International Journal of Occupational Medicine and Environmental Health, 2004; 17(1): 35—41

ON BEING CAREFUL WHAT WE WISH FOR: SOME DIFFICULTIES WITH OPERATIONALIZING THE PRECAUTIONARY PRINCIPLE

DAVID OZONOFF

Department of Environment Health
Boston University School of Public Health
Boston, MA, USA

Abstract. An important part of the Precautionary Principle is that taking action is justified for protecting public health even when there is some scientific uncertainty. We examine here the two components of this central feature of the precautionary principle, scientific uncertainty and decision making. In order to operationalize the principle we should examine the consequences of its decision rules and how they perform under various conditions. The performance of decision rules in disease screening is measured by the sensitivity and specificity of the rule, but the consequences for the patient are given by the positive and negative predictive values, determined not only by the performance of the rule but the prevalence of the disease in the population. We look at positive and negative predictive value of the Precautionary Principle from the standopoint of the costs to one or other parts of society, show that the usual rule which tends to maximize sensitivity in favor of specificity may have unexpected consequences, and demonstrate that it is sometimes possible to trade sensitivity and specificity off against each other in a way that improves both positive and negative predictive value, or worse, degrades both.

Key words:
Precautionary Principle, Disease screening, Sensitivity, Specificity, Predictive value, False positives, false negatives

INTRODUCTION

In the very first chapter of Raffensberger and Tickner's volume on the Precautionary Principle, Jordan and O'Riordan call attention to the fact that the Precautionary Principle, rather than being a coherent and internally consistent set of principles, is more an expression of a general distrust by consumers of private forces and their public sector enablers who act in ways independent of (and often inimical to) the well-being of the general public, the environment and the biosphere [1]. From the complicated mélange that is the Precautionary Principle, Jordan and O'Riordan extract a small number of common themes, among the most important of which are a willingness to take action in the face of continuing scientific uncertainty, and a belief that "in the long run" the precautionary approach provides more benefits than costs if non-monetized benefits are appropriately counted. These themes involve, among other things, making decisions in the face of uncertainty; and balancing risks and benefits within a specific context and orientation.

Leaving the Precautionary Principle at the level of a vague notion or slogan is unsatisfactory to many and there are persistent attempts to "operationalize" it. This involves yet another trade-off: trading off vagueness for

This work was partially supported by grant 2P42 ESO7381 from the National Institute of Environmental Health Sciences (NIEHS) with funds from the US Environmental Protection Agency (EPA).
Its contents are solely the author's responsibility and do not necessarily represent the official views of the NIEHS or the EPA.
Received: January 19, 2004. Accepted: January 30, 2004.
Address reprint requests to Prof. D. Ozonoff, Department of Environmental Health, Boston University School of Public Health, 715 Albany Street, Boston, MA 02118, USA (e-mail: dozonoff@bu.edu).

definiteness. It means, in essence, we must set out some basic strategies for transparent decision-making, i.e., instructions on how to construct explicit decision-rules in various circumstances to allow informed participation and accountable outcomes.

RISK-BENEFIT AND THE MEDICAL SCREENING ANALOGY

We present here a useful analogy between decision making using the Precautionary Principle and medical screening for disease. The task in screening for disease is to make a decision, based usually on a single test about a diagnosis that normally requires multiple tests. It is therefore a decision made under an uncertainty born of efficiency and economy. In screening for disease, positive outcomes of a screen are usually followed by increasingly costly and invasive confirmatory tests because a screening test is usually imperfect. In the same way, the task for the Precautionary Principle is to screen out, in the face of incomplete information, technologies or developments that might lead to serious consequences for health or the environment. The confirmatory tests of the disease screening analogy are the shifted burden on the producer of the technology to show its safety before proceeding.

There are many issues associated with these decisions, such as who participates, who benefits and who pays the costs. Leaving these aside, we draw attention to some simple logical consequences of the screening model that are unsettling but forced on us to the extent that the screening model is a faithful reflection of the decisions made when operationalizing the Precautionary Principle.

MEASURES OF ADEQUACY

Our argument involves an arbitrary decision rule rather than any specific rule, and therefore applies to any and all decision rules, not just particular ones.

We want any decision rule to perform well. This means we want to be right about what we decide not to do (e.g., move away from nuclear power) and right about what we decide it is acceptable to do (e.g., develop solar power, fuel cells or wind power). In the language of disease screening, we want high sensitivity and high specificity.

As a reminder, in a screening test specificity is the proportion of non-diseased individuals correctly classified as non-diseased by our test, while sensitivity is the proportion of truly diseased picked up or detected by our test. If we were to cross-classify true disease state with test outcome we would get a two by two table, with the right answers in the diagonal boxes and the false positives and false negatives in the off diagonal boxes:

	Positive test	Negative test	
Diseased	A	B	A + B
Healthy	C	D	C + D
	A + C	B + D	A + B + C + D

The sensitivity of the test is A/(A + B) and its specificity is D/(C + D). From the patient's perspective, however, this is not the important information. Patients want to know the chances they have the disease if they come up positive in the test (or less commonly, the chances they are free of the disease if they come up negative). These are the proportions A/(A + C) and D/(B + D). These patient-centered values are called the predictive values of a positive or negative test. To determine them one needs to know more than the performance of the test itself as measured by its sensitivity and specificity. One also needs to know something about the population you are performing the test on, in particular, the prevalence of the condition you are screening for. This is frequently not known.

We can express these measures in terms of a decision-rule for some implementation of the Precautionary Principle, i.e., the criteria which tell you whether to invoke the Precautionary Principle and not implement a technology, or decline to invoke it and let a technology proceed. The positive predictive value of the decision-rule tells you the probability you did the right thing if you stop a technology. The burden falls on the producer while any possible benefit goes to consumers. One minus the positive predictive value represents the chances you stopped something from going forward that in reality did not pose a serious threat to health or the

Table 1. General set-up of Precautionary Principle decison

		Precautionary Principle decision		
		Don't go forward	Allow to go forward or implement	
"Truth"	Poses true hazard	s_1pN	$(1 - s_1)pN$	pN
	Not true hazard	$(1 - s_2)(1 - p)N$	$s_2(1 - p)N$	$(1 - p)N$
		$s_1pN + (1 - s_2)(1 - p)N$	$(1 - s_1)pN + s_2(1 - p)N$	N

environment. These are the false positives, representing a cost both to producers and consumers. The negative predictive value represents the probability that you correctly let something go forward when it posed no extra risks. This represents a benefit to both producer and consumer, while one minus this value, the false negatives, is the case where you let something go forward that ultimately turned out badly. This may be a benefit to the producer (at least in the short term) but a cost (possibly a huge one) to consumers and the environment. These are the "late lessons from early warnings" [2]. Characterizing "costs" and "benefits" in this fashion is simplistic, but corresponds to the qualitative nature of the burdens, suggesting if you are to maximize benefits to consumers, then, the strategy is to maximize both the positive and negative predictive values. If you have a choice, you might choose to focus on maximizing negative predictive value since that measures the probability that our decision to let a technology be implemented was correct. Sensitivity and specificity are given by the rule. The value of p, the prevalence, is exogenous to the rule. Together they determine the outcome.

We want to set our "threshold" for rejection at some socially appropriate place. The precautionary point of view is to opt for high sensitivity in our decision procedure, tending toward the "better safe than sorry" end of the spectrum. Indeed, for many this is the practical embodiment of the Precautionary Principle itself. For any specific decision rule, however, there is a trade-off between sensitivity and specificity: as we increase one we tend to decrease the other. As it will turn out, and this was quite surprising, the form this trade-off takes is of critical importance.

It is most efficient to solve this problem in full generality, rather than exhibiting numerical examples. We call the prevalence of policies or technologies that will turn out to be truly bad down the line, p. We don't know what that proportion is. Clearly it is not likely to be anywhere near 100% unless we have specially selected technologies to produce an unusually high yield. Then the selection rule itself would have been quite an accurate one and we should be using it as part of our decision rule. We should be including the full range of alternatives such as solar power, wind power or fuel cells as well as nuclear power or cell phone towers [3]. Indeed, widening the options with other alternatives is an important benefit of the Precautionary Principle. When the full range of alternatives is considered it is reasonable to assume that the policies or technologies that will turn out to be truly problematic are relatively small as a proportion of the whole.

The general set-up to analyze this is given in Table 1, where s_1 and s_2 are the sensitivity and specificity, respectively, of the decision rule, while p is the proportion or prevalence of truly bad policies or technologies from the Precautionary Principle point of view. N represents the total number of decisions, i.e., the number of technologies falling under the view of the Precautionary Principle.

We can use this general scheme immediately to calculate the positive and negative predictive values of the Precautionary Principle decisions.

$$PV+ = \frac{s_1 pN}{s_1 pN + (1 - s_2)(1 - p)N} = \frac{s_1}{s_1 + (1 - s_2)\dfrac{(1 - p)}{p}}$$

$$PV- = \frac{s_2 N}{s_2 N + (1 - s_1)\dfrac{p}{(1 - p)}JN} = \frac{s_2}{s_2 + (1 - s_1)\dfrac{p}{(1 - p)}}$$

We see the number of decisions, N, cancels out in the numerator and denominators and is irrelevant.

We can see what happens when these values change. Consider the following case:

$$p = .8, \qquad s_1 = 90\%, \qquad s_2 = 90\%$$

Here we are making our decisions about policies or technologies that are highly suspicious, i.e., enriched in terms of a high percentage that really need to be halted (p = 80%). Moreover we have a decision rule that is very accurate, correctly picking out the truly harmful cases 90% of the time and leaving the more benign ones alone 90% of the time. We calculate the predictive value of a positive decision (halt the technology), PV+, and the predictive value of negative decision (let it go forward), PV-:

PV+ = 97% PV- = 69%

For this high positive predictive value, the cost of correct decisions to invoke the Precautionary Principle falls on the producer while consumers benefit. On the other hand, the negative predictive value tells us there is a high probability that you correctly let something go forward 69% of the time when it posed no extra risks, benefiting both producer and consumer, but there was still an appreciable chance (31%) that you let something go forward that ultimately turned out badly, possibly very badly.

"p," the prevalence, enters into these computations. Fig. 1 is a plot of positive and negative predictive values as the prevalence changes from very low (1%, on the left) to very high (99%, on the right).

Notice that for very high prevalence or very low prevalence one or the other of the positive or negative predictive values is low while the other is high. It is important to realize that in each instance exactly the same decision process is being used, i.e., our thresholds do not change. Only the questions, that is, the list of technologies it screens, are different. For 80% prevalence we are missing 31% of cases we really do not want to miss. What can we do about that? We might think that we could boost

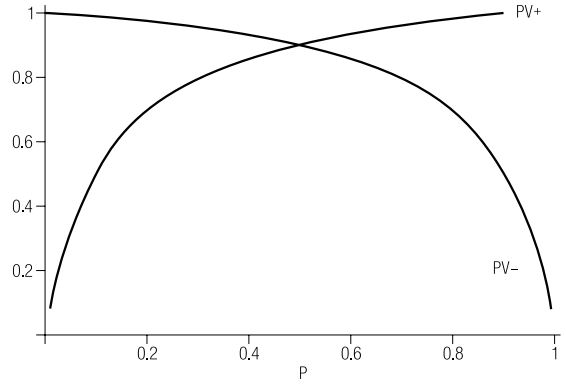

Fig. 1. Trade offs between PV+ and PV- as prevalence (horizontal axis) increases for a test with 90% sensitivity and 90% specificity.

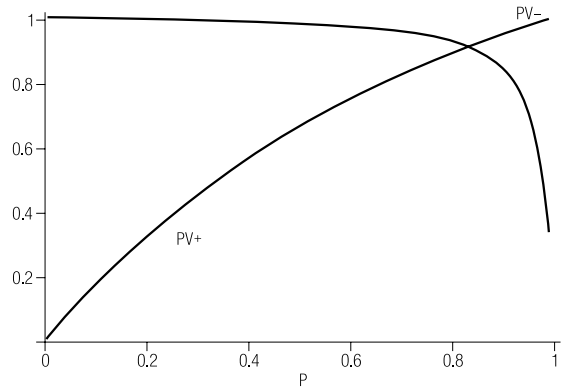

Fig. 2. Trade offs between PV+ and PV- as prevalence (horizontal axis) increases for a test with 99% sensitivity and 50% specificity.

the sensitivity by changing our "threshold." When we do this, however, we also lower the specificity, the amount of lowering depending upon the specifics of the decision rule. Fig. 2 shows what happens to the positive and negative predictive values when we lower specificity to 50% by raising sensitivity to 99%:

There is high negative predictive value until the prevalence rises to about 80% but then it drops precipitously. The positive predictive value, however, exhibits a steady, almost

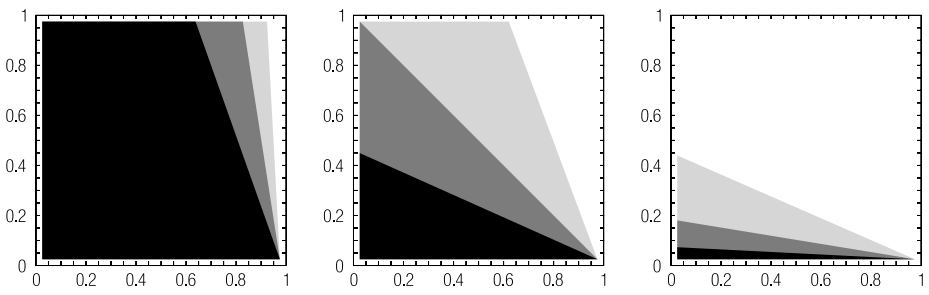

Fig. 3. Contour plots, in quintiles, of specificity (vertical axis) vs. sensitivity (horizontal axis) for three prevalences: 80% (left), 40% (middle) and 10% (right). White areas are highest quintile (>80%).

linear increase, with increasing prevalence. Again, unless you are lucky enough to be near the "cross-over" point you will do badly on one of the predictive value measures as prevalence varies. In Fig. 3 we see contour plots, in quintiles, of specificity (vertical axis) versus sensitivity (horizontal axis) for three prevalences: 80% (left), 40% (middle) and 10% (right). The white areas are negative predictive values above 80%. We want to be in the white area.

Here we see that predictive value is a measure of benefit to producers and consumers both, while its complement is a long run cost to everyone and only a short term benefit to producers. Good positive predictive value is a cost to producers but is a benefit to consumers, while poor positive predictive value is a cost to producers and a potential opportunity cost to consumers. From the Precautionary Principle point of view, then, we would sacrifice positive predictive value for negative predictive value.

Notice that for high prevalence situations (left) the determining factor is sensitivity. Even modest drops cause substantial drops in negative predictive value. In such a high prevalence list we could adopt a simple decision rule: stop most things on the list. This gives us close to 100% sensitivity. The opposite is true for low prevalence situations (right). Moreover, under low prevalence most combinations of sensitivity and specificity give us high negative predictive value. In the case of low prevalence lists we should opt for rules with higher specificity, since this almost guarantees us good negative predictive value whereas maximizing sensitivity helps us less. This is counter-intuitive if one assumes the Precautionary Principle should always prefer being "better safe than sorry."

Unfortunately we do not know the prevalence of bad technologies on the list, but we are helped if that prevalence is low, at least as far as negative predictive value is concerned. This is another argument for widening the list of considered alternatives. Positive predictive value, however, will suffer. Our dilemma is that if we start with a completely unselected list we have no *a priori* way to know the prevalence of bad technologies, and the desirability of the actual operation of any decision rule depends upon the prevalence. Once we start to make selected lists we are incorporating into the selection process part of the decision rule. There is one additional feature of selecting decision rules and changing thresholds that is of interest. For any particular rule, we are usually able to adjust "thresholds" thereby increasing sensitivity, but that has the effect of decreasing specificity, or at best, leaving it unchanged. The only way to increase both at the same time is to adopt a different rule. In addition, the relationship of s_1 to s_2 as one changes is also a feature of a specific rule. The exact nature of this co-variation can have a dramatic effect on negative predictive value. Consider the following three generic examples of inverse relationship of specificity with sensitivity (Fig. 4).

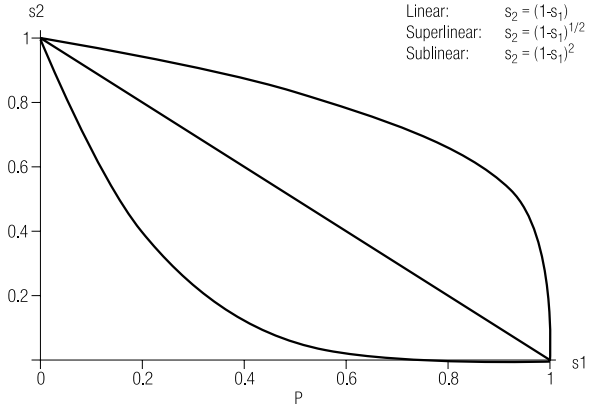

Fig. 4. Supralinear, linear and sublinear relationships between s_1 and s_2.

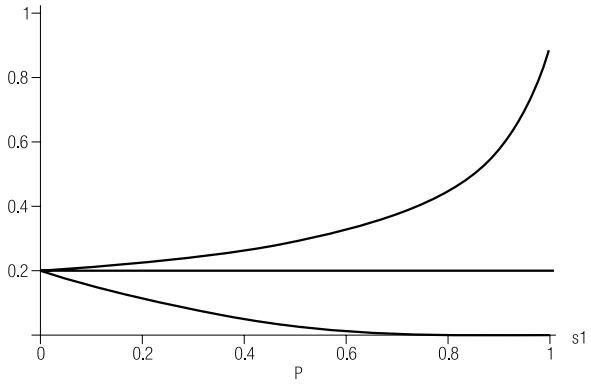

Fig. 5. What happens to PV- as the decision process is changed to increase s_1 when s_2 decreases linearly (flat line), supralinearly (the increasing line), or sublinearly (the decreasing line).

In all three instances s_2 decreases as s_1 increases (and vice versa). But the effect on PV- is different in each case (Fig. 5).

What is happening is that the exact way sensitivity and specificity trade off against each other is critical to the

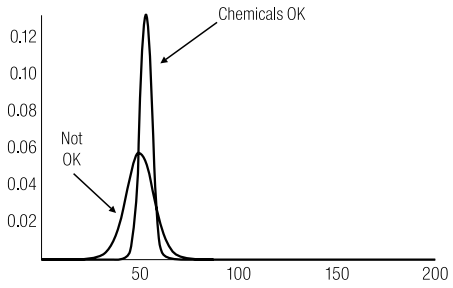

Fig. 6. Distribution of "scores" for chemicals that are truly OK vs. those that are not.

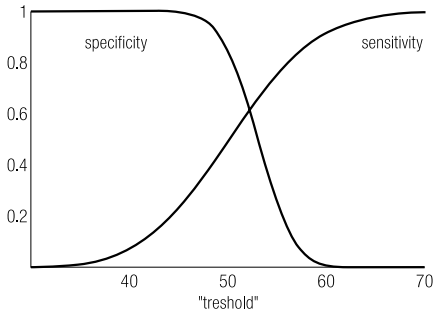

Fig. 7. Sensitivity and specificity vs. threshold score for the distribution of scores seen in Fig. 6 when true prevalence is 10%.

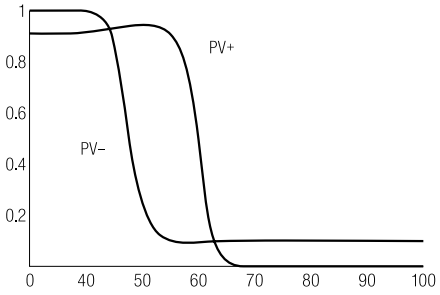

Fig. 8. PV+ and PV- for the same example.

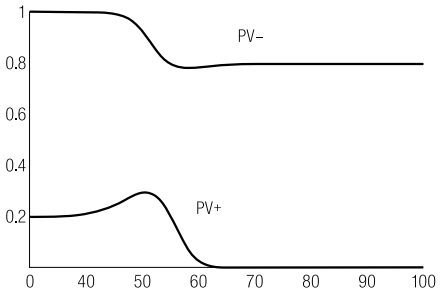

Fig. 9. PV+ and PV- for a true prevalence of 80%.

behavior of the negative predictive value. Consider the following example. Suppose we choose some feature of a chemical, and calculate a "score." The better the score the more likely we are to allow the technology. We measure the criterion along the x-axis, with values to the left worse than those to the right. Chemicals will naturally have a distribution of scores. If we have not chosen particularly good criteria to calculate a score we might have a situation where, on average, the scores of the bad chemicals are worse, but they are also spread out over a wide band that also includes many harmless chemicals (Fig. 6).

As we ratchet up our standards for an acceptable score interesting things start to happen to the positive and negative predictive value. Clearly we want to get at least as high as 40 since at that point we have encompassed some of the bad chemicals and few of the harmless ones. As we go higher and higher with our threshold, that is, as we increase sensitivity, we succeed in getting more of the bad actors, but we lose specificity. Fig. 7 is a graph of the sensitivity and specificity as we increase our threshold ($\mu_1 = 50, \sigma_1 = 7; \mu_2 = 53, \sigma_2 = 3$).

If we calculate positive and negative predictive value for this example we find this happens (Fig. 8). We lose both negative and positive predictive value, the worst of both worlds. This example was for a prevalence of 10%. If the true prevalence is 80% it looks like this (Fig. 9).

In this case, too, we lose out if we increase our sensitivity by too much. The final 80% negative predictive value is just the prevalence of bad chemicals, not a feature of the decision rule. In fact, all possible combinations are possible (stretches where both increase, both decrease or each goes in opposite directions). This is a simple, and univariate example. We would expect things to get much more complicated as we use multiple criteria.

The lesson we can draw from this exercise is the following: When we try to make the Precautionary Principle explicit, we can no longer rely on our intuitions, or worse, our inclinations. The world is complicated, and at times counter-intuitive. Perhaps the Precautionary Principle is better left at the level of an adage, such as "Look before you leap" or "Better safe than sorry." To do otherwise may put us in mind of another adage, "Be careful what you wish for." If the Precautionary Principle does nothing else than encourage the liberal use of common sense, which regrettably is not very common, it will still have done an important job.

REFERENCES

1. Jordan A, O'Riordan T. *The Precautionary Principle in contemporary environmental policy and politics.* In: Raffensperger C, Tickner J, editors. *Protecting Public Health and the Environment: Implementing the Precautionary Principle.* Washington DC: Island Press; 1999. p. 15–35.

2. EEA. *Late lessons from early warnings: the precautionary principle 1896–2000.* Environmental Issue Report No. 22. Copenhagen, Denmark: European Environment Agency; 2002.

3. O'Brien M. *Making Better Environmental Decisions: an Alternative to Risk Assessment.* Cambridge, MA: MIT Press; 2000.