Human and Ecological Risk Assessment, 12: 120–129, 2006

ISSN: 1080-7039 print / 1549-7680 online
DOI: 10.1080/10807030500430476

Statistical Issues in Ecological Risk Assessment

David R. Fox
Australian Centre for Environmetrics, Parkville Victoria, Australia

ABSTRACT

Ecological risk assessment (ERA) is concerned with making decisions about the natural environment under uncertainty. Statistical methodology provides a natural framework for risk characterization and manipulation with many quantitative ERAs relying heavily on Neyman-Pearson hypothesis testing and other frequentist modes of inference. Bayesian statistical methods are becoming increasingly popular in ERA as they are seen to provide legitimate ways of incorporating subjective belief or expert opinion in the form of prior probability distributions. This article explores some of the concepts, strengths and weaknesses, and difficulties associated with both paradigms. The main points are illustrated with an example of setting a risk-based "trigger" level for uranium concentrations in the Magela Creek catchment of the Northern Territory of Australia.

Key Words: trigger values, Bayesian statistics, natural resource management, statistical inference.

INTRODUCTION

Environmental risk assessment is not new. An early application can be found in the setting of permissible occupational exposure limits for chemicals in the workplace back in the 1930s (Eduljee 2000). However, it was not until much later that the "risk paradigm" was institutionalized and mandated by the environmental protection agencies of the world. There is little doubt that environmentalism of the 1970s achieved a great deal, as noted by Sunstein (2002). However, the "command and control" approach did little to elevate our understanding of the ecosystem and resulted in data rich–information poor agencies that were ill equipped to make more comprehensive and holistic assessments of the environment. The 1980s saw the emergence of risk assessment as a regulatory paradigm, although the ensuing decade was dogged by a lack of agreement on what constituted a risk assessment, a confused lexicon, and inconsistent methodologies. In particular, many quantitative risk assessments were little more than an assignment of subjective probabilities to various adverse outcomes, where the assigned probabilities were manipulated by an oftentimes dubious and concealed calculus. In addition, the terms *hazard* and *risk*

Address correspondence to David R. Fox, Australian Centre for Environmetrics, Parkville Victoria, Australia. E-mail: david.fox@unimelb.edu.au

were often used interchangeably and synonymously or defined mathematically as *risk* = *hazard* × *exposure*. In my view, neither is correct.

The central element of risk is *uncertainty*—it is a probabilistic concept although this is not a shared view. Duckworth (1998), for example, believes that risk is a qualitative term and "is not in itself a measurable quantity and the term should not be used synonymously with probability (p. 10)." He notes that "to 'take a risk' is to allow or cause exposure to the danger (p. 10)." The counter view is that in the absence of uncertainty (about timing and consequences) there is no risk, only defined events having entirely predicable and known consequences. This is consistent with Bridges (2003), who defines risk as "a science-based process for establishing the likelihood of adverse effects (p. 1347)" and Gentile *et al.* (1993), who state "risk assessment is the process for determining the probability, with associated uncertainty, of a particular event occurring as a result of a specific agent or stressor (p. 242)."

During the 1980s, risk assessments became the purview of the technical elite and agencies tended to adopt what has been referred to as the DAD approach—Decide, Announce, and Defend (Kwiatkowski 1998) based (in part) on increasingly technical risk assessments. By the 1990s the emphasis had shifted so that environmental protection was based on more holistic concepts of ecosystem science, whereby a systems understanding was sought that looked at multiple stressors and multiple endpoints, their relationships with each other and their interaction in a bigger landscape. Environmental risk assessment was embedded within this framework, but was no longer an end in itself. In 1992 the USEPA published its environmental risk assessment framework and this was followed by the publication of its environmental risk assessment guidelines in 1996.

Australia has been widely recognized as being at the forefront of development of risk management frameworks (McCarty and Power 2000; Milke 2003). Current thinking and practice is exemplified in The Australia/New Zealand Standard for Risk Management AS/NZS 4360 (Standards Australia 1999) and the ANZECC/ARMCANZ water quality guidelines (ANZECC and ARMCANZ 2000).

In this article we explore some of the statistical aspects of ERA that are both impeding and aiding the development of quantitative risk assessments. We commence with a brief discussion of risk metrics before moving on to consider risk calculus and related statistical methodologies. Finally, with the use of some examples we illustrate the use of Bayesian and frequentist methods for analyzing chronic and acute toxicity data in the context of aquatic ecosystem protection.

RISK METRICS

The U.K. Department of Health (DoH) has assigned narrative terms to various levels of risk associated with death in any year from various causes (DoH 1996). These are reproduced in Table 1. As can be seen, this construct clearly equates "risk" with probability as argued in this article.

The Society of Petroleum Engineers (SPE) has defined "acceptable" environmental risks in terms of the frequency of occurrence for various damage categories (Klovning and Nilsen 1995). These damage categories and risks are shown in Table 2.

The data in Tables 1 and 2 are not directly comparable, although a mapping can be constructed as follows.

Table 1. Risk of death in any year from various causes (DoH 1996).

Term used	Risk estimate	Example
High	>1:100	
Moderate	1:100–1:1,000	Smoking 10 cigarettes/day 1:200
		All natural causes for 40 year old 1:850
Low	1:1000–1:10,000	All kinds of violence and poisoning 1:3,300
		Influenza 1:5,000
		Road accidents 1:8,000*
Very low	1:10,000–1:100,000	Leukemia 1:12 000
		Accidents at home 1:26,000
		Accidents at work 1:43,000
		Homicide 1:100,000
Minimal	1:100,000–1:1,000,000	Railway accident 1:1,000,000
Negligible	1:1,000,000–1:10,000,000	Hit by lightning 1:10,000,000
		Radiation leak from nuclear plant 1:10,000,000

*More recent estimate is 1:16,000.

Let $p = P$[at least one incident in n years] $= 1 - P$[no incidences in n years]. Furthermore, define θ as the probability of an incident in any given year. Then $p = 1 - (1-\theta)^n$ (assuming "incidences" are independent from year to year) and hence $\theta = 1 - (1-p)^{\frac{1}{n}}$ and letting $p = \frac{1}{n}$ we obtain values of θ for various n (Table 3).

The risks in Tables 1 and 3 and their corresponding labels have been plotted on a logarithmic scale for ease of comparison (Figure 1).

From Figure 1 we see the mismatch between the SPE's definition of "acceptable" environmental damage and the DoH scale of risk to humans. For example, the SPE's risk for "serious" environmental damage is about two orders of magnitude greater than the most serious DoH risk category. It is precisely this sort of ambiguity and inconsistency in the application and interpretation of risk metrics that prompted at least one professional society to try to standardize the risk metric.

In his June 1996 presidential address to the Royal Statistical Society, Adrian Smith suggested that the public needed some simple measure of risk to alleviate the irrational behavior associated with individuals' perception of risk. He coined the term "riskometer" and campaigned for the development of a one-dimensional risk scale in a spirit similar to the Fujita scale for tornadoes, the Richter scale for earthquakes, the Beaufort scale for winds, and the decibel scale for sound intensity.

Table 2. Society of Petroleum Engineers 'acceptable' environmental risk.

Damage category	"Acceptable" environmental risk: once every
Minor	10 years
Moderate	40 years
Significant	100 years
Serious	200 years

Table 3. Imputed risk probabilities (probability of incident in any given year).

Damage category	"Acceptable" environmental risk
Minor	0.0105
Moderate	0.000633
Significant	0.0001005
Serious	0.0000251

Statistician Frank Duckworth (co-developer of the Duckworth/Lewis method for cricket scoring) took up the challenge and devised a "risk number" (Duckworth 1998). A central concept to Duckworth's risk number is the concept of a *loss function.* He defines the loss function $L(\cdot)$ to be the expected life shortening expressed as a proportion of normal life expectation. More formally

$$L(y, s, e) = \frac{S(y, s, e)(1 - q)}{E(y, s, 0)}$$

where S is the expected shortening of normal life as a function of the subject's age, y and sex, s for degree of exposure e; $E(\cdot)$ is the expectation of life for a healthy person of the same age and sex but with zero exposure, and q is a quality of life factor for the period of infirmity (death or coma $q = 0$; good health $q = 1$).

The risk metric proposed by Duckworth is based on the "risk magnitude," R, defined as:

$$R = L(y, s, e) \exp(-rt)$$

where t is the time in years before the consequence is felt and r is a discount rate. For risks associated with chronic exposure, Duckworth suggests integrating R over the subject's remaining expected lifetime, that is $R = \int L \exp(-rt)\, dt$. Finally, the "risk number" is obtained by adding eight to the log (base 10) of R (the rationale being that the most unlikely events we would ever contemplate have estimated risks of about 10^{-8} and thus adding 8 to the log of this risk sets the origin of the scale of

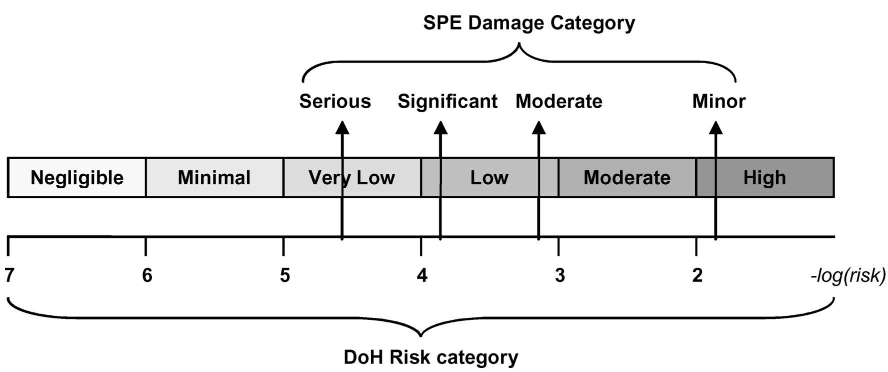

Figure 1. Mapping of Petroleum Industry's damage categories and Department of Health risk categories.

the risk number at zero). Formally, the risk number ($\Re$) is defined as:

$$\Re = 8 + \log_{10} R$$

Duckworth has computed the risk number for a variety of events. These range from $\Re = 0.3$ for a 100-mile rail journey (in the UK) to $\Re = 8.0$ for suicide.

RISK CALCULUS

Environmental risk assessment is about trade-offs. In assessing the risk to the environment posed by a certain activity, there are strong parallels with statistical process control (SPC) methodologies that have been utilized by the manufacturing industries since the 1930s. Fox (2001) refers to "green" and "brown" statistical paradigms to reflect the schism between industrial and environmental statistics, arguing that there should be a greater degree of cross-talk between these two areas. The Australian and New Zealand *Guidelines for Fresh and Marine Water Quality* (ANZECC and ARMCANZ 2000) helped move the Australian water industry further down the risk path and advocated the use of SPC tools such as control charts for water quality monitoring and greater reliance on percentiles rather than averages. Not only are percentiles often more appropriate as indicators of water quality but by definition, they have a simple probabilistic interpretation and are thus potentially more amenable to a risk-analytic approach.

Frequentist Statistics

"Classical" or frequentist statistics is based on the notion of repeated sampling and sequences of infinite realizations of repeatable events. As noted by Root (2003), environmental protection agencies adopt the logic of the courtroom in making environmental assertions, but that "the logic of the courtroom operates under the handicap of working with non-repeatable events."

The word *probability* appears most commonly as a "p-value" in the context of statistical hypothesis tests. The predominant view among scientists is that probability is the quantification of uncertainty. In fact, the p-value of a test is the probability associated with the observed data under the assumption that the null hypothesis is true. If the null hypothesis is true, and if an experiment is repeated many times, the p-value is the proportion of experiments that would give less support to the null than the experiment that was performed.

Null-hypothesis tests are routinely misinterpreted by scientists. Widespread flawed practices have been documented in many disciplines including ecology and medicine (see Anderson *et al.* 2000). Conventional modes of inference are particularly error-prone when Type II errors are costly. For example, conventionally Type II errors are ignored in null-hypothesis tests, implying it is unimportant to detect an impact when in fact, there is one. Large impacts with costly environmental impacts are overlooked. Despite these difficulties, food and drug regulatory authorities, environmental protection agencies, law courts and medical trials all accept null-hypothesis testing as an appropriate method of inference. Reliance on traditional methods of inference leads to logical errors in interpreting data. Environmental applications are particularly error prone. Environmental risk assessments attempt to remediate the situation by applying methods that take into account the chances of incorrectly concluding there are important environmental impacts, and of concluding incorrectly that there is no important impact.

Bayesian Statistics

Increasingly frustrated with purely data analytic approaches to environmental assessment, many natural resource managers are turning to Bayesian methods as this framework is seen to alleviate some of the concerns associated with the binary decision-making process that characterizes classical Neyman-Pearson hypothesis testing. Although the Bayesian approach provides a logical and consistent method for melding prior probabilities with evidence in the form of data, the omnipotent issues concerning choice of priors and parameterization of complex hierarchical models invariably arise (Bier 1999). In addition, Bayesian risk assessments have been the subject of debate and strong criticism as they have been seen to hinder rather than help in courts of law. Much of the concern stems from the misrepresentation of statistical evidence via an error of logic referred to as the "prosecutor's fallacy" whereby the two conditional probabilities (hypothesis given evidence) and (evidence given hypothesis) are confused (Donnelly 1994). Another stumbling block for the Bayesians is the perception that this is a highly technical methodology that is not readily understood by the lay person. As reported in *The Times* (3 November 1997), the London Court of Appeal reaffirmed its position on the role of probability and statistics in assessing weight of evidence cases:

> Introducing Bayes Theorem, or any similar method, into a criminal trial plunges the jury into inappropriate and unnecessary realms of complexity, deflecting them from their proper task"

Although some difficulties in interpretation exist, I argue that these can be overcome through better communication and education. Nevertheless, as we seek to refine existing risk paradigms and develop new ones, there are some clear take-home messages that must be heeded if environmental risk assessment tools are to find a prominent place in the natural resource manager's toolkit.

EXAMPLE: DERIVING RISK-BASED TRIGGERS FOR AQUATIC ECOSYSTEM PROTECTION

In this section we illustrate the use of both "conventional" (frequentist) and Bayesian approaches to the setting of a "trigger" value for uranium concentrations in the Magela Creek in the Northern Territory. Uranium mining in the Magela Creek catchment has been undertaken for more than 20 years. The Department of Environment and Heritage (DEH) used the statistical extrapolation method recommended in the Australian and New Zealand *Guidelines for Fresh and Marine Water Quality* (2000) to obtain a site-specific trigger value (to protect 99% of species) for uranium of 5.8 $\mu g\ L^{-1}$. This value is higher than the historical site-specific guideline value for Magela Creek of 3.8 $\mu g\ L^{-1}$, and is about two orders of magnitude above natural background concentrations (DEH 2001). The data used by the DEH are taken from the 2000–01 Annual Report (DEH 2001) and are reproduced in Table 4.

Table 4. DEH (2001) NOEC data used to derive Uranium trigger concentration.

Species	Test endpoint	NOEC* (μg L^{-1})
Chlorella sp.	Cell division rate	129
Moinodaphnia macleayi	Reproduction	18
Hydra viridissima	Population growth	150
Mogurnda mogurnda	Mortality	400
Melanotaenia splendida inornata	Mortality	810

*NOEC: no-observed-effect concentration.

As can be seen from Table 4, the NOECs range from 18 μg L^{-1} to 810 μg L^{-1}. In deriving the trigger of 5.8 μg L^{-1} the DEH associated all the NOECs in Table 4 with chronic toxicity. Mortality is associated with acute toxicity whereas effects on cell division, reproduction, and growth are associated with chronic toxicity. In order to "standardize" the data, it is conventional practice to apply an acute to chronic ratio prior to analysis (J. Stauber, personal communication). This typically involves dividing the acute mortality data by 10. The computation of trigger values as recommended in the Australian and New Zealand *Guidelines for Fresh and Marine Water Quality* (ANZECC and ARMCANZ 2000) uses a variant of the approach suggested by Aldenberg and Slob (1993). Using the *BurrliOz* software (available at http://www.cmis.csiro.au/Envir/burrlioz/) supplied with the Australian and New Zealand Guidelines with the standardized data of Table 3 (*i.e.*, 129, 18, 150, 40, 81), a value of 3.11 μg L^{-1} is obtained for the 99% trigger value. This is a little over half the value adopted by the DEH and very close to the historical value for Magela Creek of 3.8 μg L^{-1}.

The preceding analysis illustrates some of the difficulties with the derivation of risk-based trigger levels for contaminants in aquatic environments. Not only will

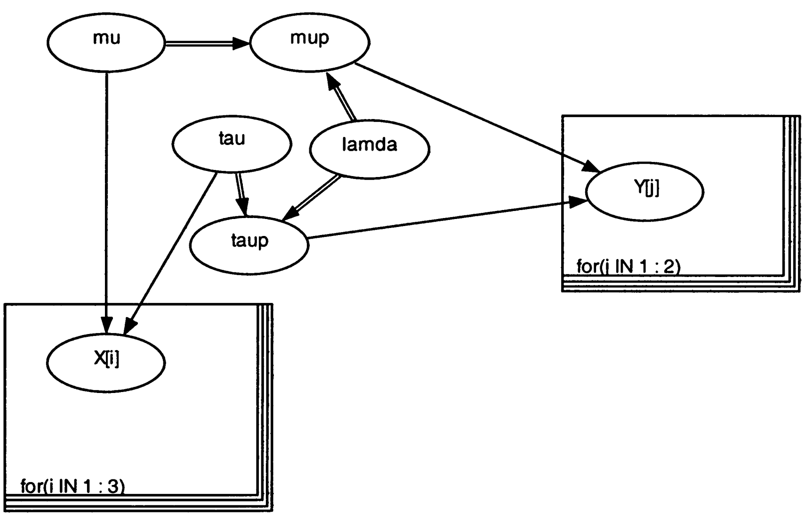

Figure 2. Directed Acyclic Graph for Uranium NOECs example.

Table 5. Summary statistics from posterior distribution of λ.

node	mean	stdev	$P_{2.5}$	median	$P_{97.5}$	sample
lamda	7.324	3.036	4.075	6.624	14.92	50,000

different results be obtained depending on the statistical model employed but the acute to chronic ratio of 10 is quite arbitrary. An alternative approach is to "let the data speak for themselves" so as to find an acute to chronic ratio that maximizes the likelihood of the joint data set. A brief description of the method follows.

Let X denote a chronic NOEC having probability density function (*pdf*) $f_X(x;\theta)$ where θ is a vector of parameters and let Y denote an acute NOEC. It will be assumed that the distribution of Y/λ is the same as the distribution of X, where λ is the acute to chronic ratio. Given a sample of n_1 observations on X and n_2 observations on Y the maximum likelihood estimator (*mle*) for λ is that value that maximizes the likelihood function $L(\lambda) = \prod_{i=1}^{n_1} f_X(x_i;\theta) \prod_{j=1}^{n_2} f_Y(y_j/\lambda;\theta)$. For the data in Table 4, we have $x = \{129, 18, 150\}$ and $y = \{400, 810\}$ with $n_1 = 3$ and $n_2 = 2$. Assuming $f_X(x;\theta)$ is a logistic distribution the *mle* for λ is found to be 7.451. Using $\lambda = 7.451$, the re-scaled uranium data in Table 4 becomes $\{129, 18, 150, 53.68, 108.71\}$ and the revised 99:50 trigger value is estimated to be 5.34 μg L^{-1}.[1]

Bayesian methods provide an alternative mode of inference by allowing us to specify a *prior* distribution forλand then updating this on the basis of the data at hand. The prior distribution may be "non-informative" if we have no particular belief about the likely value of λ or can be chosen to reflect a "best guess." Our model is represented by the directed acyclic graph (DAG) as shown in Figure 2. As before, both X and Y are assumed to follow a logistic distribution.

In Figure 2, X has parameters identified by the stochastic nodes "mu" and "tau" whereas Y's parameters are the stochastic nodes "mup" and "taup" where $mup = mu{\cdot}\lambda$ and $taup = tau/\lambda$.

We have chosen a Gamma(2, 0.1) as the prior distribution for λ. This is a positively skewed distribution that has a mean of 20. Using Gibbs sampling and the *WinBUGS* software tool 50,000 values were generated from the posterior distribution of λ. These were used to obtain summary statistics (Table 5) and an empirical density (Figure 3).

From Table 5 we see that the posterior density for λ has a mean of 7.324 and a median of 6.624. This result agrees well with the maximum likelihood estimate of 7.451. A Bayesian 95% credibility interval for λ is from 4.075 to 14.92 suggesting that the previously assumed "default" value of 10 is plausible (for these data).

Using the estimated median of the posterior distribution of $\lambda = 6.624$ the re-scaled uranium data from Table 4 becomes 129, 18, 150, 60.39, 122.28 and the revised 99:50 trigger value is estimated to be 6.64 μg L^{-1}.

[1]It is acknowledged that the uncertainty in the *estimated* scaling parameter λ has not been accounted for in this analysis. This could be done, although the additional complexity is unlikely to enhance the subsequent interpretation.

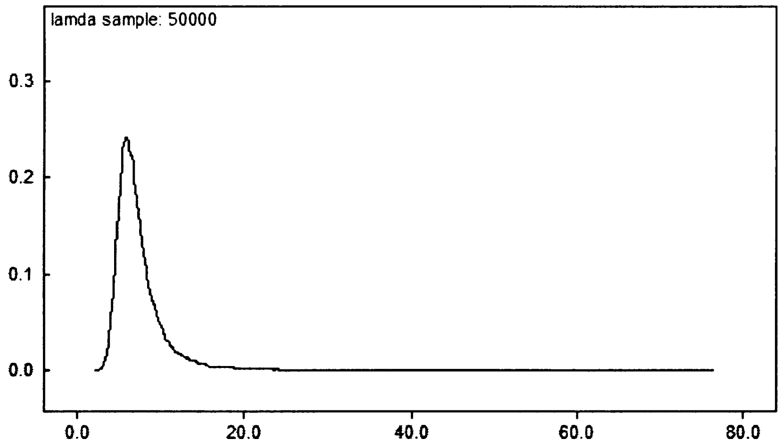

Figure 3. Empirical posterior density for λ based on 50,000 Gibbs samples.

CONCLUSIONS

Ecological risk assessment is an evolving science that attempts to provide a consistent, rational, and scientifically defensible approach to environmental decision-making under uncertainty. Standard tools of (frequentist) modes of estimation and inference provide a natural framework for quantitative ecological risk assessments and although their utility is not questioned, important issues remain unresolved. Of the most pressing is the lack of a universally agreed metric for "risk" and an agreed calculus for assigning and manipulating risk estimates. Bayesian methods of estimation and inference are becoming increasingly popular in ERA due to their inherent ability to introduce subjective belief and/or expert opinion in the form of *prior* probability distributions. Although this is certainly an attractive feature in the context of natural resource management, the omnipotent issues of arbitrariness of choice of prior and parameterization of complex hierarchical models invariably arise.

Some of the advantages and disadvantages of both the frequentist and Bayesian approaches have been illustrated in the context of determining "trigger" values for uranium concentrations in the Magela Creek in the Northern Territory. It has been shown that the resulting trigger level is dependent on both the statistical framework adopted and the method by which acute and chronic toxicity data are combined.

ACKNOWLEDGMENT

The author gratefully acknowledges the helpful suggestions of Ray Correll and Mary Barnes during the preparation of the final version of this article.

REFERENCES

Aldenberg T and Slob W. 1993. Confidence limits for hazardous concentrations based on logistically distributed NOEC toxicity data. Ecotox Environ Safety 25:48–63

Anderson DR, Burnham KP, and Thompson WL. 2000. Null hypothesis testing: problems, prevalence, and an alternative. J Wildlife Management 64:912–23

ANZECC and ARMCANZ. 2000. Australian and New Zealand Guidelines for Fresh and Marine Water Quality. Paper No. 4. Australian and New Zealand Environment and Conservation Council & Agriculture and Resource Management Council of Australia and New Zealand, Canberra, Australia

AS/NZS. 1999. Risk Management. AS/NZS 4360. Standards Australia, Homebush, NSW, Australia

Bier V. 1999. Challenges to the acceptance of probabilistic risk analysis. Risk Anal 19(4):703–10

Bridges J. 2003. Human health and environmental risk assessment: The need for a more harmonised and integrated approach. Chemosphere 52:1347–51

CWQG (Canadian Water Quality Guidelines). 1999. Canadian Water Quality Guidelines Appendix B: Ecological Benchmarks. Available at http://www.tiem.utk.edu/~sada/eco_appendix_b.pdf (accessed 14 January 2005).

DEH. 2001. Supervising Scientist Annual Report 2000–01. Environment Australia. Available at http://www.deh.gov.au/about/annual-report/00-01/ss3environmental.html

DoH. 1996. On the State of the Public Health: Annual Report of the Chief Medical Officer of the Department of Health for the Year 1995. Her Majesty's Stationery Office, London, UK

Donnelly P. 1994. The prosecutor's fallacy. RSS News 22(1):1–2

Duckworth F. 1998. The quantification of risk. RSS News 26(2):10–2

Eduljee GH. 2000. Trends in risk assessment and risk management. Science Total Environ 249:13–23

Fox DR. 2001. Environmental Power Analysis—a new perspective. Environmetrics 12:437–49

Gentile JH, Harwell MA, van der Schalie WH, *et al.* 1993. Ecological risk assessment: a scientific perspective. J Hazardous Materials 35:241–53

Klovning J and Nilsen EF. 1995. Quantitative environmental risk analysis. Society of Petroleum Engineers 30686:461–70

Kwiatkowski RE. 1998. The role of risk assessment and risk management in environmental assessment. Environmetrics 9:587–98

McCarty LS and Power M. 2000. Approaches to developing risk management objectives: an analysis of international strategies. Environ Science Policy 3:311–9

Milke MW. 2003. Improving our ability to manage risks. Waste Management 23(2):iii–iv

Root DH 2003. Bacon, Boole, the EPA and scientific standards. Risk Anal 23(4):663–68

Standards Australia 1999. Risk Management. AS/NZS 4360: 1999. Standards Association of Australia, Strathfield, Australia

Sunstein C. 2002. Risk and Reason. Cambridge University Press, Cambridge, UK

The Times. Monday, November 3, 1997, "Evidence of theorem recipe for confusion; Law report"

USEPA (US Environmental Protection Agency). 1998. Guidelines for Ecological Risk Assessment. EPA/630/R-95/002F. Washington, DC, USA