplementary criteria and guidance that acknowledge that such a situation is a normal and expected aspect of accounting for uncertainty will be valuable in responding to these questions.

Summary

The alternatives to supplementary criteria concerning conditional risks that exceed the criterion are either an expectation that all conditional risks will lie below the regulatory criterion, or a requirement for regulatory judgement as to the acceptability of a particular set of assessment calculations. The former, as noted above, is equivalent to the adoption of worst-case or arbitrarily conservative assumptions or parameters in a deterministic assessment and is contrary to adopting a realistic approach to assessment. Regulatory judgements will need to be made throughout the examination of a safety case and performance assessment, but reasonable supplementary criteria or guidance that can be established as a basis for these judgements could be of benefit.

2.6.3 Different performance measures

One of the arguments put forward for restricting the time-scale considered in assessments of dose or risk is the increasing level of uncertainty in the behaviour of the disposal system over long time-scales. In particular, large uncertainties in the state of the biosphere in the distant future are transferred directly to large uncertainties in the calculated dose or risk. An alternative to simply curtailing the period considered in assessments, is to use dose or risk as measures of health effects for the period when uncertainties are reasonably well characterised, and to use indicators that do not involve the biosphere to provide a measure of the safety of the disposal system beyond this period. Such alternative indicators are also subject to uncertainty, but to a lesser extent than measures of dose or risk. To be useful in the context of regulatory decision-making, however, there must be a link between such alternative indicators and the measures of health effects.

The International Atomic Energy Agency's Sub-Group on Principles and Criteria for Radioactive Waste Disposal has considered the role of alternative safety indicators (IAEA 1994). The hierarchy of environmental safety indicators considered include:

- Concentrations in surface and near-surface environment
- Flux from geosphere
- Waste package containment time
- Transfer time through geosphere or other barriers
- Engineered barrier system fluxes

Each of these alternatives has different advantages and disadvantages. The principal advantages are independence from changes in the biosphere. The IAEA Sub-Group also argues that they may be easier to communicate. The principal disadvantages are that they may not be directly related to safety, that reference levels may be difficult to define, and that there are no direct comparators in the case of artificial radionuclides. Two alternative performance measures that are used in regulations governing the disposal of radioactive waste are environmental concentrations and radionuclide fluxes. Radiotoxicity has also been proposed as a possible performance measure. Issues concerning the use of these are discussed below.

Environmental concentrations

There are two approaches to defining appropriate environmental concentrations for use as alternative performance measures. The first of these is by reference to concentrations of naturally-occurring radionuclides. A release of radionuclides from a disposal facility will increase concentrations in the environment, and thereby increase exposures to radiation. The performance measure must therefore be a measure of the acceptable increase in concentration, rather than an absolute concentration. For example, it may be acceptable to increase the concentration of uranium isotopes in the soil by 50% without increasing background radiation to unacceptable levels. However, this is not a universal measure, because environmental concentrations may be low in particular geological environments and already unacceptably high elsewhere. Relative increases also cannot be defined for radionuclides that are not present naturally, including both transuranic elements and artificial isotopes of other elements. Comparisons for these radionuclides could be based on radiotoxicity or by comparison to natural radionuclides (e.g., comparing the total concentration of alpha emitters with that of natural alpha emitters).

An alternative approach to determining environmental concentration criteria is to use a dose assessment model to calculate doses from unit releases of radionuclides from the near-field. By assuming a linear relationship between release and dose, or by re-calculating with higher releases, the release rate that gives a “just acceptable” dose can be determined, together with the corresponding concentrations in intermediate regions of the modelled domain (e.g., in the surface and near-surface environments).

This method of linking radionuclide concentrations and health effects requires a set of assumptions about the behaviour of an exposed group and other biosphere characteristics. Because it is the best understood and characterised, it is likely that the present-day biosphere would be used as the basis for these assumptions. This is in spite of the fact that the aim of using environmental concentrations is to obviate the large uncertainties regarding the state of the biosphere in the far future.

Furthermore, if there is a capability of determining dose in order to “back calculate” the equivalent radionuclide concentrations, then there is no necessary calculational reason for use of an alternative performance measure. In this case, the driver for an alternative measure is the extent to which it can be more readily communicated than dose or risk. This may be a very important reason if it is considered that the increasing level of uncertainty in the biosphere will reduce confidence in the overall assessment of the disposal system.

An example of the use of environmental concentrations as performance measures are the regulations in the US promulgated by the EPA for the WIPP site and for Yucca Mountain. Both these regulations include groundwater protection criteria, which set limits for environmental concentration of radionuclides in underground sources of drinking water. In the case of Yucca Mountain, the aquifer in the region could supply a larger population than currently use it. Instead of conjectural assumptions involving increases in the local population or transport of the water to centres of population, the EPA considers it preferable in terms of public acceptability to impose concentration limits that protect the groundwater at source. However, these limits are in addition to the individual dose criterion, rather than being a substitute for them.

Radionuclide fluxes

Environmental concentrations may provide an alternative performance measure because it is these concentrations that govern the extent of exposure to radiation and hence dose and risk. Similarly, environmental concentration are themselves governed by the fluxes of radionuclides entering and moving within the environment. Concentrations in the biosphere, for example, are controlled by fluxes at the geosphere-biosphere interface and in various parts of the biosphere.

It is more difficult to establish independent criteria for radionuclide fluxes. Natural radionuclide fluxes in the biosphere arise through processes such as erosion, dissolution in water, water flow, sedimentation, resuspension and transport by wind, which lead to a continuous flow of natural radionuclides from land, through rivers and lakes, and into the sea. This flux leads to exposures to humans and other biota that are, at least largely, acceptable. A comparison of radionuclide fluxes from a disposal facility with this natural flux is a potential performance measure because these fluxes are independent of assumptions regarding factors such as the environment, pathways and receptors. Changes in these factors would be the same for both the waste-related and natural fluxes. However, determining the natural flux, through either modelling or measurement, is subject to large uncertainties, and hence the basis for comparison is uncertain. As with environmental concentrations, there are also problems in determining what fractional increase in the natural flux would be acceptable, and in how to deal with radionuclides for which there is no natural flux.

Another radionuclide flux that has potential use as a performance measure is the release of radionuclides from the near-field. However, this flux is a better measure of repository performance than of health effects, because it neglects any barrier function (e.g., sorption, matrix diffusion) of the far-field and biosphere. Regulatory criteria for environmental concentrations and biosphere fluxes will be, at least to some extent, generic but can be based on assumptions about the regional biosphere. Criteria for fluxes from the near-field, unless established after site characterisation and modelling studies, will not be able to account for site-specific barriers and so any assumptions about these barriers will be conjectural.

Although not included in the regulatory guidance reviewed in this report, because the corresponding regulations do not include a risk criterion, recent Finnish regulatory guidance provides an example of the use of radionuclide fluxes as constraints (see Box 1). These constraints effectively apply to activity releases from both the expected evolution scenario and unlikely disruptive event scenarios (including human intrusion). In the case of unlikely disruptive events, it is the expectation value of the activity release (i.e., the release multiplied by the probability of the event) that is to be compared to the constraints. It is also the expectation value that is used for comparison in the case of the expected evolution scenario, but the probability of this scenario is assumed to be one.

These activity release constraints apply to releases that occur after several thousand years and are not applicable to potential releases in the short-term. The Finnish regulatory guidance also limits the temporal averaging to a maximum of 1,000 years. For the expected evolution scenario, however, releases are expected to be gradual rather than pulse events and the purpose of temporal averaging in this case is unclear. Finally, the guidance requires that the sum of the ratios between nuclide-specific activity releases and the respective constraints shall be less than one. This means that releases for particular radionuclides can be higher if there are no other releases, but can never exceed the constraints. The regulatory guidance does not detail the basis for the activity release constraints or how they have been derived.

Box 1 Activity Release Constraints defined by the Finnish Radiation and Nuclear Safety Authority (STUK)

The nuclide-specific constraints for the activity releases to the environment are as follows:

- 0.03 GBq/a for the long-lived, alpha emitting radium, thorium, protactinium, plutonium, americium and curium isotopes
- 0.1 GBq/a for the nuclides Se-79, I-129 and Np-237
- 0.3 GBq/a for the nuclides C-14, Cl-36 and Cs-135 and for the long-lived uranium isotopes
- 1 GBq/a for Nb-94 and Sn-126
- 3 GBq/a for the nuclide Tc-99
- 10 GBq/a for the nuclide Zr-93
- 30 GBq/a for the nuclide Ni-59
- 100 GBq/a for the nuclides Pd-107 and Sm-151.

(STUK 2001)

A variant on the radionuclide flux as a performance measure is the cumulative release criterion set in the regulations for the WIPP site in the US. Because it is cumulative release rather than flux, a period must be specified - in this case it is 10,000 years. This limit also reduces the need for conjecture about system evolution. The boundary for determining cumulative release is a region termed the "controlled area". This represents a concept not used in other regulations - it acknowledges that there will be releases from the facility itself, but that these will not pose a health risk if there is limited migration in the geosphere. The surface expression of the controlled area is under Federal ownership and institutional controls will limit the extent of development in this region. The cumulative release criterion applies to a range of radionuclides and the specific limit for a particular radionuclide depends on the other nuclides released. The limit is also a function of the total inventory, so that effectively it is a measure of barrier efficiency rather than environmental concentration or dose.

Radiotoxicity

A further performance measure included in the IAEA review, and also included by the UK Environment Agencies as a potential performance measure, is radiotoxicity.

There are several indicators of radiotoxicity, including total activity, specific activity and the number of annual intake limits (ALIs) contained in the waste. These indicators may be expressed in a dimensionless form by comparison with a reference material such as the activity in the mass of uranium ore required to produce the disposed waste or the activity of the uranium destroyed by fission to produce the waste.

These radiotoxicity measures indicate when radioactive decay will have reduced the waste inventory to such an extent that it no longer poses a significant hazard. Unlike the other performance measures, they do not account for the distribution of radionuclides within the disposal system. Thus even when the radiotoxicity is above an acceptable level, the disposal system may still be safe if the inventory is contained within the near-field (or conversely widely dispersed and dilute). The principal use of radiotoxicity is, therefore, as an indicator of when further assessment of disposal system performance is no longer required.

Summary

The discussion of alternative performance measures above indicates that the principal reason for adopting any measures other than health risks or dose would be to increase confidence in assessment results by reducing conjecture. Uncertainties are not reduced through the use of these alternatives, although the role of the uncertainties in the regulation - assessment cycle does change. Use of alternative performance measures shifts some of the onus of assessment from the proponent to the regulator. A risk criterion is a generic criterion, which is not specific to radioactive waste disposal, but can be based on the level of risk generally found acceptable or tolerable within society. Corresponding dose limits can be determined directly from the acceptable level of risk and do not require any level of disposal system assessment. The onus for demonstrating that a disposal facility will perform within these criteria lies with the proponent. However, the alternative performance measures are, to a greater or lesser extent, site- or concept-specific. Setting of these criteria cannot be done without making some assumptions about the biosphere or the behaviour of the disposal system barriers. These assumptions, and the corresponding assessments to derive the criteria, must be made by the regulator, and the onus of making and *justifying* these assumptions no longer lies with the proponent.

3 Conclusions and Suggestions for Further Work

3.1 Introduction

This report has reviewed a range of regulations, regulatory guidance and assessments, all of which require or utilise probabilistic calculations to determine risk or other performance measures. This review has identified several areas where the different regulatory regimes have similar requirements, and has also noted areas where there are differences or additional requirements. On the basis of these comparisons, several topics have been identified where guidance to the proponent on the regulator's expectations would be of value. Such guidance would not be prescriptive, but if followed it would facilitate review and regulatory decision-making.

The following sections summarise the principal conclusions of the review and make suggestions for topics that we consider could be included in additional guidance from SKI and SSI. Some of these topics will require additional studies by the regulators to establish, for example, supplementary quantitative criteria. In other cases, there is sufficient information available now to allow guidance to be drafted.

3.2 Suggestions for Further Work

3.2.1 Models and codes

The Swedish regulators have access to a range of models and codes for developing system understanding of a disposal system and for assessing overall performance. Few of these, however, were designed and implemented with the specific requirements of a risk assessment in mind.

As noted in Section 2.5, the flexibility and computational requirements of probabilistic assessment codes are greater than those for codes used solely for deterministic calculations. Imposing these requirements on existing codes would lead, at the least, to a significant development effort. For some codes, suitable development may not be possible. This applies to commercial codes without open source-code, and to codes implementing complex models that cannot be appropriately generalised.

Overall, decisions on future developments of the regulators' modelling capability require as much a strategic evaluation of the types and extent of calculations that the regulators will undertake as a technical evaluation of model and code capabilities. For example, because it remains the proponent's responsibility to conduct and present assessment calculations for comparison with regulatory criteria, the regulators may not have a requirement for an integrated system performance code capable of full-scale assessment calculations. However, developing an understanding of the risk assessment process, and assessing the effects of issues such as model simplification and different approaches to the treatment of uncertainty, will likely require some development of the existing capability.

A strategic assessment of regulatory needs is outside the scope of this report. The regulators have started the process through recent workshops and may continue the process through the appointment of Clearing Houses to review key aspects of the assessment system. This review, and the recent workshops, have, however, identified a few key areas where some model and code development may be needed:

- **Sub-system interfaces**. Risk assessments require an integrated system analysis, rather than separate studies to develop sub-system understanding. This means that behaviour at the sub-system interfaces (i.e., the near-field / far-field boundary, and the geosphere – biosphere interface) become important elements of the assessment. In addition to an appropriate representation of the processes at the interface, there must be a coherent linkage of model domains and time-scales.

- **Biosphere - other receptors**. Recent Swedish regulations have established an individual risk criterion for the protection of human health. Standards for the protection of other species could be established through supplementary criteria or guidance. Quantitative criteria may require biosphere modelling and the development of additional modelling capabilities.

- **Biosphere - alternative end-points**. Doses and risks arising from releases at the present-day are assessed through the identification of the critical group – the group of individuals with similar habits and lifestyles that is most exposed. For releases taking place in the future, assessments must consider a wider range of individuals or members of potentially exposed groups. In the far future, the assumptions required to specify these potentially exposed groups become increasingly conjectural. An alternative to calculating risk is to consider environmental end-points that do not require definition of a potentially exposed group. However, for regulatory consistency, any such environmental end-points must be clearly related to the individual protection criterion. Setting alternative criteria is therefore likely to require specific biosphere modelling using codes that are compatible with both options. International programmes such as the EC project FASSET could provide useful inputs to such studies.

We suggest that SKI and SSI jointly establish an outline protocol for calculations in support of regulatory review. This protocol could form the basis of a requirements specification for code development. We also suggest that SKI and SSI initiate Clearing Houses or similar groups of experts, so that a review of existing code capabilities against this requirements specification can be conducted across as wide a range of codes as possible. This review should include topics such as availability, documentation, quality assurance (QA) pedigree, usability and development potential as well as technical capabilities.

3.2.2 Scenarios

The recent Swedish regulations already address one of the key issues regarding the identification and treatment of scenarios in performance assessments. Section 2.6 of the supporting commentary (SSI 1999) notes that the individual protection criterion applies to the normal or expected evolution of the disposal system, and that separate analyses will be required of the system's performance following human intrusion. This means that one source of aleatory uncertainty, which would otherwise involve arbitrary and conjectural assumptions, can be omitted from calculations of risk.

Although omitted from calculations of the risk associated with undisturbed performance, the issue of human intrusion will need to be addressed in a safety case, and the SSI regulations require the proponent to report the consequences of intrusion events. The protective capability of the repository after intrusion must also be reported. We suggest that SKI and SSI give early consideration to developing additional guidance on their expectations about what intrusion events should be considered, how the likelihood and timing of these events should be defined, and how they will judge the acceptability or otherwise of a design or proposal with respect to human intrusion.

The Swedish regulations do not exclude other uncertain disruptive events from consideration in performance assessments. There is currently no guidance on how this type of event should be assessed. On the assumption that a probabilistic methodology will be used for calculations of risk, one approach would be to include uncertain events in a single, all-encompassing scenario, but to sample the occurrence and timing of these events from pdfs. This would account appropriately for the uncertainty of event occurrence but, if they are only included in a few simulations, associated uncertainties may not be fully explored. An alternative approach would be to develop separate scenarios for any uncertain, disruptive events. This would allow further exploration of the uncertainties, but might focus attention unduly on relatively uncertain events.

The regulators will need to determine both how they will explore scenario uncertainty and whether they will provide guidance on their expectations for a safety case. In the previous section, there is a presumption against the development of an independent probabilistic assessment code for regulatory analysis. This suggests that, if the effects of disruptive events are to be explored, some stand-alone analyses will be required. If regulatory analyses show that particular disruptive events may be significant, then the regulators will need to assess the proponent's treatment of the event in detail. An adequate level of review may only be possible if the proponent has undertaken detailed analyses rather than incorporating disruptive events into a single system model.

At the current stage of SKB's disposal programme, the key issues and potentially significant disruptive events cannot necessarily be identified. The argument presented above indicates, however, that regulatory review will be easier if uncertain, disruptive events are treated as separate scenarios.

A possible way forward would be for SKI and SSI to consider providing additional guidance on the treatment of scenario uncertainty. We suggest that the regulators reinforce their expectation that a safety case will include a probabilistic assessment of risk, clarify their expectation that the principal focus of such an assessment should be the normal or expected evolution of the disposal system, and indicate that uncertain, disruptive events should be the subject of separate analyses. Where quantification of scenario uncertainties can be achieved, the results of these separate analyses can be integrated with the results of the normal evolution scenario to give an overall estimate of risk.

3.2.3 Parameters

SKI and SSI have commissioned work on the use of judgements in performance assessments, including the way in which expert elicitation can be used to derive distributions for parameter values that cannot be readily obtained in any other way. SKI and SSI are also proposing to continue this work by conducting an elicitation using appropriately selected experts. These studies and exercises will provide the regulators with direct experience of the benefits and disadvantages of the elicitation process, and enable an informed review of future elicitations undertaken by SKB. They will also, for a few parameters, provide elicited distributions that can be used in sensitivity studies.

We consider that SKI and SSI should continue with their studies of judgements and elicitations, but that the number of actual elicitations undertaken is limited. Elicitation is a time-consuming and expensive process, and there would be little benefit to the regulators of pre-empting studies that should be undertaken by the proponent. Because of the time and cost involved, there could be programmatic implications if SKB conducted elicitations in a manner that did not satisfy the regulators. We therefore suggest that SKI and SSI consider publishing supplementary guidance containing the experience gained from their studies in the form of requirements for elicitation.

In order to conduct an informed review of a probabilistic assessment, SKI and SSI will need to understand the derivation of pdfs from poorly characterised data, and the extent to which different assumptions and perspectives can affect the derived distributions. We therefore suggest that SKI and SSI use raw site characterisation data from SKB to derive a number of parameter pdfs. A comparison of these with the equivalent pdfs derived by SKB will help to develop an understanding of the judgement process and the importance of different assumptions. Documentation of the parameter derivation process will also indicate the level and detail of justification that the regulators consider appropriate.

We suggest that SKI and SSI develop supplementary guidance which reinforces their expectation that SKB adopts an iterative approach to performance assessment, and that interim evaluations of the disposal system will be made available for review. This guidance should stress that the role of early iterations should be the identification of key uncertainties rather than the development of finalised parameter distributions.

3.2.4 Risk criteria

The recent regulations established in Sweden for disposal of spent fuel include a risk criterion, but there is no further guidance on how risk should be calculated, other than the specification that risks arising from human intrusion should be assessed separately. The review in this report has shown that there are some additional criteria or guidance that will help to ensure that calculations of risk undertaken as part of a performance assessment will be an appropriate basis for regulatory decision-making.

A way forward would be for SKI and SSI to develop supplementary guidance to clarify that the comparison with the regulatory criterion will be done on the basis of the expectation value of the risk, calculated as a function of time. In order to ensure that the expectation value is properly representative of the calculated risks, and that neither isolated high risks nor risk dilution unduly influence this value, assessment results must include all the calculated conditional risk distributions, and there must be a robust demonstration that the results are statistically converged.

3.2.5 Alternative performance measures

The use of alternative performance measures may serve to increase confidence in assessment results by reducing conjecture. However, the adoption of alternative performance measures as criteria will require a significant regulatory effort both in formulating the basis for deriving the performance measures and in developing quantitative criteria.

The use of alternative performance measures as criteria will shift some of the onus of assessment from the proponent to the regulator. Retaining confidence in the independence and competence of the regulators will be critical to the overall success of the regulatory process. We therefore suggest that SKI and SSI consult as widely as possible on the benefits to be gained in terms of confidence in assessment results through use of alternative measures, and also on how the involvement of the regulators in the assessment process will be viewed by stakeholders.

Notwithstanding the results of any consultation, we also believe that SKI and SSI should further examine the extent of assumptions that would be required in order to derive robust alternative performance measures. This study should also consider whether the data required to support these assumptions and to derive quantitative criteria are generic or site-specific, whether SKB's site characterisation plans will address any site-specific data needs, and whether any future site characterisation data or performance assessment results could invalidate generic criteria.

As part of the review of models and codes suggested in Section 3.2.1, we suggest that SKI and SSI assess the model requirements for developing alternative performance measures.

3.3 Summary

The key elements of further work by SKI and SSI that we consider would prepare the regulators for a review of a probabilistic risk assessment are:

- Develop an outline protocol for calculations in support of regulatory review.
- Review model and code requirements and existing capabilities.
- Provide regulatory guidance on their expectations for SKB's performance assessment, including:
    - The need for iterative assessments showing an increase in system understanding and a decrease in uncertainties.
    - Use of probabilistic techniques to calculate risk.
    - Presentation of all conditional risk calculations.
    - Use of the expectation value for comparison with the regulatory criterion.

- A demonstration of convergence.

- Consult on the benefits and disadvantages associated with the adoption of criteria based on alternative performance measures.

- Develop a methodology for formulating and quantifying alternative performance measures.

4 References

Ababou, R, Sagar, B and Wittmeyer, G. 1992. Testing procedures for spatially distributed flow models. Advances in Water Resources, **15**, 181-198.

AECB (Atomic Energy Control Board). 1985. Regulatory policy statement. Deep geological disposal of nuclear fuel waste: Background information and regulatory requirements regarding the concept assessment phase. Atomic Energy Control Board Regulatory Document R-71.

AECB (Atomic Energy Control Board). 1987. Regulatory policy statement. Regulatory objectives, requirements, and guidelines for the disposal of radioactive wastes - long-term aspects. Atomic Energy Control Board Regulatory Document R-104.

DOE (U.S. Department of Energy). 1990. WIPP Test Phase Plan: Performance Assessment. DOE/WIPP-89-011, Rev. 0. U.S. Department of Energy, Waste Isolation Pilot Plant, Carlsbad, NM.

DOE (U.S. Department of Energy). 1996a. Title 40 CFR Part 191 Compliance Certification Application for the Waste Isolation Pilot Plant, October 1996. DOE/CAO-96-2184.

DOE (U.S. Department of Energy). 1996b. Title 40 CFR Part 191 Compliance Certification Application for the Waste Isolation Pilot Plant: Appendix Codelink; The WIPP 1996 performance assessment codes and their linkages. DOE/CAO-96-2184.

DOE (U.S. Department of Energy). 1998. Viability assessment of a repository at Yucca Mountain, DOE/RW-0508, December 1998.

Environment Agency, Scottish Environmental Protection Agency, and Department of the Environment for Northern Ireland. 1997. Radioactive Substances Act 1993 - Disposal Facilities on Land for Low and Intermediate Level Radioactive Wastes: Guidance on Requirements for Authorisation. Environment Agency, Bristol.

EPA (U.S. Environmental Protection Agency). 1993. 40 CFR Part 191 Environmental Radiation Protection Standards for the Management and Disposal of Spent Nuclear Fuel, High-Level and Transuranic Radioactive Wastes; Final Rule. Federal Register, Vol. 58, No. 242, pp. 66398 - 66416, December 20, 1993. Office of Radiation and Indoor Air, Washington, D.C.

EPA (U.S. Environmental Protection Agency). 1996. 40 CFR Part 194: Criteria for the Certification and Re-Certification of the Waste Isolation Pilot Plant's Compliance with the 40 CFR Part 191 Disposal Regulations; Final Rule. Federal Register, Vol. 61, No. 28, pp. 5224 - 5245, February 9, 1996. Office of Air and Radiation, Washington, D.C.

EPA (U.S. Environmental Protection Agency). 2001. 40 CFR Part 197: Public Health and Environmental Radiation Protection Standards for Yucca Mountain, NV; Final Rule. Federal Register, Vol. 66, No. 114, pp. 32073-32135, June 13, 2001. Office of Air and Radiation, Washington, D.C.

Goodwin, B.W., McConnell, D.B., Andres, T.H., Hajas, W.C., LeNeveu, D.m., Melnyk, T.W., Sherman, G.R., Stephens, M.E., Szekely, J.G., Bera, P.C., Cosgrove, C.M., Dougan, K.D., Keeling, S.B., Kitson, C.I., Kummen, B.C., Oliver, S.E., Witzke, K., Wojciechowski, L. and Wikjord, A.G. 1994. The disposal of Canada's nuclear fuel waste: Postclosure assessment of a reference system. AECL-10717, COG-93-7. AECL, Pinawa, Manitoba.

Hodgkinson, D. P. 1992. An overview of fracture flow modelling at Stripa. Proceedings of the Final Stripa Symposium, Stockholm, 1992.

IAEA (International Atomic Energy Agency). 1994. Safety Indicators in Different Time Frames for the Safety Assessment of Underground Radioactive Waste Repositories, First report of the INWAC Subgroup on Principles and Criteria for Radioactive Waste Disposal, IAEA-TECDOC-767, IAEA, Vienna.

Mackay, R. 1993. A study of the extent of site investigation upon the estimation of radiological performance: Overview. UK Department of the Environment Report DoE/RR/93.053.

Norris, S., Bailey, L.E.F., Askarieh, M.M., Hickford, G.E. 1997. Nirex 97: An Assessment of the Post-closure Performance of a Deep Waste Repository at Sellafield. Overview. UK Nirex Report S/97/012. Nirex, Didcot.

NRC (U.S. Nuclear Regulatory Commission). 2001. 10 CFR Parts 2, 19, 20, 21, etc.: Disposal of High-Level Radioactive Wastes in a Proposed Geologic Repository at Yucca Mountain, Nevada; Final Rule. Federal Register, Vol. 66, No. 213, pp. 55731-55816, November 2, 2001.

OECD/NEA. 1991. Disposal of radioactive waste: Review of safety assessment methods. Report of the Performance Assessment Advisory Group of the Radioactive Waste Management Committee, OECD Nuclear Energy Agency. OECD/NEA, Paris.

OECD/NEA. 1992. Systematic Approaches to Scenario Development. OECD Nuclear Energy Agency, Paris.

SKB (Swedish Nuclear Fuel and Waste Management Company). 1999. SR 97 Main Report Volumes I and II. SKB Technical Report TR-99-06. SKB, Stockholm.

SSI (Swedish Radiation Protection Institute). 1999. Health, Environment and Nuclear Waste: SSI Regulations and Comment. SSI Report 99:02. SSI, Stockholm.

SKI (Swedish Nuclear Power Inspectorate). 1997. SKI SITE-94 Deep Repository Performance Project: Summary. SKI Technical Report 97:5. SKI, Stockholm.

Stenhouse, M.J., Miller, W.M, and Chapman, N.A. 2001. System studies in PA: Development of process influence diagram (PID) for SFR-1 repository Near-field + Farfield. SKI Technical Report 01:30. SKI, Stockholm.

STUK (Radiation and Nuclear Safety Authority). 2001. Long-term safety of disposal of spent nuclear fuel. Guide YVL 8.4. Helsinki.

Sumerling, T.J. (Editor). 1992. Dry Run 3 - A trial assessment of underground disposal of radioactive wastes based on probabilistic risk analysis - Volume 10: Overview. UKDOE/ HMIP Report DoE/HMIP/RR/92.039.

Wingefors, S., Andersson, J., Norrby, S., Eisenberg, N. A., Lee, M. P., Federline, M. V., Sagar, B., and Wittmeyer, G. W. 1999. Regulatory Perspectives on model validation in high-level radioactive waste management programs: A joint NRC/SKI White Paper. SKI Technical Report 99:02. SKI, Stockholm.

Wilmot, R.D. 2001. Presenting numerical risk estimates. Environment Agency Technical Report P3-037/TR. Environment Agency, Bristol.

Wilmot, R.D. and Crawford, M.B. 2000. A review of expert judgement and treatment of probability in SR 97. In: Opinions on SKB's Safety Assessments SR 97 and SFL 3-5: A review by SKI consultants. SKI Technical Report 00:04. SKI, Stockholm.

Wilmot R.D., and Galson D.A. 1994. Validation: what should a regulator look for in a safety case? In *Proceedings of the Second International Conference on Probabilistic Safety Assessment and Management - PSAM II* (San Diego, CA, 20-24 March 1994), Elsevier, NY.

Wilmot, R.D. and Galson, D.A. 2000. Expert Judgement in Performance Assessment SKI Technical Report 00:47. SKI, Stockholm.

Wilmot, R.D. Galson, D.A. and Hora, S.C. 2000. Expert Judgements in Performance Assessments. Report of an SKI/SSI Seminar. SKI Technical Report 00:35. SKI, Stockholm

Zimmerman, D. A. and Gallegos, D. P. 1993. A comparison of geostatistically-based inverse techniques for use in performance assessment analyses at the WIPP site - Results from Test Case No. 1. Informal paper presented at the NEA PSAG Meeting, Santa Fe, January 20-22, 1993. NEA/PSAG/DOC(93)3.

Appendix A
Regulations and Regulatory Guidance

A.1 United Kingdom

Environment Agency for England and Wales, Scottish Environmental Protection Agency, and Department of the Environment for Northern Ireland. 1997. Radioactive Substances Act 1993 - Disposal Facilities on Land for Low and Intermediate Level Radioactive Wastes: Guidance on Requirements for Authorisation. Environment Agency, Bristol.

In the UK, the guidance published by the regulators on their expectations includes a risk target, and notes (Para. 5.12) that the assessment of risk "… may be based on probabilistic techniques." Further guidance (Para. 6.15) defines radiological risk in such a way as to require some use of probabilistic techniques:

> "Radiological risk to a representative member of a potentially exposed group is the product of the probability that a given dose will be received and the probability that the dose will result in a serious health effect, summed over all situations that could give rise to exposure to the group."

The second probability in this definition is the dose-risk conversion factor, and it is in determining the probability that a given dose will be received that probabilistic techniques are likely to be required. The Agencies' assumption that probabilistic techniques will be necessary to support an application is further illustrated in Para. 8.21, which sets out some of the presentational requirements:

> "overall results from probabilistic risk assessments of the disposal system which explore the relevant uncertainties;
>
> suitable breakdowns of such risk assessments to show, for example, the probability distribution of doses and the contribution of important radionuclides;"

Despite this assumption that probabilistic assessments will be used, the Agencies also recognise (Para. 8.16) that alternative approaches to the treatment of uncertainty are appropriate in some circumstances:

> "… Other uncertainties may be eliminated from further consideration by making simple deterministic assumptions based on reasoned arguments. … Some uncertainties may be quantified and incorporated into numerical assessments of probability or risk. Quantification of other uncertainties may be inappropriate. Where such uncertainties are important to the case, they may be treated by making deterministic assumptions and exploring the effects of varying these."

Importantly, the Agencies also recognise that an overall safety assessment is not solely a question of comparing calculated risks with a regulatory target (Para. 8.20). Further, qualitative assurance measures are also expected:

> "… the Agencies take the view that sufficient assurance of safety is likely to be achieved only through considerations rather broader than purely the

evaluation of numerical values of risk, although this remains an important component of achieving such assurance."

A.2 United States

A.2.1 Waste Isolation Pilot Plant

> 40 CFR Part 191 Environmental Radiation Protection Standards for the Management and Disposal of Spent Nuclear Fuel, High-Level and Transuranic Radioactive Wastes
>
> 40 CFR Part 194: Criteria for the Certification and Re-Certification of the Waste Isolation Pilot Plant's Compliance with the 40 CFR Part 191 Disposal Regulations

In the USA, the Department of Energy (DOE) has responsibility for the disposal of transuranic wastes from defence-related sites and activities, and has developed the Waste Isolation Pilot Plant (WIPP) in southern New Mexico as a deep disposal facility for these wastes.

The WIPP Land Withdrawal Act (LWA) 1992, as amended, required the EPA to determine whether the WIPP will comply with EPA's standards for the disposal of radioactive waste. These standards include the generic standards at 40 CFR part 191, and also site-specific criteria for demonstrating compliance at 40 CFR part 197.

Regulatory standards

The principal regulatory measure used in assessments of the site is cumulative release to the accessible environment over 10,000 years. The accessible environment comprises the atmosphere, land surfaces, surface waters, and the lithosphere outside the controlled area. The controlled area is a region around the site that is in Federal ownership and in which activities such as drilling will be prevented during the period of institutional controls.

The cumulative release standard is a function of the radionuclide inventory disposed in the facility, rather than an absolute limit. Because there are different release limits for each nuclide or group of nuclides (see Table 1 below), a normalised release limit (the sum of the ratios between the release and the release limit for each radionuclide) is used to determine compliance. Compliance with this standard requires that the disposal system

> "… shall be designed to provide a reasonable expectation, based upon performance assessments, that the cumulative releases of radionuclides to the accessible environment for 10,000 years after disposal from all significant processes and events that may affect the disposal system shall:
>
> (1) Have a likelihood of less than one chance in 10 of exceeding the quantities calculated according to Table 1; and

(2) Have a likelihood of less than one chance in 1,000 of exceeding ten times the quantities calculated according to Table 1.

…

TABLE 1—RELEASE LIMITS FOR CONTAINMENT REQUIREMENTS [Cumulative releases to the accessible environment for 10,000 years after disposal]	
Radionuclide	**Release limit per 1,000 MHTM or other unit of waste (curies)**
Americium-241 or -243	100
Carbon-14	100
Caesium-135 or -137	1,000
Iodine-129	100
Neptunium-237	100
Plutonium-238, -239, -240 or -242	100
Radium-226	100
Strontium-90	1,000
Technetium-99	10,000
Thorium-230 or -232	10
Tin-126	1,000
Uranium-233, -234, -235, -236 or -238	100
Any other alpha-emitting radionuclide with a half-life greater than 20 years	100
Any other radionuclide with a half-life greater than 20 years that does not emit alpha particles	1,000

…"

The WIPP-specific criteria require that the results of performance assessments are presented in the form of complementary cumulative distribution functions (CCDFs) that represent the probability of exceeding various levels of cumulative release. These CCDFs can then be readily compared to the compliance criteria.

The regulatory requirements for the WIPP require the use of probabilistic techniques to demonstrate compliance, and there are other requirements that govern how the assessments are performed and presented:

- Probability distributions for uncertain disposal system parameter values must be documented.

- Techniques that draw samples from across the entire range of the probability distributions must be used in generating the CCDFs.

- The number of CCDFs generated shall be large enough such that the maximum CCDF generated exceeds the 99th percentile of the population of CCDFs with at least a 0.95 probability.

- Any compliance application shall display the full range of CCDFs generated.

- Any compliance application shall provide information which demonstrates that there is at least a 95 percent level of statistical confidence that the mean of the population of CCDFs meets the containment requirements.

The cumulative release criteria apply to releases arising from *all* significant processes and events that may affect the disposal system disturbed conditions. In the case of undisturbed performance, which excludes disruptive natural events as well as human intrusion, there are other compliance criteria.

- An annual committed effective dose limit of 0.15 mSv for any member of the public in the accessible environment, received through all pathways. The WIPP-specific criteria require assessment to assume that an individual resides at the single geographic point on the surface of the accessible environment where that individual would be expected to receive the highest dose.

- A groundwater protection requirement. Levels of radioactivity in any underground source of drinking water should not exceed the limits specified in the National Primary Drinking Water Standards (40 CFR part 140). These limits include:

    - combined Ra-226 and Ra-228: 5 picocuries per liter;

    - gross alpha particle activity, including Ra-226 but excluding radon and uranium: 15 picocuries per liter;

    - annual dose equivalent to the total body or any internal organ from the average annual concentration of beta particle and photon radioactivity from man-made radionuclides: 4 millirem per year.

All of these criteria apply over the compliance period of 10,000 years.

Compliance period

The EPA established a compliance period of 10,000 years for land disposal of spent nuclear fuel, high level radioactive waste and transuranic radioactive waste in their generic regulation at 40 CFR part 191. This compliance period was set for three principal reasons:

- After 10,000 years, there is concern that the uncertainties in compliance assessment become unacceptably large.

- There are likely to be no exceptionally large geologic changes at selected sites during the next 10,000 years.

- The use of time frames of less than 10,000 years does not allow for valid comparisons among potential sites. For example, for 1,000 years, any well-selected site would contain the waste approximately equally because of long ground water travel times at well-selected sites, and because any well-designed engineered barrier systems would have comparable containment capabilities over this period.

Human intrusion

The regulatory criteria for demonstrating the compliance of the WIPP prescribe many of the assumptions to be made in assessments of disturbed performance. In particular, these criteria limit the scope of human intrusion events that need to be considered to mining, deep drilling, and shallow drilling that may affect the disposal system during the regulatory time frame. Additional criteria include:

- The most severe intrusion scenario that needs to be considered is inadvertent and intermittent intrusion by drilling for resources (other than those resources provided by the waste in the disposal system or engineered barriers designed to isolate such waste).

- Rates for deep drilling must be determined on the basis of drilling in the Delaware Basin over the past 100 years, for all resources. Future drilling at this rate shall be assumed to take place randomly in time and space over the entire Delaware Basin. A similar approach is used for shallow drilling.

- Future drilling practices and technology should remain consistent with practices in the Delaware Basin at the time a compliance application is prepared. Such practices should include: the types and amounts of drilling fluids; borehole depths, diameters, and seals; and the fraction of such boreholes that are sealed by humans.

- The integrity of the boreholes and any seals emplaced should be assumed to be affected by natural processes over the regulatory time frame.

- Assessments need not analyse the effects of any resource recovery techniques used after a borehole is drilled.

- Mining can be assumed to have a probability of one in 100 in each century of the regulatory time frame, and assessments of mining effects may be limited to changes in the hydraulic conductivity of the hydrogeologic units of the disposal system from excavation mining for natural resources.

A.2.2 Yucca Mountain

40 CFR Part 197: Public Health and Environmental Radiation Protection Standards for Yucca Mountain, NV; Final Rule 10 CFR Part 63: Disposal of High-Level Radioactive Wastes in a Proposed Geologic Repository at Yucca Mountain, Nevada; Final Rule

The United States Nuclear Regulatory Commission (USNRC) is responsible for the regulation of disposal sites for high-level radioactive waste, including spent fuel, from the civilian and defence nuclear programmes in the United States. A generic regulation (10 CFR part 60), originally issued in 1983 but since updated, includes criteria for disposal in any geological repository. However, the Energy Policy Act, promulgated in 1992, directed USNRC to modify its technical requirements and criteria to be consistent with health and safety standards to be issued by the Environmental Protection Agency (EPA) specifically for Yucca Mountain. The EPA standards in turn, were to be based on the recommendations of the National Academy of Sciences (NAS), which were published in 1995. The EPA published its Public Health and Environmental Radiation Protection Standards for Yucca Mountain at 40 CFR part 197 on 13 June 2001, and the NRC published site-specific license criteria for the proposed repository at 10 CFR part 63 on 2 November 2001.

Protection standard

The individual protection standard established by the EPA is an annual dose to the reasonably maximally exposed individual (RMEI) of less than 0.15 mSv. Although the NAS recommended the adoption of a risk standard, the EPA elected to use a dose standard for a number of reasons:

- The Energy Policy Act explicitly required the setting of a dose standard, although this Act also required the EPA to issue standards “based upon and consistent with” the NAS’s findings and recommendations.

- A dose standard allows a convenient comparison with existing dose guidelines and standards, including the EPA’s general criteria at 40 CFR part 191.

- The dose standard is risk-based, in that the EPA judged that a risk of about 8.5 fatal cancers per million members of the population per year is acceptable. The dose standard is derived from this risk using a dose-risk conversion factor of 0.075 per Sv.

The dose-risk conversion factor accounts for one key uncertainty - the probability of developing fatal cancer after exposure to a radiation dose. The EPA’s dose standard does not necessarily require any other uncertainties to be accounted for, as the criteria for demonstrating compliance with the standard are the responsibility of the NRC. However, the EPA anticipates the use of performance assessments for estimating doses, and for accounting for uncertainties. Furthermore, the EPA anticipates that the annual committed effective dose equivalent incurred by the RMEI will be calculated based on releases arising through all significant features, events, processes, and sequences of events and processes, weighted by their probability of occurrence. The

EPA is therefore anticipating the use of probabilistic techniques to calculate dose, and specifies that numerical compliance should be based on the mean of the distribution of projected doses.

The NRC provides additional assumptions that the DOE must make in defining the RMEI. The RMEI is a hypothetical person who lives in the accessible environment above the highest concentration of radionuclides in the plume of contamination, has a diet and living style representative of the people who now reside in the Town of Amargosa Valley, Nevada, and is an adult with metabolic and physiological considerations consistent with present knowledge of adults. The RMEI drinks 2 litres of water per day from wells drilled into the ground water at the point of highest concentration, and the well water used has average concentrations of radionuclides based on an annual water demand of 3,000 acre-feet (approx. 3.7 x 10^6 m^3).

Compliance period

Although the 40 CFR part 191 standards apply to the same types of waste and type of disposal system as will be present at Yucca Mountain, the WIPP Land Withdrawal Act exempts Yucca Mountain from the 40 CFR part 191 standards and so a different compliance period could have been established. However, in developing the Yucca Mountain regulations, the EPA determined that the issue of uncertainties over long time frames, and the use of performance projections over those time frames for regulatory decision making, had not changed significantly since their earlier analysis. The EPA therefore determined that a 10,000 year compliance period for Yucca Mountain would be appropriate for consistency with the earlier regulations.

The initial analyses undertaken by the DOE on the performance of the Yucca Mountain facility suggest that most radionuclides would not reach currently populated areas within 10,000 years, because of the expected performance of the engineered barrier system. This seems to indicate that the compliance period for Yucca Mountain should be longer than 10,000 years if exposures are to be considered when they are calculated to occur. The NAS noted that, although there are significant uncertainties in a performance assessment and that the overall uncertainty increases with time, "… there is no scientific basis for limiting the time period of the individual-risk standard to 10,000 years or any other value". Also, NAS stated that many of the uncertainties in parameter values describing the geological system are not due to the length of time but rather to the difficulty in estimating values of site characteristics that vary across the site. Thus, NAS concluded that the probabilities and consequences of the relevant features, events, and processes that could modify the way in which radionuclides are transported in the vicinity of Yucca Mountain, including climate change, seismic activity, and volcanic eruptions, "… are sufficiently boundable so that these factors can be included in performance assessments that extend over periods on the order of about one million years".

Despite the NAS's conclusion, the EPA determined that there is still considerable uncertainty as to whether current modelling capability can provide sufficiently meaningful and reliable projections over a time frame up to tens-of-thousands to hundreds-of-thousands of years. Importantly, the EPA noted that, simply because models can provide projections for those time periods does not mean those projections are meaningful and reliable enough to establish a rational basis for regulatory decision making.

The final rule applicable to Yucca Mountain (40 CFR part 197) was based on a consideration of both the technical and policy issues connected with establishing a compliance period. As a result, the EPA established a 10,000-year compliance period with a quantitative limit and a requirement to calculate the peak dose, using performance assessments, if the peak dose occurs after 10,000 years.

Human intrusion

In 40 CFR part 197, the EPA requires that the DOE consider human intrusion in its assessment of Yucca Mountain. Intrusion need only be considered at times beyond the point at which the waste will have degraded to such an extent as to not be recognised by drillers.

The type of intrusion to be considered in the analysis is prescribed as a single borehole drilled for groundwater exploration, which passes through the degraded waste and into the underlying aquifer. The borehole is drilled using present-day techniques, but is left unsealed and allowed to degrade naturally. Only transport of radionuclides downwards from the repository, through the borehole and into the underlying saturated zone need be considered in the intrusion analysis. In 10 CFR part 60, the NRC clarified this restriction to mean only radionuclides transported by water, and not those transported as particles.

As with the individual protection standard, the requirements are different depending on when the intrusion is projected to occur. If the intrusion is projected to take place after 10,000 years (or the exposure of the RMEI does not take place until after 10,000 years), there is no numerical standard, but the DOE must include the results of its analyses in the Environmental Impact Statement as an indicator of long-term safety. If intrusion is projected to occur at or before 10,000 years after disposal, then the DOE must demonstrate that there is a reasonable expectation that the RMEI receives no more than an annual committed effective dose equivalent of 0.15 mSv as a result of the intrusion.

Reference biosphere

The NRC in 10 CFR part 63 requires the DOE to define a reference biosphere for use in calculating doses for the assessment of both the individual protection and human intrusion standards. The features, events, and processes that describe the reference biosphere must be consistent with present knowledge of the conditions in the region surrounding the Yucca Mountain site, and the biosphere pathways assessed must be consistent with arid or semi-arid conditions. Reasonable assumptions that are consistent with present knowledge of the factors that could affect the Yucca Mountain disposal system must be made in order to assess the evolution of the geological, hydrogeological and climatic systems over the next 10,000 years. However, changes in society, the biosphere (other than climate), human biology, or increases or decreases of human knowledge or technology must not be assumed and these factors should remain as they are at the time of submission of an application.

Groundwater protection

In addition to the individual dose criteria, the EPA standard also includes standards for groundwater protection, based on radionuclide concentrations (see Table A.1

below). The EPA recognises that the individual-protection standard includes a drinking water exposure pathway, and that to some extent the groundwater protection standard is redundant. It is, however, a general EPA policy to provide separate protection for groundwater resources, and the standard developed for Yucca Mountain is consistent in this respect to the standard (40 CFR part 194) developed for the WIPP site. The EPA also notes that the aquifers in the Yucca Mountain area could potentially supply a larger population than currently uses them, either through a growth in the local population or through transport to other locations. Developing scenarios to encompass such potential exposure pathways, and assessing the resulting health effects, would involve significant speculation. The EPA believes that it is more appropriate to assure the resource is not contaminated in the first place.

Table A.1. Limits on radionuclides in the representative volume specified in the 40 CFR part 197.

Radionuclide or type of radiation emitted	**Limit**	**Is natural background included?**
Combined radium-226 and radium-228	5 picocuries per litre	Yes
Gross alpha activity (including radium-226 but excluding radon and uranium)	15 picocuries per litre	Yes
Combined beta and photon emitting radionuclides	40 microsieverts (4 millirem) per year to the whole body or any organ, based on drinking 2 litres of water per day from the representative volume	No

A.3 Canada

Atomic Energy Control Board Regulatory Document R-71. Deep geological disposal of nuclear fuel waste: Background information and regulatory requirements regarding the concept assessment phase.

Atomic Energy Control Board Regulatory Document R-104. Regulatory objectives, requirements, and guidelines for the disposal of radioactive wastes - long-term aspects.

In Canada, regulatory responsibilities have recently changed, with the creation of the Canadian Nuclear Safety Commission (CNSC) and the introduction of the Nuclear Safety and Control (NSC) Act in May 2000. The regulations made pursuant to the NSC Act for nuclear facilities are general in nature, and there are no specific regulations for waste management. In April 2001, a draft Nuclear Fuel Waste Act (Bill C-27) was introduced, which will require nuclear utilities to form a waste management organization and establish a segregated trust fund to finance long-term nuclear fuel waste management activities.

CNSC staff are in the process of reviewing the regulatory documents (policies and guides) that relate to radioactive waste disposal, and are in the process of drafting a replacement regulatory guide for the existing Regulatory Policy Statement R-104 (entitled "Regulatory objectives, requirements and guidelines for the disposal of radioactive wastes - long-term aspects"). A draft for public consultation should be prepared by mid-2002.

In the absence of the revised guidance, the following summary of the regulatory guidance in Canada is based on the guidance in place in 1994, at the time of the most recent detailed assessment. The regulator at that time was the Atomic Energy Control Board (AECB). The AECB Regulatory Document R-71 provides guidance on the assessment of disposal concepts, and AECB Regulatory Document R-104 provides additional guidance on the requirements for assessments of specific sites. Assessments of concepts (R-71) require estimates of effective dose, while site-specific assessments (R-104) require determinations of risk. R-104 defines risk as:

> " … the probability that a fatal cancer or serious genetic effect will occur to an individual or his or her descendants. Risk, when defined in this way, is the sum over all significant scenarios of the products of the probability of the scenario, the magnitude of the resultant dose and the probability of the health effect per unit dose."

There is, therefore, an implicit assumption that site-specific performance assessments will use probabilistic techniques. R-104 provides guidelines on the application of the basic radiological requirements to the post-closure assessment. Guideline 2 provides guidance on scenario probabilities:

> "The probabilities of exposure scenarios should be assigned numerical values either on the basis of relative frequency of occurrence or through best estimates and engineering judgements.
>
> The use of subjective probability is appropriate as long as the quantitative values assigned are consistent with the quantitative values of the actual relative frequencies in situations where more information is available. The uncertainty of the probability assigned should also be estimated."

Guideline 2 provides guidance on the calculation of risk:

> "Calculations of individual risk should be made by using the risk conversion factor of 2 x 10^{-2} fatal cancers and serious genetic effects per sievert and the probability of the exposure scenario with either:
>
> (a) the annual individual dose calculated as the output from deterministic pathways analysis; or

(b) the arithmetic mean value of annual individual dose from the distribution of individual doses in a year calculated as the output from probabilistic analysis.”

Appendix B
Approaches to Risk Assessment

B.1 United Kingdom: HMIP Dry Run 3

Sumerling, T.J. (Editor). 1992. Dry Run 3 - A trial assessment of underground disposal of radioactive wastes based on probabilistic risk analysis - Volume 10: Overview. UKDOE/ HMIP Report DoE/HMIP/RR/92.039.

Thorne, M.C. 1992. Dry Run 3 - A trial assessment of underground disposal of radioactive wastes based on probabilistic risk analysis - Volume 8: Uncertainty and bias audit. UKDOE/ HMIP Report DoE/HMIP/RR/92.040.

Dry Run 3 was a trial probabilistic assessment of the deep disposal of radioactive wastes, undertaken by Her Majesty's Inspectorate of Pollution (HMIP)[5], as part of its independent regulatory oversight responsibility. Dry Run 3 was not a full assessment, but rather a demonstration of a time-dependent probabilistic assessment methodology. A key aspect of this methodology was the use of the TIME4 / VANDAL modelling system, which accounts for the uncertainties associated with possible future evolutions of the natural environmental system.

The HMIP assessment methodology was also subsequently applied to an assessment of the Sellafield site that was considered by UK Nirex for construction of an underground laboratory. Some model development was undertaken for this later assessment but the overall methodology was the same as that used for Dry Run 3, and there is less publicly available documentation of the Sellafield assessment.

HMIP's assessment methodology comprised a Reference assessment based on the use of Monte Carlo techniques and following a well-defined procedure, and a parallel assessment of "uncertainty and bias".

Hypothetical future evolutions of the environmental system were constructed using the TIME4 computer code. TIME4 includes models for climate change, sea-level change, glaciation, glaciation effects, surface hydrology, and denudation. These sub-models are coupled, and the output of TIME4 analyses is used to provide a time series of time-dependent boundary conditions (e.g., recharge, hydraulic heads) for individual runs of the radionuclide transport simulation code, VANDAL.

Four climate states are defined within TIME4, and within each climate state climate-related data (e.g., precipitation) are sampled and used to calculate parameters such as surface water flows, recharge, erosion and deposition. Within VANDAL, groundwater flow calculations are performed on the assumption that the conditions governing flow are constant within each climate state. Within each climate state, time-dependent calculations of radionuclide transport are performed, taking account of the distribution of radionuclides resulting from groundwater flow and radionuclide transport in previous states, and radioactive decay and ingrowth.

The biosphere submodel within VANDAL (DECOS-MG) simulates the movement of radionuclides in the aqueous and solid phases in groundwater, springs, and wells, and

---

5 In 1996, HMIP was incorporated within the newly formed Environment Agency for England and Wales.

between compartments representing subsoils, surface soils, rivers, estuaries, and marine environments. Inter-compartment transfer rates are calculated from parameters provided by TIME4. Individual doses to members of critical groups are calculated based on radionuclide concentration in specified compartments, and dose-rate conversion factors that depend on the climate state in which the dose is received. The conversion factors account for exposure via ingestion, inhalation, and external irradiation.

Annual doses from each VANDAL simulation are converted to risk using the ICRP dose - risk conversion factor, and the mean risk from all simulations is calculated at specified time points. Convergence criteria based on the mean annual risk are used to establish that sufficient simulations have been conducted. Simulations that contribute most to the mean annual risk are re-examined to ensure that they do not represent unrealistic combinations of sampled values or other model inadequacies.

Explicit distinction is made in the HMIP methodology between **uncertainties**, which can generally be incorporated into an assessment through the use of parameter value distributions, and **biases**, which arise from limitations in the conceptualisation of the assessment system or in the tools available for its analysis. Biases in this sense could include “uncertainty” associated with the selection and implementation of conceptual models used in the assessment, including the choice of alternative futures to be modelled.

As part of Dry Run 3, an uncertainty and bias audit (UBA) methodology was developed and parts of it were tested. In particular:

- An Expert Group was used to advise on phenomena that should be included in a comprehensive assessment of a repository for low- and intermediate-level waste at the Harwell site, and to determine priorities for modelling.

- Scoping calculations were conducted to provide illustrative results that were used to demonstrate the potential biases inherent in the trial assessment, through the omission of particular processes from the system model, or to investigate alternative parameter or model choices. For example, calculations were made for meteorite impact, gross incision through glacial action, gas transport, and human intrusion.

- Expert judgement was used to assess the scope of Dry Run 3 relative to a “minimal assessment”, as defined by the experts. A minimal assessment was defined to mean the assessment with the lowest degree of complexity that would not exhibit “significant” bias because of excluded phenomena, given that available models and data were adequate for representing those phenomena that were included. “Significant” in this context was apparently defined as an order of magnitude difference in the calculated consequences.

The conclusions of the Dry Run 3 exercise suggest that, for the particular assumptions of the trial assessment, the sequences of environmental change were important, and led to results that could not have been obtained by combining results from simulations for different constant environmental conditions (time-independent scenarios). In fact, the risk estimate from the fully time-dependent approach was about two orders of magnitude greater than that obtained by a nominally equivalent simulation for constant (present-day) environmental conditions. The extent to which this conclusion applies to other sites and disposal concepts, and hence the requirement for coupled models of environmental change, remains uncertain.

B.2 United Kingdom: Nirex 97

Nirex 97: An Assessment of the Post-closure Performance of a Deep Waste Repository at Sellafield. Overview. Norris, S., Bailey, L.E.F., Askarieh, M.M., Hickford, G.E. UK Nirex Report S/97/012.

UK Nirex has responsibility for the development of disposal strategies for solid intermediate-level and some low-level radioactive waste. In 1991, Nirex selected an area near Sellafield in Cumbria as the focus for further investigations relating to deep disposal, and in 1994 applied for planning permission for a Rock Characterisation Facility (RCF). Permission for the RCF was refused in 1997 after a planning inquiry, and Nirex's site characterisation and assessment programmes were curtailed. A description of the assessment methodology that had been developed, and some provisional results, were published by Nirex as a demonstration of the capability that had been developed to assess the radiological performance of candidate repository sites. The assessment approach developed by Nirex and described in "Nirex 97" is summarised below.

Nirex recognises performance assessment as playing a central role in an iterative cycle that builds system understanding and identifies future site characterisation and research priorities. A principal requirement of the performance assessment is the calculation of risk to an individual from releases of radionuclides from the repository. Other, qualitative, factors associated with the assessment calculations that help to build confidence are also important requirements.

The acknowledgement and analysis of uncertainty forms part of Nirex's assessment methodology. Five sources of uncertainty are recognised:

- the limited characterisation that can he achieved of the present state of the facility and its surroundings;
- incomplete knowledge of future environmental conditions and of the natural events that will affect the system performance;
- a degree of unpredictability in future human behaviour, as it affects not only potential exposures to radioactivity but also perturbations to the behaviour of the system;
- the existence of alternative defensible choices of conceptual and mathematical models; and
- the effects of approximations made in modelling the system.

Some of these sources of uncertainty are treated in the assessment through use of probabilistic techniques. These involve the definition of probability density functions to characterise uncertainty and the use of Monte Carlo simulation to develop distributions of output quantities (e.g., dose). At the time of Nirex 97, the assessment models allowed for uncertainties in site characterisation data and in some aspects of environmental evolution and human behaviour. Other aspects of environmental evolution and some aspects of human behaviour were treated through alternative scenarios, in each of which environmental conditions were assumed to be constant.

Uncertainties arising from the existence of alternative conceptual models for parts of the overall disposal system were treated in several ways. In some cases, a "scenario-like" approach was used, with separate assessment calculations undertaken for each alternative. In other cases, a pdf representing the "degree-of-belief" in each model was developed. Each simulation in a set of assessment calculations then sampled this pdf to determine which alternative model was used for that simulation. The final approach was to not explicitly incorporate alternative conceptual models in the assessment, but to specify a parameter representing their output. By specifying a pdf for this parameter that encompasses the range of outputs from all of the alternative models, the effect of the alternative models can be incorporated in the assessment.

In addition to the probabilistic assessment calculations, Nirex also undertook some deterministic calculations. These necessitated the selection of conservative models and parameter values such that risks or other performance measures were over-estimated. The conservative approach was recognised as problematic in a number of respects:

- The combination of a large number of conservatisms may lead to the prediction of unacceptable system performance.

- Calculated performance will poorly represent expected performance and therefore be unsuitable for design optimisation.

- It may not be possible to specify models or data that are conservative under all conditions.

- Even conservative assumptions will not absolutely bound system performance. Selected values are effectively those for which there is only a low probability of exceedance. If this probability is assessed, the conservative approach effectively becomes a probabilistic analysis of the upper tail of the consequence distribution.

Deterministic modelling was undertaken for a number of subsystems, such as a single disposal vault and groundwater transport through the geosphere. The results of these were used in the development of simplified models and / or definition of probability density functions (pdfs) for use in the probabilistic framework.

Nirex 97 was based on the use of the MASCOT code, which allows for the linking of sub-models from a library and also allows input values to be specified as pdfs or as algebraic functions of other parameters. This last capability allows parameter correlations to be introduced. The computing efficiency of MASCOT was achieved by restricting sub-models and their parameters to being time-invariant; a time-dependent approach was under development for use in the next phase of assessment calculations.

The output of MASCOT is a contingent dose versus time curve for each simulation, with the dose contingent upon the sampled parameter values. Nirex used a process of structured elicitation of expert opinion to derive the pdfs from which values were sampled. The contingent dose was converted to risk using the ICRP dose-risk factor. Post-processing of the results, to give summary measures such as mean risk, or to examine correlations and sensitivities, was performed using the MOP code.

B.3 United States: Compliance Certification Application for the Waste Isolation Pilot Plant

Title 40 CFR Part 191 Compliance Certification Application for the Waste Isolation Pilot Plant, October 1996. DOE/CAO-96-2184, U.S. Department of Energy, Carlsbad NM.

The performance assessment of the Waste Isolation Pilot Plant (WIPP) used a methodology based on the work of Kaplan and Garrick (1981), which was developed to estimate the effects of uncertain but characterisable futures. In Kaplan and Garrick's procedure, each of the possible futures is associated with a probability of occurrence and a consequence of occurrence. Preliminary performance assessments of the WIPP used this procedure (for example, see Sandia National Laboratories 1991; 1992-1993), but the definition of the futures as discrete entities resulted in a great number of possible futures being defined.

For the assessment presented in the Compliance Certification Application (CCA), future were defined using direct probabilistic sampling of the possible events leading to uncertain futures rather than by an *a priori* definition of possible futures. This approach is not inconsistent with the concept of risk developed by Kaplan and Garrick and did not affect the outcome of the overall analysis. Rather, the new procedure was prompted by two practical considerations. First, it is difficult to define futures as literal entities and to develop probabilities for each one. Second, generation of the futures by probabilistic methods allows for greater resolution in a complementary cumulative distribution function (CCDF) . Furthermore, the new procedure provided a structure for the calculations, making them more efficient, and provided a means for improved presentation of the results.

The CCA performance assessment was based on the use of scenarios, which were defined as subsets of the set of all possible occurrences within the 10,000-year regulatory time frame. Each scenario comprises a subset of similar future occurrences. Scenarios were determined through a formal process similar to that proposed by Cranwell *et al.* (1990), and the process used in preliminary performance assessments for the WIPP. This process involved four steps:

(1) FEPs (features, events, and processes) potentially relevant to the WIPP were identified and classified.

(2) Certain FEPs were eliminated according to well-defined screening criteria as not important or not relevant to the performance of the WIPP.

(3) Scenarios were formed from the remaining FEPs, in the context of regulatory performance criteria.

(4) Scenarios were specified for consequence analysis.

Steps (1) and (2) of the scenario development process were aimed at identifying "all significant processes and events that may affect the disposal system" as required by the regulations (40 CFR § 191.13(a) and 40 CFR § 194.32).

An initial list of FEPs was developed for the CCA. This list was intended to ensure that the identification of significant processes and events was complete, that potential interactions between FEPs were not overlooked, and that responses to possible questions were available and well documented.

For practical purposes, the calculation of stochastic uncertainty (related to aggregation and stochastic variation), represented in an individual CCDF, was separated from subjective uncertainty (uncertainty because of, for example, measurement difficulties or incomplete data), represented by the family of CCDFs. This approach can be represented mathematically as a double integral, although because an analytical solution was not available, the computational framework provided a means of calculating a double sum:

$$\sum_{su}\sum_{st} F(x)$$

A separate probabilistic analysis was performed to evaluate each sum. For example, uncertainty in the number and time of intrusion boreholes was included in the inner, stochastic, sum, and the outer, subjective, sum included a probabilistic characterisation of site properties, such as the permeability of specific rock types.

In certain cases, it is not obvious whether a particular uncertainty should be classified as subjective or stochastic. For example, whether currently observed geologic properties persist through time could be thought of as either subjective or stochastic uncertainty. For the CCA, uncertainty associated with significant future human actions was treated as stochastic (for example, drilling for natural resources), and uncertainty in disposal system properties that are subject to ongoing physical processes was treated as subjective (for example, climate change or gas generation).

Monte Carlo analysis was used for the probabilistic analysis of the WIPP. This involved five key steps: (1) selection of the variables to be examined and the ranges and distributions for their possible values, (2) generation of the samples to be analysed, (3) propagation of the samples through the analysis, (4) uncertainty analysis, and (5) sensitivity analysis. Within this general Monte Carlo framework, the performance assessment used two methods for generating the samples propagated through the model system. One method was used for the assessment of stochastic uncertainty, and another method was used for the characterisation of subjective uncertainty. Each of these methods used the same five steps but differed in methodology in Steps 2 through 5.

Information about the ranges and distributions of possible values was drawn from a variety of sources, including field data, laboratory data, and literature. In instances where sufficient data were not available, documented expert judgement was used. Judgement of the investigators and analysts involved was also important in the review process that led from the available data to the construction of the distribution functions used in the performance assessment. In part, this review process addressed the scaling of data collected at experimental scales of observation to the development of the parameter ranges applied to scales of interest in the disposal system.

Various techniques are available for generating samples from assigned distribution functions for the variables, including random sampling, stratified sampling, and Latin Hypercube Sampling (LHS). The performance assessment for the CCA used both random sampling and LHS.

Random sampling of the occurrence of possible future events was used to generate the possible futures (probabilistic futures) that formed a CCDF. This sampling was used to select values of uncertain parameters associated with future human activities, or in other words, to incorporate stochastic uncertainty into the assessment. This approach was used so as to reduce errors from aggregation that can arise with other sampling methods.

LHS, in which the full range of each variable is subdivided into intervals of equal probability and samples are drawn from each interval, was used to select values of uncertain parameters associated with the physical system being simulated. In other words, LHS incorporated subjective uncertainty into the WIPP performance assessment.

A Software Configuration Management System (SCMS) was developed to facilitate the calculations performed by the model system and to store the input and output files from each program. These files included sampled input values and the model system predictions and were saved for use in uncertainty and sensitivity studies.

The regulations that apply to the WIPP site require separate assessments of undisturbed and disturbed performance. Undisturbed performance is the predicted behaviour of a disposal system, including consideration of the uncertainties in predicted behaviour, if the disposal system is not disrupted by human intrusion or the occurrence of unlikely natural events. No potentially disruptive natural events or processes were identified as likely to occur during the regulatory time frame. The undisturbed performance scenario therefore accounted for all of the naturally occurring FEPs remaining after screening.

The only future human-initiated events and processes retained after FEP screening were those associated with mining and deep drilling within the controlled area after institutional controls have failed. The consequences of disturbed performance were evaluated through analysis of a mining scenario, a deep drilling scenario, and a mining and drilling scenario.

- The disturbed performance mining scenario assumed future mining within the controlled area. Regulatory criteria limit the effects of potential future mining within the controlled area to changes in hydraulic conductivity in the overlying aquifer unit.

- The disturbed performance deep drilling scenario involved at least one deep drilling event intersecting the waste disposal region. Regulatory criteria led to the assumptions that future drilling practices and technology will remain consistent with present-day practices in the Delaware Basin, and that natural processes will degrade or otherwise affect the capability of boreholes to transmit fluids over the regulatory time frame.

- Mining at the WIPP site (the M scenario) and deep drilling (the E scenario) may both occur in the future. A future in which both of these events occur was assessed as an ME scenario. The occurrence of both mining and deep drilling was assumed not to create processes in addition to those for the M and E scenarios.

References

Cranwell, R.M., Guzowski, R.V., Campbell, J.E., and Ortiz, N.R. 1990. Risk methodology for geologic disposal of radioactive waste: scenario selection procedure. NUREG/CR-1667, SAND80-1429. Sandia National Laboratories, Albuquerque, NM.

Kaplan, S., and Garrick, B.J., 1981. On the quantitative definition of risk. Risk Analysis, Vol. 1, No. 1, pp. 11 - 27.

Sandia National Laboratories. 1991. Preliminary Comparison with 40 CFR Part 191, Subpart B for the Waste Isolation Pilot Plant, December 1991. Volume 1: Methodology and Results. SAND91-0893/1. Sandia National Laboratories, WIPP Performance Assessment Division, Albuquerque, NM.

Sandia National Laboratories. 1992-1993. Preliminary Performance Assessment for the Waste Isolation Pilot Plant, December 1992. SAND92-0700, Vols. 1 - 5. Sandia National Laboratories, WIPP Performance Assessment Division, Albuquerque, NM.

B.4 United States: Total System Performance Assessment for the Yucca Mountain Viability Assessment

Viability assessment of a repository at Yucca Mountain. Volume 3: Total System Performance Assessment. DOE/RW-0508. U.S. Department of Energy.

The term Total System Performance Assessment (TSPA) is applied to a performance assessment in which all of the components of a system are linked into a single analysis. This type of assessment can be thought of as representing the uppermost level of a performance assessment "pyramid", in which the lower levels comprise abstracted performance assessment models, conceptual and process models, and design and site data. Each step up the pyramid represents successively more distillation and abstraction of information. The performance assessment cycle represented in this way also provides for feedback from the results of higher level calculations to subsequent iterations of site characterisation and detailed modelling studies.

The Energy and Water Development Appropriations Act (1997) required a TSPA to be conducted for an assessment of the viability of the Yucca Mountain site, using the design concept and scientific data available on September 30, 1997. The aim of this TSPA was to describe the probable behaviour of the repository in the Yucca Mountain geological setting, and to provide a basis for decisions on the potential of the Yucca Mountain site to fulfil the regulatory requirements under development by the EPA and NRC (see Section A.2.2).

The disposal concept at Yucca Mountain is based on four key attributes:

- Limited water contacting the waste packages.
- Long waste package lifetimes.
- Low rates of radionuclide release from breached packages.
- Radionuclide dilution during transport from the packages.

These four attributes provide the basis for the overall design, site characterisation, and conceptual and process models. For example, the attribute relating to package lifetime is supported by models of the near-field geochemical environment and waste package degradation, and the attribute relating to low release rates is supported by models of waste package and waste form degradation, radionuclide mobilisation and transport within the engineered barrier. Additional models not directly related to the key attributes focus on potentially disruptive events that could affect system performance.

The TSPA for the Viability Assessment was based on these attributes and model structure, and comprised five key steps:

- Develop and screen scenarios.
- Develop models.
- Estimate parameter ranges and uncertainties.
- Perform calculations.
- Interpret results.

The principal scenario was the base case or expected evolution of the disposal system, which includes the features, events and processes (FEPs) associated with each of the key attributes that are expected to occur. Other FEPs can occur that could affect the behaviour of the system, but these have a sufficiently low probability of occurrence over the period of interest that they were not considered as part of the expected evolution scenario. These unanticipated processes and events, such as igneous activity, seismic activity, criticality events and human intrusion, were examined in alternative scenarios and sensitivity studies.

The evolution of the disposal system involves a variety of coupled processes (thermal-hydrological-chemical and thermal-hydrological-mechanical) acting in three dimensions in a variety of materials and varying with time. This meant that a level of simplification was required even at the level of detailed process models, and that further simplifications and model abstractions were required in developing the performance assessment models. Key aspects of the model development process were ensuring that information passed up the model "pyramid" was consistent, and ensuring that parameters that most affect performance in detailed process models were appropriately represented in subsystem and system models. Data consistency was ensured by the use of multi-dimensional "response surfaces" describing the output from detailed models and used, with interpolation, as inputs to assessment models. Parameter sensitivities in the detailed models could not be assessed against the final performance measure (dose or risk), but were assessed against "surrogate" measures such as the amount of fracture flow in the unsaturated zone.

The overall assessment calculations was conducted using an integrating shell linking all the various component codes. This shell (RIP) coupled component models in four ways:

- External function calls to detailed process codes.
- RIP cells (equilibrium batch reactors) that provide a reasonable description of certain processes.
- Response surfaces.
- Inclusion of functional or stochastic representations of a component model in the shell.

Because much of the computational work was done outside of RIP, the multiple realisation runs required to incorporate uncertainty through probabilistic sampling were performed efficiently. Probabilistic sampling was used to account for both parameter uncertainty and alternative conceptual models. Index parameters that allow different conceptual models to be incorporated were used in cases where weighting of the alternatives could be justified (lumping). In other cases, a particular conceptual model was used in the base case calculations and the effects of alternatives were explored in separate calculations (splitting). Examples of these separate calculations included widened parameter ranges for cladding degradation models, alternative process models for unsaturated flow, one-dimensional flow and transport models for the saturated zone, and disruptive events such as volcanism.

Results were presented as complementary cumulative distribution functions (CCDFs) of peak dose, which was seen as the most important measure of system performance. The most probable behaviour was assumed to be the mode of the CCDF, although other points such as the mean or median are more generally regarded as representative of system behaviour. Dose rate versus time curves were also used to illustrate system behaviour. A single realisation, using the expected value of all input parameters, was used to show the effects of component interactions.

In addition to the expected-value realisation, the behaviour of realisations that lie close to the tails of the various input parameters was also examined. For each parameter in turn, a realisation using the 5th or 95th percentiles and the expected value of other parameters was used to illustrate the potential behaviour of the system. Similarly, realisations using alternative models or parameter sets were used to show sensitivities to different assumptions.

Other approaches to sensitivity analysis were also used, including scatter plots, regression analysis and differential analysis. Scatter plots of dose rate against input parameters showed trends for sensitive parameters. Plots of performance measures such as dose rate against subsystem outputs were also useful in revealing whether subsystems are important to performance. Regression analysis, specifically stepwise linear regression, was used to determine the effect of parameters on the variance of the output and provide a ranking and quantitative measure of the impact of parameters on performance. Differential analysis, using partial derivatives of the performance measures with respect to each input parameter, was also used to rank parameters, although this method is limited to a particular part of the overall parameter space and does not necessarily indicate sensitivities across the whole parameter range.

B.5 Canada: Postclosure Assessment of a Reference System

The disposal of Canada's nuclear fuel waste: Postclosure assessment of a reference system. Goodwin, B.W. *et al.* Atomic Energy of Canada Limited. AECL-10717, COG-93-7.

Atomic Energy of Canada Limited (AECL) developed an Environmental Impact Statement (EIS) for a disposal concept to manage Canada's nuclear fuel waste. The EIS included descriptions of engineering, barriers, site screening and evaluation approaches as well as assessments of pre-closure and post-closure performance. The EIS was submitted for public review, with the intent that, if the concept was considered broadly acceptable, more detailed studies would lead to the selection and detailed analysis of a particular site.

Because the EIS and its components focussed on a disposal concept rather than on a particular site, some elements of the assessment were based on assumptions rather than on actual data or information. Nevertheless, the structure of the assessment followed the approach that would be used for an assessment of an actual site and included a review of features, events and processes, scenario development, derivation of conceptual models and implementation of a probabilistic system model for calculating risks.

The disposal concept comprised a vault 500 m below the surface in the granitic rocks of the Canadian Shield. In this design, the vault contains 434 disposal rooms, each with 288 boreholes containing a single titanium canister. Sand and bentonite forms a buffer within the boreholes and glacial clay and crushed rock are used to backfill the rooms and access tunnels.

FEP screening involved the assessment of about 300 factors against several criteria, including their probability of occurrence, their effect on estimates of risk, and their relevance to the context of the assessment. Using this criteria, about 150 factors were screened out of the analysis and the remainder were used to define scenarios. Three types of scenarios were developed:

- **SYVAC scenarios**; these comprised most of the screened-in factors and were focussed on the groundwater pathway. For assessment purposes, the probability of these scenarios was assumed to be unity.

- **Open-borehole scenarios**; these described a situation in which one or more open boreholes pass near to the disposal vault. All the factors from the SYVAC scenarios were also included in these scenarios. No probability was assigned to these scenarios, but quality assurance procedures are assumed to limit the possibility of an open borehole. These scenarios were therefore assessed as not contributing significantly to the radiological risk.

- **Disruption scenarios**; one disruptive event, inadvertent human intrusion, was identified as likely to contribute significantly to the risk over 10,000 years. A probability of less than 5×10^{-6} per year was proposed for these events, based on the assessed probabilities of proposing to drill a borehole within the repository footprint, institutional controls failing and intersecting a canister.

Quantitative assessments of these scenarios were carried out for 10,000 years. Beyond this time, reasoned arguments were used to supplement the quantitative analysis.

The SYVAC model used in the assessment was a probabilistic system model integrating models of the vault, the geosphere and the biosphere. Parameter uncertainty was taken into account through the use of probability density functions (pdfs) derived from measurements and observations. Scaling factors were also used to incorporate uncertainties in parameters such as groundwater velocities derived from other models. These factors allowed the SYVAC calculations to account for much larger ranges of groundwater velocities than detailed hydrogeological models. Extensive use was also made in SYVAC of switches to control the selection of mutually exclusive options. For example, a switch was used to select the use of a well as a source of water, and a switch was used to select the use of lake sediment as soil. Each different combination of options can be considered as a simple scenario.

Sensitivity analysis was used to determine the key parameters and switches, and these were re-examined using further probabilistic analysis but restricting the range of switches used. These analyses showed two parameters to have correlation coefficients greater than 0.5. One of these, tortuosity of the lower rock zone, is a continuous parameter. The second, the source of domestic drinking water, is a switch. A further six parameters and switches were assessed as being important in terms of dose estimates.

Each SYVAC simulation provided an estimate of dose as a function of time, contingent upon the options selected by switches and the sampled parameter values. For determining the performance of the reference system, 40,000 simulations were performed and the maximum annual dose determined. Since each simulation was deemed equally likely, the expectation value of the resulting maximum dose histogram was used for comparison with the regulatory criterion. Although the criterion is expressed as a risk of less than 10^{-6} a^{-1}, the comparisons were all made with the corresponding dose of 0.05 mSv a^{-1}.

The system model was used to assess the effects of design variants, such as depth of the vault and different conceptual models, such as different soil conditions. The expectation value of dose for each variant, from 500 simulations, was normalised to the dose from the reference cases for comparison purposes. Parameters such as container thickness had comparatively little effect on dose estimates, but locating the vault above a major fracture zone increased calculated doses by more than two orders of magnitude. Decreases in calculated doses resulted from longer container lifetimes, increasing the distance of the vault from the fracture zone and locating the drinking water well further from the facility.

In the case of the open borehole scenarios, results from the SYVAC model were used to determine the minimum distance between the vault and an open borehole that would result in a significant increase in dose. In fact, the minimum distance analysed was 30 m, which gave an increase of 1 x 10^{-6} mSv a^{-1}, so that this a pessimistic value for the minimum acceptable distance. The aim of this analysis was to provide a reasonable basis for deriving quality assurance procedures that would prevent open boreholes. Preventing them from occurring within 30 m of the vault is seen as an achievable goal for such procedures.

Doses arising from the human intrusion scenarios were calculated in a similar manner to the other scenarios. Very high doses were calculated for early intrusions, so in order to demonstrate compliance with the criterion an assessment was made of intrusion probabilities. This was done using a fault-tree approach to examine the various events that lead up to an intrusion. AECL considered it more reasonable to apply judgement to the probabilities of these events than to elicit an overall probability.