

Risk Management and Precaution: Insights on the Cautious Use of Evidence

Steve E. Hrudey¹ and William Leiss²

¹Department of Public Health Sciences, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta, Canada; ²Risk Communication and Public Policy, Haskayne School of Business, University of Calgary, Calgary, Alberta, Canada

Risk management, done well, should be inherently precautionary. Adopting an appropriate degree of precaution with respect to feared health and environmental hazards is fundamental to risk management. The real problem is in deciding how precautionary to be in the face of inevitable uncertainties, demanding that we understand the equally inevitable false positives and false negatives from screening evidence. We consider a framework for detection and judgment of evidence of well-characterized hazards, using the concepts of sensitivity, specificity, positive predictive value, and negative predictive value that are well established for medical diagnosis. Our confidence in predicting the likelihood of a true danger inevitably will be poor for rare hazards because of the predominance of false positives; failing to detect a true danger is less likely because false negatives must be rarer than the danger itself. Because most controversial environmental hazards arise infrequently, this truth poses a dilemma for risk management. *Key words:* complacency, false negatives, false positives, futility, positive predictive value, zero risk. *Environ Health Perspect* 111:1577–1581 (2003). doi:10.1289/ehp.6224 available via <http://dx.doi.org/> [Online 10 June 2003]

Risk management has been widely advocated as a rational means for environmental decision making, assisting us in dealing with the wide array of dangers that we face in our uncertain world, ranging from pathogens in drinking water to terrorist attacks. Done well, risk management is inherently precautionary in the sense that it should make use of effective risk assessment to predict, anticipate, and prevent harm, rather than merely reacting when harm arises.

Some of the insights that have supported the movement toward making better use of available evidence for medical decision making, particularly in the field of diagnostic screening, have important but usually overlooked insights for environmental decision making. In particular, the use of four key concepts used for judging the quality of evidence in medical diagnosis—sensitivity, specificity, positive predictive value, and negative predictive value—are relevant to the assessment of environmental hazards, especially those that have low probabilities of occurrence. Applying these concepts rigorously allows us to see more clearly both the value and the limitations of the precautionary approach, as well as to reveal more quantitatively the logical flaw in the notion of “zero risk.” The key question, we suggest, is not whether to be precautionary, but how precautionary we ought to be in specific cases, in relation to the quality of our screening evidence.

Interpreting Evidence about Hazards

Our premise can be illustrated by considering an analogy with airport security. Suppose that we have acquired impressive new scanning technology with the following detection capabilities: *a*) when someone is carrying a

dangerous weapon, 99.5% of the time it will respond positively, and *b*) when someone is not carrying such a weapon, 98% of the time it will respond negatively. If our best intelligence indicates that about 1 in 10,000 passengers screened will be carrying a detectable, dangerous weapon, we can ask how well the screening evidence will allow us to manage this risk. In particular, we can ask, if we get a positive result how likely is that detection to be correct? Given the properties described, common intuition will lead us to expect that this detection should be reliable.

The answer to our question depends on considering that, on average, we will need to screen 9,999 unarmed passengers to find the 1 who is carrying a weapon. The characteristics described provide for a false-positive rate of 2% (98% of the time unarmed passengers will show up as negative). This means that, on average, we will get 199.98 or, effectively, 200 false positives detected for every true positive. Consequently, the answer about how likely it is that a positive detection will be correct turns out to be only 0.5% (1 in 201).

Of course, these numbers are hypothetical, but they likely overestimate both the realistic capability of such technology and the frequency of passengers truly carrying weapons. The frequency of the hazard we are seeking to detect is a critical determinant of the ability of any screening evidence to predict danger accurately.

The basic rationale for the interpretation of screening evidence can be presented as a 2 × 2 table in which the rows relate to the evidence and the columns relate to the reality we seek to understand, but will never know with absolute certainty (Gordis 2000). Figure 1 shows that four quantitative characteristics [sensitivity (Se), specificity (Sp), positive predictive value

(PPV), and negative predictive value (NPV)] are defined by two properties of the screening method (the false-positive rate, α , and the false-negative rate, β) and by a property of the danger being assessed, the frequency of true hazards, $P[H]$, where H is a true hazard.

Considering reality (the columns), Se is the conditional probability, $P[EH|H]$ (where EH is evidence of a hazard), that the evidence will identify a true hazard, given a true hazard:

$$Se = \frac{TP}{TP + FN} = 1 - \beta, \quad [1]$$

where TP is true positive and FN is false negative.

Sp is the conditional probability, $P[EnH|nH]$ (where EnH is evidence of a nonhazard, and nH is a nonhazard), that the evidence will identify no hazard, given a true nonhazard:

$$Sp = \frac{TN}{FP + TN} = 1 - \alpha, \quad [2]$$

where TN is true negative and FP is false positive.

Considering the evidence (the rows), PPV is the conditional probability, $P[H|EH]$, that something is a true hazard, given that the evidence identifies it as a hazard:

$$PPV = \frac{TP}{TP + FP}. \quad [3]$$

NPV (negative predictive value) is the conditional probability, $P[nH|EnH]$, that something is truly no hazard, given that the evidence identifies it as a nonhazard:

$$NPV = \frac{TN}{FN + TN}. \quad [4]$$

Address correspondence to S.E. Hrudey, Department of Public Health Sciences, University of Alberta, 10-102 Clinical Sciences Building, Edmonton, AB, T6G 2G3 Canada. Telephone: (780) 492-6807. Fax: (780) 492-7800. E-mail: steve.hrudey@ualberta.ca

We thank K. Brand, P. Fedorak, S. Newman, S. Rizak, D. Saunders, and A. Senthilselvan for reviewing drafts of this manuscript. This work was inspired by Justice O'Connor's public inquiry into the Walkerton tragedy.

This study was supported by the Natural Sciences and Engineering Research and Social Sciences and Humanities Research Councils of Canada.

The authors declare they have no conflict of interest.

Received 21 January 2003; accepted 10 June 2003.

Some of the foregoing terminology, used widely in various forms in the health and medical sciences, may be confusing to environmental scientists (chemists, biologists, toxicologists, and engineers). Environmental scientists are familiar with sensitivity in a subtly different sense—as the lowest level of a hazard (e.g., a pathogen in drinking water) that can be accurately detected. The distinction is that, in environmental science, sensitivity is typically expressed as a concentration, whereas “sensitivity” as defined in Equation 1 is a conditional probability concerning the ability to detect a hazard when it is truly present.

Likewise, to an environmental scientist, “specificity” refers to the ability of evidence to discriminate a particular hazard (e.g., a particular pathogen) from other factors (e.g., other microbes that may be present). In environmental science, this concept may be used interchangeably with selectivity. As with sensitivity, the environmental science usage is subtly different. In environmental science, “specificity” is the resolution or selective capability of the monitoring method, whereas in medical science “specificity” is the conditional probability of declaring no hazard given that the hazard is truly absent. In this article, we have retained the terminology adopted by the health and medical sciences and seek to show its potential for application to risk management in the environmental sciences.

As Hoffrage et al. (2000) documented, the intuitive application of the conditional probabilities from Figure 1 is still often misinterpreted by medical practitioners. In particular, Se is commonly confused with PPV, which amounts to inverting inherently unequal conditional probabilities ($P[EHI/H] \neq P[H/EH]$). This inequality of conditional probabilities will be recognized by those familiar with Bayesian methods, which provide an alternative approach to deriving these relationships (Gill 2002).

$P[EHI/H]$ or Se is the probability that a hazard will be identified by the evidence, given that there is truly a hazard. $P[H/EH]$ or PPV is the probability that a hazard truly exists, given that one has been identified by the evidence. Clearly, it is PPV that should be considered for interpreting the meaning of evidence, whereas Se is important, as is Sp, for selecting a screening method for collecting evidence. The main factor that creates the asymmetry between these two conditional probabilities is the frequency of true hazards ($P[H]$). In fact, these two conditional probabilities will only be equal for the special case of $P[H] = 0.5$ and $\alpha = \beta$.

Serious problems in the use of these conditional probabilities have also been formally recognized in legal decision making (Balding and Donnelly 1994). In particular, two questions may be asked about DNA evidence as applied to determining guilt of an accused

person: What is the probability that an individual will have a DNA match, given that he or she is innocent? and What is the probability that an individual is innocent, given that there is a DNA match?

As shown above, despite the apparent similarity of these statements, these two conditional probabilities are not equal because one is the inverse of the other in terms of which condition is given. Inverting them by taking the answer to the first question, which the forensic evidence is able to provide, as being the answer to the second question, which the court must decide, has been described as the “prosecutor’s fallacy.” A murder conviction was overturned by the British Court of Appeal because forensic evidence was represented to the jury as if the first conditional probability from the forensic evidence also directly gave the probability for answering the second question about innocence (Balding and Donnelly 1994).

A fact recognized in the health and medical sciences is that, regardless of the apparent capability of a screening method, diminishing returns will appear when screening is applied to relatively rare hazards. For example, universal screening (as opposed to targeting high-risk groups or individuals) for AIDS (acquired immunodeficiency syndrome) would be futile in North America because of the preponderance of false positives that would be generated among the entire population where AIDS is still relatively rare.

The current ongoing health care debate about the value of breast cancer screening programs for women has acknowledged the

important influence of disease incidence on the PPV. For example, Meyer et al. (1990) reported that the PPV for malignant tumors based on mammography screening was 31.5% (with about two false positives for every true positive) for women over 50 years of age, but it was only 8.8% (with about 11 false positives for every true positive) for women under 50 years of age. This difference is mainly driven by the much lower incidence of breast cancer in the younger age group. This limitation is illustrated in Figure 2, which shows the PPV as a function of $P[H]$, for a screening method with an Sp of 99% (i.e., a low false-positive error rate $\alpha = 1\%$) and an Se ranging from only 50% (i.e., a high false-negative error rate, $\beta = 50\%$) up to 99% (i.e., a low false-negative rate $\beta = 1\%$).

Figure 2 (which is derived from spreadsheet calculations using Figure 1 with Equation 3 for PPV and the designated values for α , β , and $P[H]$) shows that PPV becomes increasingly dependent on $P[H]$ for low values, ultimately becoming linearly related for true hazard occurrence frequencies < 0.01 .

$$PPV \approx \frac{P[H]}{\alpha} \times 100\% \text{ as } P[H] \ll \alpha \quad [5]$$

In practical terms, this shows that the PPV, our ability to correctly predict danger with a single source of screening evidence, will inevitably be poor for rare hazards (i.e., for small $P[H]$). Effectively, for PPV to support an expectation of a positive being more likely correct than not ($PPV > 50\%$), the false-positive

		Reality		
		True hazard (H)	True nonhazard (nH)	
Evidence	Indicates a hazard (EH)	True positives (TP) with rate = Se	False positives (FP) with rate = α	[TP + FP]
	Indicates a nonhazard (EnH)	False negatives (FN) with rate = β	True negatives (TN) with rate = Sp	[FN + TN]
		Cases of hazard $P[H] \times n$ = [TP + FN]	Cases of nonhazard $P[nH] \times n$ = $(1 - P[H]) \times n$ = [FP + TN]	$n = [TP + FP + FN + TN]$

Figure 1. Framework for hazard detection and judgment of evidence.

rate, α , must be less than double the likelihood of the hazard, $P[H]$, a specification for a monitoring method that is challenging to provide for very low-frequency hazards. This also shows that the only characteristic of a screening method that has any influence on PPV for low $P[H]$ is the specificity ($Sp = 1 - \alpha$). The problem is that we become overwhelmed by false positives in relation to the relatively rare true positives. An analogy may be made with statistical hypothesis testing, so this can be termed a type I or false-positive error.

Clearly, we must expect to have a large proportion of false positives for any single, practical screening procedure that we use to gain evidence for managing risks from important and well-characterized but rare hazards. For the airport security case described above, we would gladly choose to inconvenience only 200 passengers to catch 1 in 10,000 who is actually carrying a lethal weapon. However, we should know what level of precaution we are applying in any given circumstance so that we can design appropriate follow-up for the positives detected, knowing that they will be mainly false positives. Before resorting to drastic measures, we should apply another test on this smaller sample of initial positives that will now have a higher hazard frequency, $P'[H]$ of 1 in 201, providing a 50-fold improvement from the initial screening, thereby allowing us to achieve a much lower proportion of false positives for the second screen.

Roots of Complacency

Perhaps the high degree of precaution that is inherent to interpreting common monitoring results can help explain the complacency that was displayed by the operators and regulators of the Walkerton, Ontario, Canada, drinking water system, which contributed to a fatal waterborne disease outbreak (Hrudey and Hrudey 2002). In May 2000, 7 individuals died, the youngest a 30-month-old infant, and more than 2,000 became ill after drinking water contaminated with *Escherichia coli* O157:H7 and *Campylobacter jejuni* (O'Connor 2002). A total of 27 people, with a median age

of 4 years, suffered hemolytic uremic syndrome, a serious kidney condition that may carry life-long implications.

The shallow groundwater source had not been adequately disinfected to cope with manure contamination from a nearby farm during a severe rainstorm. Moreover, that system had a long history of adverse microbial monitoring results on treated water, but these result did not trigger necessary improvements in disinfection or better source protection. The evidence of problems included recurring positive results for the indicator organisms (total coliforms, fecal coliforms, and *E. coli*). Justice Dennis R. O'Connor (2002) concluded that this tragedy should not be blamed solely on the operators, despite their many documented failings and misdeeds, since abundant regulatory failures had allowed these negative conditions to persist. This case is important because there is a risk of complacency developing among personnel in other sectors, from airport security to emergency response. They will predictably experience a preponderance of false positives in performing their routine screening responsibilities dealing with rare hazards. Complacency may be less of a concern in environmental circumstances where there is uncertainty about whether harm has happened or where feedback about the lack of harm is not sufficiently immediate to promote complacency.

Positive results for microbial indicators in drinking water most likely do not signal the presence of an infective dose of viable pathogens making an outbreak imminent, unless there is other evidence of contamination (Nwachuku et al. 2002). Rather, positive results for indicators in treated water signal ineffectiveness of the disinfection process that must inactivate bacterial pathogens because they will intermittently challenge the system. Positive results may reflect sampling or analytical errors as well. There will be many false positives in which a positive result for indicator organisms will not coincide with the presence of any viable pathogens, let alone an infective dose. Hence, operators will find by experience, much more

often than not, that the presence of indicator organisms has no immediate health consequences to the community if the operators choose to ignore adverse monitoring results.

Thus, a predominance of false-positive results from tests can foster complacency among the operators. They must understand the reality that false positives will be the norm in such types of environmental screening. Of course, they must also be convinced that effective responses are still necessary to avoid the infrequent but devastating disasters. Competent personnel equipped with the truth will be better able to respond effectively than those who are kept in the dark and told that disaster is likely to strike with any positive result.

Complacency is only one of the most obvious unintended and indirect effects that may arise from any risk management decision. Hofstetter et al. (2002) elaborated on the ripple effect metaphor, first proposed by Graham and Wiener (1995), about the many other effects that arise from an individual risk management decision. In the water safety example, ripple effects from actions such as public notification of adverse water quality results with no explanation of their meaning, or issuance of advisories to boil water when pathogens are unlikely to be present will include undermining consumer confidence in water safety and possibly encouraging some to seek alternative, but less safe, water supplies.

Exercising Caution and the Precautionary Principle

The concern expressed by advocates of the precautionary principle is about the failure to detect a problem, the chance of allowing a false negative (a type II error). Figure 3 demonstrates that hazard frequency has a similar impact, whereby even a very low sensitivity (i.e., a high false-negative rate, $\beta = 50\%$) still allows an apparently precautionary NPV ($> 99.5\%$) for those common circumstances when $P[H]$ is low (< 0.01). An NPV $> 99.5\%$ means that a hazard would truly be present $< 0.5\%$ of the time when the monitoring test indicated that there was no hazard. The reason for this direct effect is evident when considering the distribution of values in Figure 1; a low $P[H]$ corresponds to few occurrences in the left column. Even a large false-negative rate, β , will be applied to only a small number of true hazards, making the failure to detect any true hazard a rare occurrence among the total number of cases tested.

Of course, this will provide small comfort if extremely severe consequences arise from the failure to detect a rare true hazard. For example, Lewis (1996) estimated that about every million years we will experience an object impact from outer space capable of causing a 1-million-megaton explosion that would result in more than a billion casualties, if not threaten

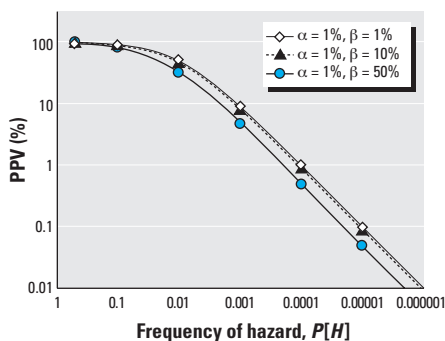


Figure 2. Quantitative dominance of PPV (%) by low frequency of hazard occurrence, $P[H]$.

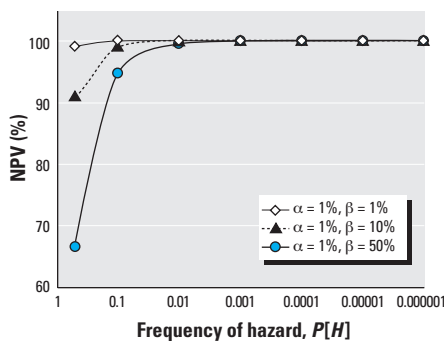


Figure 3. Quantitative dominance of NPV (%) by the low frequency of hazard occurrence, $P[H]$.

human extinction. We would want to avoid a false-negative error in predicting such an event, even if we have not yet developed any means of avoiding it. Thus, circumstances where there are severe consequences from failing to detect even a rare hazard provide the strongest rationale for invoking a substantial degree of precaution. Presumably, this is why the most widely cited version of the precautionary principle has stressed its application for those situations where serious or irreversible consequences are likely to arise unless precautionary action is taken (United Nations 1992).

These valid concerns about precaution may be better perceived by considering the chance of a false-negative error occurring when we get a negative screening result ($1 - \text{NPV}$). That chance is illustrated in Figure 4 as a function of $P[H]$. As with the influence of $P[H]$ on PPV, Figure 4 shows, as we should expect, that the chance of a false-negative error declines in direct proportion to $P[H]$, becoming a very small chance for very rare hazards, essentially equal to $\beta \times P[H]$, because the possibility of a false-negative error cannot exceed the chance of the hazard. Applying this analysis to the previous airport screening example ($\beta = 0.5\%$), the chance of missing a passenger with a weapon would be 1 out of 2 million negative results when 1 in 10,000 passengers is carrying a weapon (i.e., 0.5% of $P[H]$).

Issues surrounding precaution can be elaborated by considering a public health analogy using two drinking-water incidents. The Sydney, Australia, water crisis in 1998 has been described as a case of issuing an advisory for Sydney residents to boil water on the basis of erroneous monitoring results (Clancy 2000), what amounts to a false-positive (type I) error. Increased surveillance and targeted telephone surveys revealed no consistent evidence of any increase in diarrheal disease, nor any increase in laboratory-diagnosed cases of cryptosporidiosis, despite the apparent detection of massive numbers of *Cryptosporidium* oocysts in the treated water supplied to 3 million consumers (McClellan 1998). Recently, almost 5 years after the episode, a rebuttal has been published arguing that water contamination did occur (Cox et al. 2003). However, all parties agree that excess illness was absent in the community, and because there are some serious inconsistencies in the case for these latest claims, we accept Clancy's account (Clancy 2000).

Walkerton, on the other hand, was a case of false-negative (type II) error. Warnings were ignored for more than 20 years and tragedy resulted. Given this comparison, precaution in public health demands that we should prefer type I (false positive) over type II (false negative) errors because the consequences of the latter are usually more direct and potentially more severe. However, we also need to recognize that there are usually consequences to type I errors

as well (i.e., the ripples). In the Sydney case, several million dollars of public funds were spent on circumstances where public health was apparently not endangered. Frequent type I errors will create a "cry wolf" response with the public such that important measures such as advisories to boil water may be ignored when they are truly needed.

A precautionary approach for public health appears to be opposite to the conventional scientific bias in hypothesis testing that critics have claimed favors type II over type I errors (Cranor 1993; M'Gonigle et al. 1994). However, contrary to those criticisms that do apply when either outcome is equally likely, for situations where $P[H]$ is < 0.33 (i.e., most environmental health cases), accepting the conventional scientific levels for α (5%) and β (20%) does not create any probability bias favoring detection of false positives compared with false negatives. For $P[H] = 0.33$, 11% of positives will be false and 10% of negatives will be false. For progressively lower $P[H]$, the likelihood of detecting false positives increases while that of detecting false negatives decreases, so that we move toward a cautionary bias in probability criteria for less frequent hazards regardless of what values we may choose for α in relation to β .

We agree that the potentially greater severity of public health problems caused by false negatives, as compared with false positives, suggests that we should strive to avoid the former more than the latter. For hazards with severe consequences, we will want to focus on the chances of a false negative occurring (i.e., $1 - \text{NPV}$; Figure 4). Regardless, it is clear that our conventional risk management framework becomes progressively and inherently more cautious by generating many more false positives than false negatives as it is applied to rarer hazards. This reality places a premium on judgment for interpreting the evidence in these cases, a need that demands the fullest possible understanding of these relationships that govern the nature of evidence.

Where the precautionary principle may be most relevant is for a problem that is so poorly understood that there is no prospect for confidently assessing it within this hazard-screening framework. In such cases where there is overriding conceptual or epistemic uncertainty (Walker 1991), the greatest danger may be the chance of a type III error (Kendall 1957). This error can be described as misunderstanding a problem so completely that we may find the right answer to the wrong problem. A type III error is likely for an event that has never happened before.

Returning to our airport security example, a type III error arises if the primary hazard we seek to detect comes from passengers carrying plastic explosives that are undetectable by our screening technology designed for metallic

weapons. In these uncertain circumstances, if the stakes are sufficiently high for inaction, the precautionary principle may provide appropriate guidance toward taking some additional action. In such cases, there must be an obligation to link the application of the precautionary principle to a commitment for research to better define and understand the problem (Goldstein 1999): If a hazard is important enough to invoke precaution as a justification for priority action, it must also be important enough to understand better.

Overall, to better define any hazard-screening program, we should seek enough understanding to allow us to estimate the false-positive rate (α), the false-negative rate (β), and the likely frequency of the hazard ($P[H]$). This conclusion may appear to be circular in that the likelihood of the hazard must be estimated before analyzing how precautionary we are likely to be. This is more useful than it may appear at first glance. The need to estimate hazard frequency in order to initiate this analysis simply illustrates the inescapable requirement for following an iterative approach to evaluate complex problems, reinforces the importance of requiring evidence from more than one source, and supports the merits of Bayesian logic as a means to make rational use of available evidence (Gill 2002).

Serious problems, involving high stakes, demand sequential analyses to acquire better evidence often within the constraints of the need for rapid answers. Just as surgeons do not pursue high-risk surgery in search of a brain tumor based only on initial screening evidence, we will expect to gather improved sequential evidence when high-stakes outcomes are suggested by initial positive results. Likewise, parallel evidence consistent with hazardous conditions will support an expectation of higher $P[H]$ than we would expect in the

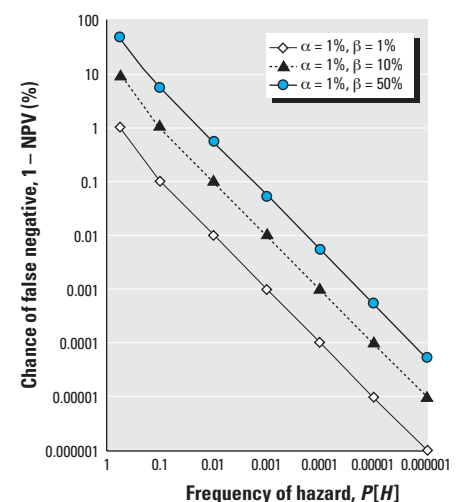


Figure 4. Quantitative dominance of the chance of a false-negative error (%) by the low frequency of hazard occurrence, $P[H]$.

absence of supporting evidence. The established public health practice of targeting higher risk groups for screening programs provides an analogy of the effective use of parallel evidence to improve PPV.

Caution versus Futility

Finally, we may use all of the above reasoning to turn the problem of precaution on its head, so to speak. What we have sought to show (by calling attention to the dominance of the frequency of hazard over inherent error rates for low-frequency hazards) is that our conventional practices of risk management will encounter a dominance of false positives, even when they rely on valid screening evidence in the case of searching for rare hazards. This outcome is inevitable because the screening method (even with a low false-positive rate) must be applied to an overwhelming proportion of nonhazards in search of a rare hazard. The result may be appropriate in many cases, if follow-up actions to positives detected are measured and suited to the circumstances so that the combined consequences of frequent false positives do not exceed the consequences of the rare hazard that we seek to avoid.

Poor risk management arises from an inadequate level of precaution, but, paradoxically, seeking an excessively high and ultimately self-defeating level of precaution for a narrowly defined rare hazard will also provide poor risk management. This occurs because of the diminishing returns achieved in pursuing the illusion of “zero risk,” a pursuit that others have critiqued for various related and differing reasons (Cross 1996; Graham and Wiener 1995).

As we pass the point of appropriate precaution, motivated by a search to detect hazards at lower and lower levels of likelihood, this “iron law” of increasing false positives defeats us. Below a certain low level of hazard frequency, we simply cannot have a reliable idea of whether what we fear is actually there or not, unless we have resources and knowledge to pursue a series of increasingly effective sequential tests to provide meaningful evidence on extremely small risks. For most environmental health issues, our limited toolbox of valid sequential tests will be rapidly depleted. Because

the interventions we devise (to combat diseases, for example) have their own risks, the wisest course of action is to avoid trying to be more precautionary than our knowledge enables us to be. In total, our analysis amounts to a mathematical elaboration of the inevitability of risk trade-offs and the compelling need for good judgment to deal sensibly with those trade-offs.

Conclusions

As we have sought to show, our current best practices for the management of risks from well-characterized low-frequency hazards will have an inevitable dominance of false positives over true positives and false negatives. This implies an inherent degree of substantial precaution, and thus practicing good risk management cannot mean deciding whether to be precautionary or not. Rather, the critical question always is, how precautionary should we be in any particular case?

We know from experience that in many cases (such as occupational exposure to asbestos or sustaining the Atlantic cod fishery) we have been less precautionary than we ought to have been. We hope that the contemporary emphasis on the precautionary principle may lessen the chance of this occurring in the future. But we have also sought to show that when we are dealing with well-characterized hazards, we sometimes unwittingly want to be more precautionary than it is possible to be, ensuring a self-defeating outcome.

Certainly, if we allow false-positive errors to reign without understanding what is happening, we risk fostering complacency to an extent that we may ultimately be ensuring eventual catastrophe through neglect. Risk management needs to maintain a healthy tension based on considering the likelihood and consequences of both false-positive and false-negative errors, seeking an appropriate balance between these opposite outcomes, rather than zealously seeking the absolute elimination of false-negative errors in a futile search for zero risk. The increasingly serious and complex challenges that we currently face on our planet demand that we find better ways to make more tangible, informed, and effective use of our knowledge, not less.

REFERENCES

- Balding DJ, Donnelly P. 1994. How convincing is DNA evidence? *Nature* 368:285–286.
- Clancy JL. 2000. Sydney's 1998 water quality crisis. *J Am Water Works Assoc* 92(3):55–66.
- Cox P, Fisher I, Kastl G, Jegatheesan, V, Warnecke M, Angles M, et al. 2003. Sydney 1998 – lessons from a drinking water crisis. *J Am Water Works Assoc* 95(5):147–161.
- Cranor CF. 1993. *Regulating Toxic Substances: A Philosophy of Science and the Law*. New York:Oxford University Press.
- Cross FB. 1996. Paradoxical perils of the precautionary principle. *Washington Lee Law Rev* 53:851–925.
- Gill J. 2002. *Bayesian Methods—A Social and Behavioural Sciences Approach*. Boca Raton, FL:Chapman and Hall/CRC.
- Goldstein BD. 1999. The precautionary principle and scientific research are not antithetical [Editorial]. *Environ Health Perspect* 107:A594–A595.
- Gordis L. 2000. *Epidemiology*. 2nd ed. Philadelphia:WB Saunders.
- Graham JD, Wiener JB, eds. 1995. *Risk Versus Risk: Tradeoffs in Protecting Health and the Environment*. Cambridge, MA:Harvard University Press.
- Hoffrage U, Lindsay S, Hertwig R, Gigerenzer G. 2000. Communicating statistical information. *Science* 290:2261–2262.
- Hofstetter P, Bare JC, Hammit JK, Murphy PA, Rice GE. 2002. Tools for comparative analysis of alternatives: competing or complementary perspectives? *Risk Anal* 22:833–851.
- Hrudey SE, Hrudey EJ. 2002. Walkerton and North Battleford—key lessons for public health professionals. *Can J Public Health* 93:332–333.
- Kendall AW. 1957. Errors of the third kind in statistical consulting. *J Am Stat Assoc* 52:133–142.
- Lewis JS. 1996. *Rain of Iron and Ice—The Very Real Threat of Comet and Asteroid Bombardment*. Reading, MA:Addison-Wesley.
- McClellan P. 1998. *Sydney Water Inquiry*. 3rd Report. Sydney, Australia:New South Wales Premier's Department.
- M'Gonigle RM, Jamieson TL, McAllister MK, Peterman RM. 1994. Taking uncertainty seriously: from permissive regulation to preventative design in environmental decision making. *Osgoode Hall Law J* 32:99–169.
- Meyer JE, Eberlein TJ, Stomper PC, Sonnenfeld MR. 1990. Biopsy of occult breast lesions. Analysis of 1261 abnormalities. *JAMA* 263:2341–2343.
- Nwachuku N, Craun GF, Calderon RL. 2002. How effective is the TCR in assessing outbreak vulnerability? *J Am Water Works Assoc* 94(9):88–96.
- O'Connor DR. 2002. Part 1. Report of the Walkerton Inquiry: The Events of May 2000 and Related Issues. Toronto, Ontario, Canada: The Attorney General of Ontario. Available: <http://www.attorneygeneral.jus.gov.on.ca/english/about/pubs/walkerton/part1/> [accessed 31 July 2003].
- United Nations. 1992. *Rio Declaration on Environment and Development*. U.N. Doc./Conf.151/5/Rev.1. New York:United Nations. Available: <http://www.un.org/documents/ga/conf151/aconf15126-1annex1.htm> [accessed 1 August 2003].
- Walker VR. 1991. The sirens songs of science: toward a taxonomy of scientific uncertainty for decisionmakers. *Conn Law Review* 23:567–626.