DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

A FRAMEWORK FOR UNREASONABLE RISK IN THE TOXIC SUBSTANCES CONTROL ACT

Talbot Page

SOCIAL SCIENCE WORKING PAPER 368

February 1981

ABSTRACT

"Unreasonable risk" is a key term in the Toxic Substances Control Act. A finding that a chemical "may present an unreasonable risk" must be made before testing can be required, under section 4a of the Act; a finding of "presents or will present an unreasonable risk" must be made before a chemical can be regulated under section 6a. In the act and legislative history, "unreasonable risk" was defined in very general terms as a balancing of the probability of harm, the severity of harm, and benefits of the chemical in question. The Environmental Protection Agency was delegated the responsibility to define a legal concept of unrasonable risk. I wrote this paper while in residence at the Office Pesticides and Toxic Substances (EPA) as part of the effort to define "unreasonable risk" for regulatory purposes.

This paper develops a concept of unreasonable risk based on economic efficiency. Three ingredients are brought together in a common framework: baseline or existing information; the characteristics of a test, if testing is to be an option; and the valuation of the costs and benefits of the various control options.

The framework is designed to accomodate "typical" characteristics of toxics problems: "zero-infinity" dilemmas and the low statistical power of tests. Because of pervasive uncertainties in both baseline information and in new information, several rules of thumb and policy directions are suggested. A complete definition of unreasonable risk would go beyond minimization of expected costs to incorporate considerations of equity.

A FRAMEWORK FOR UNREASONABLE RISK IN THE TOXIC SUBSTANCES CONTROL ACT

Talbot Page

The purpose of this paper is to put three ingredients of a concept of unreasonable risk into a common framework. The ingredients are: baseline or existing information, the characteristics of the test or combination of tests, and the valuation of the costs and benefits of the various control options. The framework permits systematic consideration of the cost of testing and the value of new information from testing. I will focus directly on some aspects of the cost and incompleteness of information but I will leave out others or just touch upon them, as they go beyond the goal here. Let me mention from the start that I do not think that it is always necessary to know very much about each of these three ingredients before making a finding of "presents an unreasonable risk." (This finding in Section 6a leads to precautionary regulation) or "may present an unreasonable risk" (this finding in Section 4a leads to testing.) When the costs of information are taken into account the framework suggests procedures which operate on highly incomplete information.

TSCA and the legislative history offer considerable discretion to the Administrator in developing a concept of unreasonable risk. The legislative history calls for a balancing of costs, benefits, and risks, but not necessarily formal, quantitative cost-benefit analysis. 1/ Although

there is increasing interest in introducing risk-benefit (or cost-benefit) analysis into regulatory decision-making and there is increasing use of quantative risk assessment for carcinogens, there seems to be some confusion as to what such analysis means in the areas of environmental risks, where information is highly incomplete. For this paper, a risk-benefit approach does not mean that for a chemical all the costs, benefits, and probabilities are to be numerically estimated in some fashion or other, toted down to some bottom line which tells us whether to regulate or not. Instead the idea is to develop a balanced view of the costs and probabilities of regulatory options--including the option of no action--and to suggest procedures so that in the long run the expected cost of regulation is minimized.

The approach here attempts to strike a balance between two risks. The first is the risk of taking precautionary action for a chemical which is safe (a regulatory false positive). The second is the risk of not controlling a chemical which is unsafe, and which would be controlled with better information (a regulatory false negative). In the majority of cases for potentially toxic chemicals, decisions are made under pervasive uncertainty. A decision to postpone precautionary action until there is better data is just as much a decision under uncertainty as a decision to take precautionary action in the meantime. In fact, the crucial

decision is what to do "in the meantime" while uncertainties are far from resolution. The central fact about decisionmaking under uncertainty is that the risk of a false positive "trades off" against the risk of a false negative. The risk of a false positive can be reduced, but at the price of increasing the risk of a false negative. The essence of the balancing process is a willingness to accept some false positives as the unavoidable means of controlling false negatives. How many false positives should be accepted for each false negative depends upon the relative costs and probabilities of the two risks. We may not be able to tell with precision just how many false positives should be accepted for each false negative, but we can sharpen our perspective by setting out the basic ingredients, considering the typical asymmetries and features of toxics problems, and putting them together in a common framework.

In general terms the approach is an extention of the one used by Learned Hand in a liability case having to do with a barge which broke loose from its pier. 2/ Hand defined three variables: the probability that the vessel will break away (P); the gravity of the injury (L); and the burden of adequate precautions (B). In this case "adequate precaution" would have meant having a bargeman in attendance. The duty of the bargeowner to have a bargeman in attendance, and hence liability, depended as Hand wrote it, on whether $B<PL$, or whether

the cost of precautionary action was less than the expected cost of the uncontrolled situation. Hand did not attempt to quantify the variables, but noted that when a storm threatens P is higher, and the duty of the bargeowner to take precautionary action clearer. In other words, Hand made use of a simple expected cost framework for a qualitative analysis. 3/

Hand's analysis dealt with the evaluation of existing information. For the management of potentially toxic chemicals the situation is complicated by the opportunity of gathering new information through testing. With the opportunity to test the question arises as how to evaluate new information and fit it together with existing information. Tests are statistical and quantitative in nature and their performance is characterized by two probabilities, both calculable for specific effects of concern. These are the probabilities of failure: (1) the probability that the test will "find" a nonexistent effect, which the probability of a test false positive and is traditionally denoted α; (2) the probability that the test will miss an existing effect, which is the probability of a test false negative and is tradionally denoted β. Thus, in the absence of testing, Hand was concerned with one probability, which characterized the existing baseline information; with the additional possibility of testing we are concerned with three.

The most distinguishing feature of Hand's approach, when applied to the toxic chemicals problem, is that it treats the roles of false positives and false negatives symmetrically. This is in contrast with other approaches which devote more explicit attention and importance, in a decision sense, to the role of false positives.

Three Characteristics for the Framework

The three ingredients correspond to three characteristics of the toxic chemicals problem. I discuss these and other characteristics elsewhere in more detail, 4/ but the following three are the most essential:

(1) The first asymmetry, between the potential costs and benefits. Many toxic chemicals problems involve adverse hypotheses, which, if true, lead to catastrophic costs. Often the costs of precaution are large, in the millions or hundreds of millions of dollars. But compared with the potential magnitude of the catastrophic effect, if the adverse hypothesis is true, the costs of precaution are relatively modest. Tris, PCBs, asbestos, and CFCs all have aspects which illustrate this asymmetry in potential costs. In terms of Hand's notation L, gravity of the injury, may typically be large compared with B, the burden of adequate precautions. "Catastrophic" is a relative concept, and we are really concerned with the ratio L/B.

The asymmetry in potential costs is not the same for every problem; it is stronger for CFCs as aerosol propellants than for CFCs as refrigerants. For some problems like that of carbon dioxide and the greenhouse effect, the cost of control (substantial decrease in fossil fuel combustion) may not be relatively small compared with the potential costs of climate change. We want a framework designed to take this potential asymmetry into account, with flexibility for different degrees of it, to cases where there may be no asymmetry at all.

(2) The second asymmetry, between the probability of the benign hypothesis and the probability of the adverse hypothesis. Often there is a low probability attached to the adverse hypothesis, especially an adverse hypothesis about a potential worse case. In the ozone depletion example, some experts have suggested that there might be about a 1 percent chance of sufficient ozone depletion to melt the ice caps, increase cancer rates significantly, and severely disrupt the climate. Similarly some have suggested that here might be a one to five percent chance that age-adjusted cancer rates could double in the next twenty years due to the presence of man-made contaminants. (Others have suggested that this likelihood may be substantially higher.)

In Hand's notation this probability or likelihood is P. This likelihood is not directly calculable, but is an

"assessed" value. It is a translation of all the existing information into a statement of the level of suspicion about the adverse hypothesis. Equivalently, we could translate the existing information into a statement of the odds against the adverse hypothesis (N).5/ We can imagine a scene where the exasperated regulator backs the scientific community against the wall and says "What is your level of suspicion? What are the odds against the effect happening?" Scientists (and lawyers) are reluctant to make numerical statements about uncertainty; nonetheless, the regulator is forced, in a real sense, to bet on the likelihood of the effect of concern whenever he makes a decision toward or away from precautionary action.

As a practical matter it may not be possible to assess P, or equivalently N, with anything close to precision. However, it is critical for the decision process to have some rough idea whether the existing information points to a high or low level of suspicion, whether the odds against the environmental effect are long or short. It is also critically important to develop processes whereby we learn how well risk assessment is done and how it can be improved.6/ To do this we need to see how the level of suspicion fits into the framework.

The relatively low probability often attached to the adverse hypothesis is the second "typical" asymmetry. The two

asymmetries taken together form what has been called the "zero-infinity dilemma" -- a "nearly" zero chance of a "nearly" infinite catastrophe. We want a framework to allow for this second asymmetry, but again to be flexible enough to allow for different degrees of it, including cases where the adverse hypothesis does not have a relatively low probability, or even where has been established with a large measure of scientific certainty.

(3) Low Power. It is common practice in testing for carcinogenic and other toxic effects to set α equal to 5 percent (the probability of a false positive is also the significance level of the test). Whether or not this makes sense for toxic chemicals depends in part on the corresponding β for the effects of concern. For the most commonly used bioassay for potential carcinogens, for which there are 50 animals in the control group and 50 in each treated group, β may be greater than 50 percent. 7/

If the test is considered positive when it is positive for any one of twenty possible sites, the true probability

of false positive substantially increases over the α for a single site; but when the historical rate of tumor in the controls is taken into account the true α is greatly diminished. Fears et al. have calculated that when the rate of tumor in the controls is 1%, the true α for a twenty site experiment is less than .2%. 7/. For a two dose experiment which is taken to be positive if the test is positive at the same site for both doses, the tradeoff between the true α and β is shown in Figure 1, for a potential 10 percent access risk, along with the nominal α, (nominal α is obtained by ignoring information on the historical rate of tumor in the controls).

The tradeoff between α and β is often established by setting nominal α equal to 5 percent. For the calculated example, this suggests that we may routinely design and interpret tests with true α far less than 1 percent and β as high as 74 percent, for a potential excess risk of 10 percent (the most upper point in Figure 1). It doesn't make much sense to have such a skewed tradeoff, so highly protective against the false positive and so weakly protective against the false negative. 8/ One of the most important features of the balancing approach, in an operational sense, is much greater attention to the calculation of false positive and false negative error rates. At present, analyses of β (and true α) are rare, both in epidemiologic studies and clinical tests. We consider below the application of the balancing

concept for a determination of the tradeoff between α and β based on minimizing expected cost.

As a general observation, it appears that for many toxics problems, adverse effects, when they exist, are hidden with the consequence that even expensive, well designed tests have low probabilities of discovering effects when they exist. It would be desirable, for example, to be able to have both α and β less than 5 percent, but for this test in Figure 1 there is no way to get both below about 25 percent. The property I have called low power, typically associated with toxics problems, can be loosely restated as saying that the (α,β) tradeoff curve for effects of concern is "far" from the origin in a characteristics diagram such as Figure 1. Again, the framework, should allow for this characteristic, encouraging explicit attention and investigation of it, while providing flexibility to handle cases where it does not obtain.

To summarize so far, the framework should be able to take into account the three characteristics with sufficient flexibility for varying degrees of each or all the characteristics. In its "pure" form, when the characteristics are strong, the problem of controlling a potentially toxic chemical can be compared with searching for a needle in a haystack. First of all the needle may not be in the haystack at all (the low probability of the adverse

hypothesis). But if the needle is in the haystack, it is likely to be hard to find (low power). Nonetheless, if the needle is present and if we miss it there will be catastrophic costs (the infinity-ness of the dilemma). The potentially catastrophic costs are high in comparison with the cost of precautionary control which would keep the needle, or what we think may be the needle, from entering the haystack in the first place.

Carcinogens in Drinking Water: An Illustration

The three ingredients are brought together in a study of carcinogenic risk in drinking water. 9/ Figure 2 shows the result of an anlysis of the statistical power, a particular regression equation in the study. The statistical power, and hence β, are shown as a function of the magnitude of the potential drinking water effect, for a given α=.05. 10/ Even though the drinking water variable is highly significant in the regression equation, the statistical power for the estimated effect, an excess of 60 cancers per million annually, is only about 45 percent. In other words, even if the statistical model were correctly specified (there is almost always some specification error as there is in this case), and even if the true effect of carcinogens in drinking water were to increase the gastrointestinal cancer rate by 60 per million, there would be less than half a chance of finding it by this statistical test (the test was considered positive if the t value was greater than 2.0, for a confidence level of 95 percent).

Further regression analysis suggests that the drinking water effect might occur in cancers of the urinary tract as well as gastrointestinal sites, and for comparison the drinking water effect is also expressed as a percent of the total gastrointestinal and urinary tract cancer rate in Figure 3. On the basis of other existing information, at the time the first regression aquations were reported, it was widely held that the drinking water effect, if it existed at all, was small. Commentators did not translate their level of suspicion about the existence of the effect into an explicit probability assessment, but it appeared that an effect in the range from 0 to 3 or 4 percent was considered most likely, a range from 5 to 20 percent plausible but unlikely, and anything over about 30 percent to be extremely unlikely.

The benefits of control are the reduction of human cancers, mutagenic and other toxic effects, and other damage to the ecosystem (humans being considered part of the ecosystem). Of this, only the cost of cancer was even crudely quantified, (but with some attention to the problem of intergenerational equity). These benefits, on a steady state basis of comparison, were taken to be $500,000 per life saved, and correspond to Hand's L. The cost of control, by

means of granular activated carbon (GAC), was estimated to be $10.3 million for a city of 1,193,000. This corresponds to Hand's B. The cost-benefit ratio, for GAC control, increases linearly with magnitude of the drinking water effect, and hence the number of deaths to be prevented.

There is a simple and direct relationship between the cost-benefit ration L/B and the cost of a false negative relative to the cost of a false positive. If preventative action is taken when there is no drinking water effect, the cost of this mistake is the unnecessary cost of carbon treatment B. If preventative action is not under taken when the drinking water effect is substantial, the cost of this mistake is the cost of the environmental harm over and above the cost it would have taken to control it, or L-B. Thus in Figure 3 the ratio of the costs of mistakes is (L-B)/B and is plotted one unit below the cost-benefit ratio, labeled D for future reference. 11/

For practical purposes decisionmaking the region of concern is very roughly from 0 to 100 excess cancers per million, where the excess carcinogenic risk is at least plausible, even if considered unlikely. If the effect were known to be less than 3 or 4 percent, it is unlikely that any action would betaken, as the cost-benefit ration for cancer risk would appear to be less than one (other control such as changing the point of chlorination might still be taken). Where the effect of concern seems plausible from

existing information, in the range of 5 to 20 percent of background, the cost-benefit ratio is favorable for preventative action. However, for most of this range the statistical power of the test is low. An effect of 60 excess cancers per million is clearly large enough to warrant considerable preventative action, but the effect is still too small to find easily by statistical tests. All this is another way of saying that managing the toxic chemicals problem is like looking for needles in a haystack.

Balancing: A Numerical Example

Having described the ingredients, we are now ready to put them together in a framework that uses the notion of least expected cost. I'll start with a concrete illustration wich incorporates the two asymmetries of the zero-infinity dilemma and the test characteristics. Even though simplified, the example remains a bit complicated, as it must to reflect the nature of the toxics problem. In the example information is more complete than ordinarily is the case in toxics problems -- this is done to help fix ideas to see the relationships among the ingredients. With the example in mind, the framework is then illustrated by means of a diagram. The framework can be interpreted in qualitative terms and we can begin to ask what are sensible courses of action when information is less complete and more costly.

The example. You are feeling a little rotten and you go to an eminent doctor with a wide practice. "Yes," says the doctor after his examination, "it is probably nothing, but there is a small chance that you may have something serious. One person in twenty with your symptoms has a particular, rare but serious cancer."

Your concern aroused, you want to know how the doctor has assigned the 19 to 1 odds against the disease, "I keep a a track record," says the doctor. "I have followed up many thousands of patients with your symptoms. When we operate we can tell right away whether or not there is cancer. For those we don't operate on we can tell one way or the other in five years." My colleagues tell me that I am a fanatic for record-keeping."

"What about the cancer?" The doctor is reassuring. "It is too early to worry about cancer. You could walk away from the office right now with the odds strongly in your favor. But, he adds quickly, "I don't recommend this, because it is a nasty tumor (in the remote chance you have it). I have a very good test; take that and in two weeks we'll decide what to do."

He describes the test, which turns out to be painful and even a little disfiguring.

"What about the operation? There is no point in taking the test if I don't go through with the operation, if the test is positive."

He describes the operation and it appears to you that the cost of the operation, in pain, recouperation, permanent impairment, and doctors bills, is small compared with the prospect of the cancer. The doctor adds that the operation is completely successful, "if we nip the tumor in the bud."

"And what happens if I do not take the test and I turn out to be that unlucky one in twenty?"

"Then you will die."

You turn your attention back to the test. "How accurate is it?" "I have kept careful records. Those who have the tumor score positive on the test 90 percent of the time. Those who do not have the tumor score negative on the test 90 percent of the time."

You ponder this information. You have no way of judging this matter precisely, but dying of the cancer is ten times worse than the operation you roughly guess. Compared with these potential costs, the cost of the test is a mere annoyance

("inconvenience" the doctor puts it). The test appears highly accurate, so you decide to follow the doctor's recommendation and take it. Two weeks later you return for the results. "Sit down," says the doctor. "You have a serious but not fatal problem--the test came out positive --I recommend the operation, as a precaution."

"But almost surely I must have cancer, since the test is positive. Why as a precaution?"

The doctor reassuringly, "I have kept careful records and made the calculation. This positive test result means only that you have a 32 percent chance of having the tumor." You are astonished, but he keeps going. "Even though the odds are still almost two to one in your favor, I recommend the operations, as a precaution."

"Isn't there some alternative," you say. "I hate to have an operation when the chances are that it is unnecessary."

"There is another alternative. We can reinterpret the test to be more sensitive against false positives. By changing the 'critical value' of the test we can halve the probability of a false positive--it was 10 percent but it can be reduced to 5 percent, which is the value used by many statisticians." Just as you perk up, the doctor continues,

"Unfortunately, in doing so, the sensitivity of the test to protect against false negatives goes down. In the test I recommend, the probability of a false negative is 0.1, but in the alternative test with the lower probability of a false positive the probability of a false negative goes up to 0.5. It's a tradeoff, and I recommend the original test. The way most people value their lives, it leads to lower costs, on average."

You decide to take the operation because you don't want to live with the prospect of even a 32 percent chance of the cancer when you can avoid the gamble at relatively low cost, and you take the doctor at his word on the trade-off of false positives and false negatives.

After the operation, at one of the routine checkups for the surgery, you thank the doctor for the skill of his knife and the candor of his advice. "By the way," you ask, "did I have the tumor?"

"No, you were one of the 66 percent for whom the operation turns out to be an unnecessary precaution."

You don't complain because you knew the odds were in your favor before the operation and there is no sense of wishing you had the tumor to "justify" the operation. You ask him further about his approach.

"I really follow a very simple strategy," the doctor responds. "With rarer tumors or ones that are less severely malignant, I wait and watch. But I take a conservative approach with this one. Even at the stage where the baseline information indicates only a five percent chance that the adverse hypothesis is true, I recommend taking the test and following it without developing further information. For this tumor there is high cost of waiting, as the operation becomes less effective the longer you wait, but I would be buried in malpractice suits if I didn't keep complete records and fully inform the patients before the operations."

"I suppose that every once in a while one of the people you operate on who turn out not to have the tumor get sore. With only a 32 percent chance of cancer on a test positive, you must expect about two unnecessary operations on people without tumors for each operation with a tumor."

"The mistake ratio is a lot worse than that. I have kept track of my two types of mistakes. With my follow-up studies I can tell how many people were diagnosed as positive, had operations, but who were actually tumor-free--my false positives. And I have kept track of the people who were diagnosed as negative, didn't have the operation, but who later died of the tumor -- my false negatives. For every patient who slips by with an undiagnosed tumor I have 19 operations on tumor-free patients. Allowing these 19 false positives is the price paid to control false negatives."

Again you express astonishment. Is this doctor really striking the best balance between the two types of mistakes? There is no trick here, the criterion of expected cost minimization, which the doctor is following, leads to more precautionary action, when there are the two asymmetries of the zero-infinity dilemma, than one might casually expect, especially in terms of the least cost ration of mistakes -- the ratio of false positives to false negatives.

The doctor has four strategies. (1) Conclude that the baseline information ("your symptoms") is strong enough to warrant the precautionary operation without the test. This corresponds to a finding of a "presents an unreasonable risk" for direct regulation under section 6(a). (2) Conclude that the baseline information is weak enough to reassure the patient and neither test nor operate. This corresponds to a finding that there is not an unreasonable risk as in section 5(g). (3) Test and follow its result, with the test standard adjusted to a 0.1 probability of both false positive and false negative. (4) Test and follow its result, with the test standard adjusted to a 0.05 probability of a false positive and a 0.5 probability of false negative. These last two correspond to a finding of "may presents" for section 4(a), leading to testing. There are other alternatives, such as paying the cost of improving the baseline information before deciding whether or not to test (in the example,

waiting for more advanced symptions or in TSCA performing an exhaustive literature search, while delaying action in the meantime on the chemical at hand and other chemicals in the queue). This too can be analyzed in an expected cost framework, but here I want to keep the details to a minimum.

Figure 4 shows why the probability of having cancer, when the test is positive, is only 32 percent. As can be seen by the following branches, the predictive power of the test is substantially weakened by the underlying rarity of the effect of concern. Although illustrated by a single numerical example, this weakening of predictive power by the rarity of effect is a general phenomenon, directly implied by Bayes Theorem. 12/

Immediate observation from Figure 4 shows that when $\alpha = \beta$, the mistake ration is simply (1-P)/P or N. Direct observation also shows that when the test is balanced with $\alpha = \beta$, we should expect more test negatives than positives when we are testing a group of chemicals each one of which we believe to have a less than even chance of having the effect of concern. Thus by keeping score on the total number of test positives and test negatives, and later discoveries of false positives and false negatives, we can develop consistency checks on our assessment of the level of suspicion (N) and the test characteristics α and β. The geometry linking these three variables is shown in Figure 5.

Returning to the numerical example, Table 1 summarizes the implications of the four strategies for a group of 1000 patients with the baseline symptoms. The "standard of proof" is defined here as the level of suspicion that the effect exists, at the time a decision is made, toward or away from precautionary action, taking into account new information from the test, if there is a test, along with the previously existing information. In Hand's case, which was decided upon existing information, there being no test, the level of suspicion justifying a liability is P, and a decision affirming the owner's duty to provide the bargeman's attendance is implied as long as $P > B/L$. Because the burden of adequate precaution is small compared with the gravity of the barge's sinking, the probability of the barge breaking loose from its mooring can be quite small and still have liability required. We are used to a high standard of proof in criminal law, where the cost of a false positive (convicting the innocent) is large compared with the cost of a false negative (acquiting the guilty). But in tort law, and in the control of toxic chemicals, these relative costs are likely to be reversed, and the standard of proof, along with the rest of the decision process reflect the relative costs.

Row 4 shows that the mistake ratio is strikingly higher for the test with $\alpha = .1$ compared with the "traditional" test with $\alpha = .05$.

TABLE 1

	STRATEGIES			
	don't test; operate	test, operate if positive; $\alpha = 0.1$ $\beta = 0.1$	test operate if positive $\alpha = .05$ $\beta = .5$	don't test don't operate
1. Expected number of operations (total positives)	1000	140	72.5	0
2. Expected number of deaths (false negatives	0	5	25	50
3. Total cost of strategy (row 1 plus ten times row 2)*	1000	190	322.5	500
4. Mistake ratio (number of false positives for each negative)	oo	19	1.9	0
5. Standard of proof **	5%	32%	72%	5%
6. Accepted risk ***	0	0.5%	2.5%	5%

* Based on a cost of cancer ten times worse than the cost of the operation. The rankings of strategies is the same where the total cost is minimized or the expected cost of mistakes.

** Probability of cancer when action is taken.

*** Probability of regulatory false negative.

Balancing: The Framework

We are now ready to see how the three ingredients fit together in a common framework. So far we have considered the only cost of the test to be the potential cost of it yielding erroneous information. At this point it is a simple matter to also take into account the resource cost of the test itself, but first let us summarize explicity the two extreme strategies where the potential cost of erroneous test information leads us to avoid testing altogether. If we place a very high cost on a regulatory false positive, we may wish to avoid the risk of a test false positive altogether by not testing and not taking precautionary action. In this case the cost of a regulatory false positive is avoided, but at the full risk of a regulatory false negative. Since all that really matters is the ratio of relative costs, we can count the cost of a (regulatory) false positive at one unit and the cost of a false negative at D units (Set D = (L-B)/B. Then the expected cost of the no test, no control strategy with the full risk of a false negative is PD, the likelihood of the false negative times its cost.

At the other extreme we may value the cost of a regulatory false negative so highly that we do not want to bear the risk that the test entails of a false negative. We can avoid risk of a test false negative by taking

precautionary action without prior testing. In this strategy we bear the full risk of a false positive. The false positive, valued at one unit has a level of suspicion of probability of (1-P), hence expected cost of (1-P).

In between these extremes we test and condition action on the results of the test. We also bear the resource cost of the test, which we can count in units T, relative to the cost of control. The expected cost of the testing strategy is $PD\beta + (1-P)\alpha + T$.

In order to adjust the scale to accommodate figure 6, we divide each of these expected costs by PD and recall that (1-P)/P = N, the odds against the effect. Simple geometry shows the relative rankings of the expected costs for each strategy. It can be seen from the geometry how the rankings of the expected costs depend on the interaction of all three ingredients: the value of new information from the test, summarized by the (α, β) tradeoff curve; the existing information translated into a statement of the level of suspicion (N); and the relative costs of precaution and the potential risk (D).

The framework is a general one. It encompasses what I have suggested is the paradigm toxics problem: low power, where the (α, β) tradeoff curve is "close" to being a diagonal; high cost of a false negative to the cost of a false positive, D "significantly" higher than one; and low

probability of the adverse hypothesis, or at least a worst case statement of it, N "significantly" higher than one. the framework also encompasses cases where any or even all these characteristics do not obtain. N or D can be less than one, the (α, β) trade-off curve close to the origin. Thus the framework is really a language to talk about the toxics problem. It gives us a vocabulary but does not force us into narrow specific assumptions. By providing a vocabulary and grammar linking the vocabulary, it encourages us to investigate empirically the degree to which actual toxics problems take on the paradigm characteristics. The framework encourages us to assess the strengths of the three ingredients for actual toxics problems to decide what, if any, precautionary action is prudent, in a balancing sense.

The balancing approach of this framework differs fundamentally from the alternative approach which requires that the risk be first established before costs are taken into account. This may mean resolving the scientific uncertainties, and when performing statistical tests, setting α at 5 percent, without calculating β. In arguing against the CFC aerosol ban, the American Chemical Society said the proposed ban would be "a very dangerous precedent" because it would be "the first regulation to be based entirely on an unverified scientific prediction." 13/ In this alternative approach consideration of the cost of precautionary control relative to the potential environmental harm comes only after the risk is established. In the words of Carmen Guarino, Water

Commissioner of Philadelphia, "If future research proves a true link between water-borne organics and cancer in humans, Philadelphia will spend whatever is necessary to cope with the problem. 14/ In Hand's notation we must first prove the P is close to one, then take precautionary action if B<L. The problem with this approach is obvious: it offers no protection against zero-infinity dilemmas. The CFC aerosol ban was an application of the balancing approach: even though ozone depletion by CFCs was indeed only a scientific hypothesis at the time of the ban, precautionary action was taken largely because the potential environmental costs were perceived to be enormous compared with the cost of the ban.

Somewhat differently, it also has been suggested that risk management proceed in two stages where in the first stage the scientists and statisticians do a risk estimation and in the second stage the economists and public policy people assess the risk in terms of the relative costs. 15/ It is conceivable that risk management could proceed along these lines with the two groups working separately, if the process is considered to be an iterative one and sufficient information is passed between them and if each respond to each other's needs, but there needs to be considerably more communication among the various disciplines than currently exists. It can be seen from a glance at Figure 5 that N (ostensibly the province of the scientists) and D (ostensively the province of the economists) play interactive and in places entirely symmetric roles.

In order for the scientists to know what chemicals to choose, how to prioritize them, how to pick the tests, and how to choose the proper tradeoff of α and β, they must have some idea of D. This does not mean that a cost benefit analysis should be done prior to the test, conditioned on all possible outcomes of the test. Requiring this does not make sense in terms of the cost of information, and is like requiring Bobby Fischer to write down in advance all his possible countermoves conditioned on each move his opponent might make. Clearly there are enormous cost savings if the game is analyzed as it goes along. Control options for toxics, like the moves in a chess game, are too numerous to analyze in full quantitative detail beforehand. Still, just as Fischer comes to the board with analyzed openings in his head, the scientists must have some partial idea of D.

The policy statement of the proposed rule requiring testing of chlorobenzenes is in accord with the framework. In terms of Figure 5, if D were known to be very small, or N known to be very large, the slope N/D would be large and any tangent to the (α, β) trade-off curve would cut the y-axis above 1, when translated upward by the test cost T. This means that the expected cost of the no test, no control strategy is less than that of testing, or taking precautionary action without testing.

Thus is N were known to be sufficiently large and D known

to be sufficiently small we could make a finding that there is no unreasonable risk to the chemical (an finding corresponding to no test and no control). However, neither N or D are known with any precision. To make a finding that the chemicals "may present an unreasonable risk" we need some assurance that N "may" be small enough and D "may" be large enough so that a tangent to the trade-off curve with slope -N/D could cut the y-axis below 1. As a practical decision rule, we require that there is at least some evidence of toxicity (some evidence that $N < \infty$) and some evidence that there will be potential exposure (some evidence that $D > -1$). The evidence can be weak in both cases, especially if the cost of the test is relatively low, compared with the potential environmental harm, and still warrant the "may presents" finding. Although the uncertainties themselves may be large. It is important that we have a "reasonable basis," or reasonable methodology in which the uncertainties are taken into account. 16/

Once a test is undertaken, the level of suspicion, N, is updated, and if the test is suggestive of control action, D is analyzed more carefully, with specific control options in mind, to see whether the "presents an unreasonable risk" finding is warranted, along with the subsequent precautionary control. If there is no test (or if there is a test and N is undated) the criterion for the "presents" finding is whether or not $N < D$, which is just a restatement of Hand's criterion $B<PL$. To be meaningful, these conditions need to be translated into specific policy direction, a few of which are suggested below.

Policy Steps Toward the Balancing Approach

1. Traditionally in epidemiologic and experimentally controlled studies of potentially toxic chemicals, α levels are pre-set, usually at 5 percent, and β levels, for the effects of concern, are not calculated. To move toward a balancing approach, probabilities of false negatives must be routinely calculated. Without such analyses there is no way of balancing the risk of a false negative against the risk of a false positive.

In recent years industry groups have increasingly asked for negative findings to be taken into account in the regulatory process. However, there is literally no information content in a negative finding unless there is an analysis of statistical power, or equivalently the probability of a false negative. Thus statistical power analyses are a necessary first step toward taking negative findings into account. Such analyses are also a requisite for deciding what is a "positive" test and how sequential tests can be designed to minimize the cost of regulation.

2. In the Environmental Protection Agency, most of the resources of cost-benefit analysis have gone into the study of the costs of control, the cost of a false positive. Because of the critical role of false negatives, equal or more resources should go into the study of the cost of potential harm (there is a move already in this direction).

At the present time quantification of the potential costs of environmental harm is rudimentary to put the matter delicately. Although such quantification is intrinsically difficult, and can be self-deceptive if carried too far, there is little doubt that the assessment of environmental costs can be improved. As can be seen by the framework, the assessment of environmental costs need not be precise. In many situations, order of magnitude estimates may be enough. It makes little sense, in decision terms, to have estimates of control costs that are much more precise than the estimates of potential environmental cost -- D is a ratio and it makes little sense to estimate the denominator with much greater precision than the numerator.

3. With greater attention on the estimation of potential environmental costs it becomes correspondingly more important to stress the equity aspects. Some environmental costs are to be avoided on the grounds of unfair distribution of risk and cost (especially long lived risks falling on succeeding generations). As we develop balancing notions of unreasonable risk that depend on the aggregative D, we must also develop equity notions which depend of the disaggregated distribution of D, in order not to weight the decision process too heavily in the aggregative direction.

4. Traditionally scientists are reluctant to translate their knowledge and their uncertainty about the effect into a statement of their level of suspicion. For many practical decisions this translation need not go all the way into a numerical statement of the odds against the effect, but a move toward a more aggregative and more explicit assessment of the likelihood of the effect is needed for a balancing approach. In some exhaustive evaluations of existing information, the reader is left not having any idea if the scientists think there is a 10 percent chance of the effect occuring or a 90 percent chance. 17/

5. As risk assessments become more explicit and quantitative, we need to keep score more carefully, to evaluate how well the assessment process is working and to suggest ways of improving it. As noted in Figure 5, N, α, and β are related to the number of positives and negatives and the number of false positives and false negatives of a test. The relationships between permit consistency checks to see of N, α, and β are close to what we think they are. The balancing approach suggests acceptable ratios of false positives to false negatives, and by keeping score we can try to uncover mistakes of both types. The numerical example suggests that we should have more false positives than false negatives, for "typical" toxics problems, when we are minimizing the expected cost. But we appear to discover more false negatives than false positives. 18/ Is this because the former are easier to discover

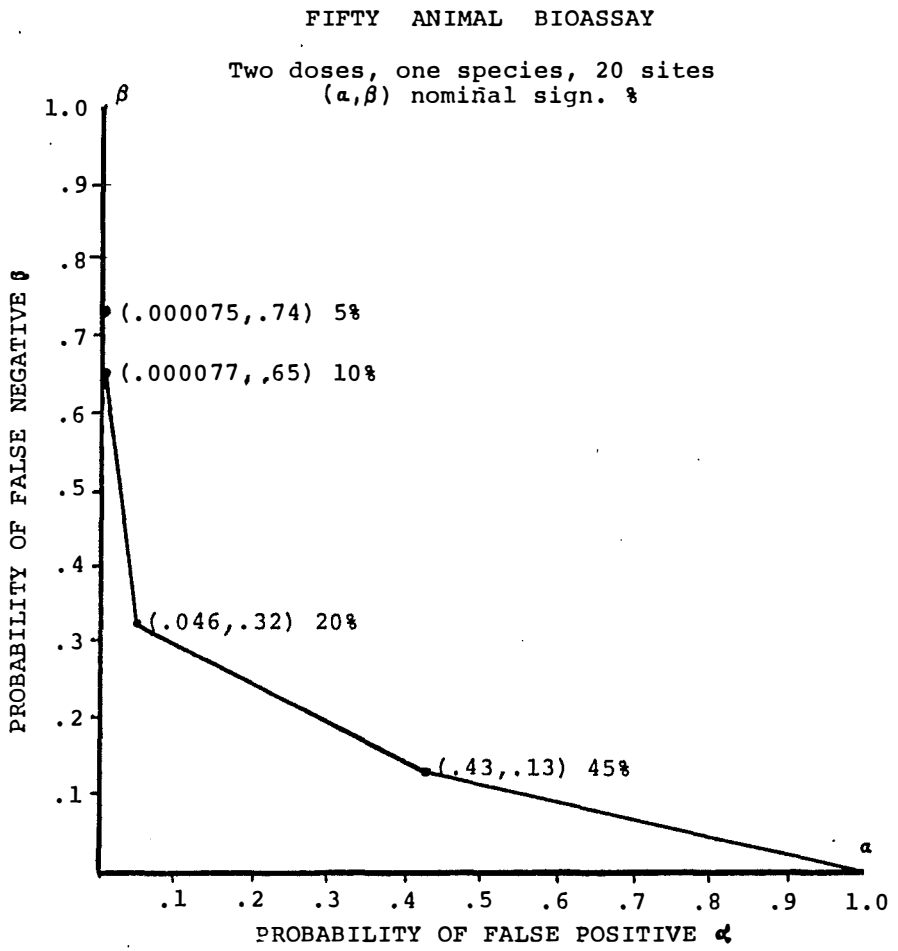
than the latter, or are we far from minimizing expected cost, or what?

6. For a balancing approach we need to focus more on the value of information and its cost for new chemicals. We need to ask what is the "best" base set of tests costing $25,000, $50,000, or $100,000, where "best" is construed at least qualitatively in terms of N, D, α, and β, and minimization of costs.

7. The essential principle of a balancing concept of unreasonable risk is that there must be a willingness to accept regulatory false positives as the price of controlling false negatives. This principle should be applied to the definition of categories generally in TSCA and to the literature searches prior to requirements for testing. For categories the risk of a false positive is the risk of drawing the category boundary too broadly so that action is taken on a chemical which is really undeserving of the action. The risk of a false negative is the risk of drawing the category boundary too narrowly so that some precautionary action (for example testing) is not taken but which in fact really warrants the action. For literature searches, the risk of a false positive means requiring a test of a chemical, when in fact the test will be unnecessary or duplicative.

In the balancing approach we acknowledge that we are making decisions under uncertainty and that the risk of a regulatory false negative must be weighed against the risk of a regulatory false negative. The regulatory false negative, is that we may take too long on the search on one chemical, pre-empting resources from other chemicals, and unnecessarily spend tax dollars on dry holes.

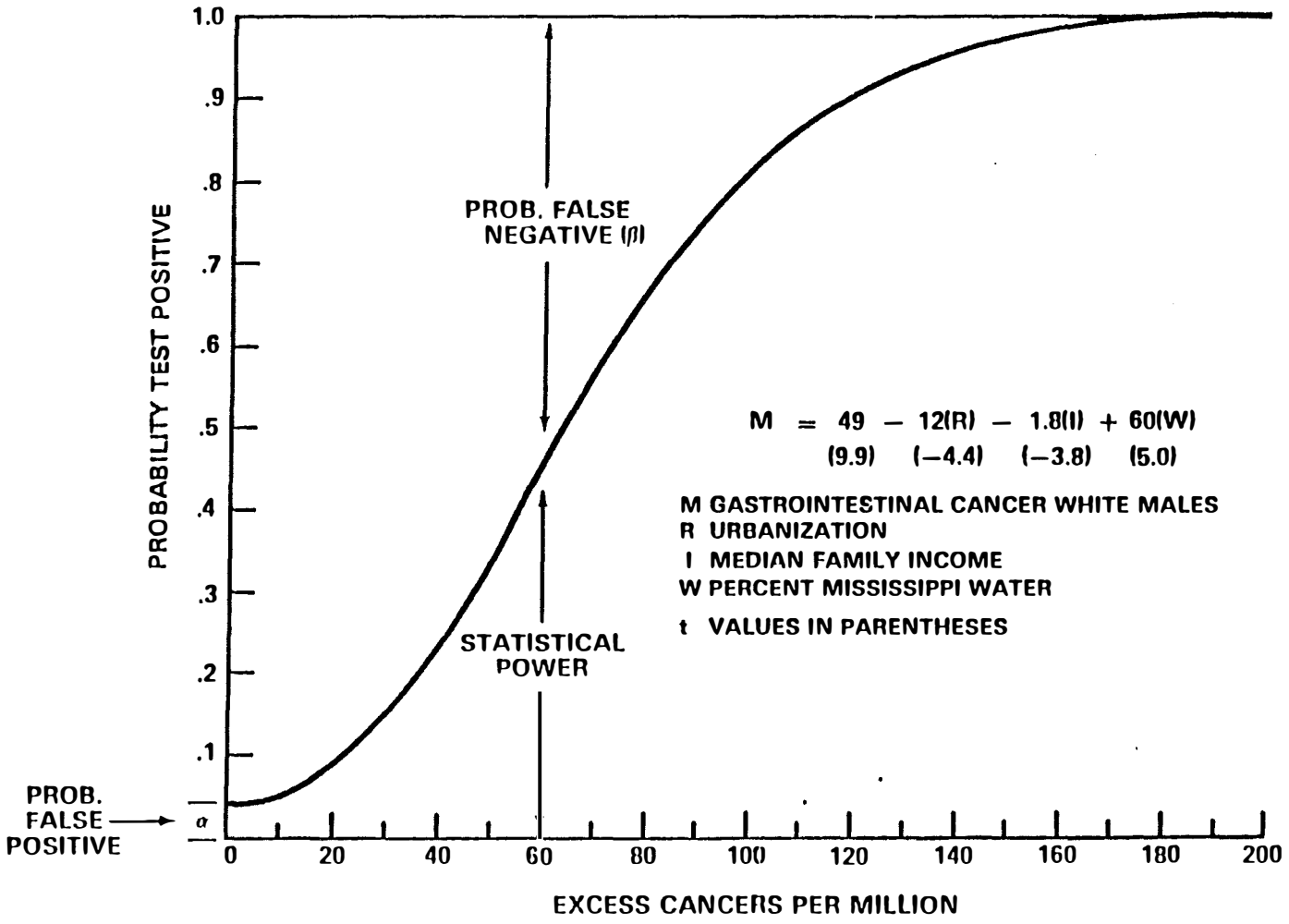
Figure 1



Test positive for a site if positive for both doses
Excess risk 10%
True tumor rate in controls 2%
Nominal critical values from one-tailed Fisher exact test

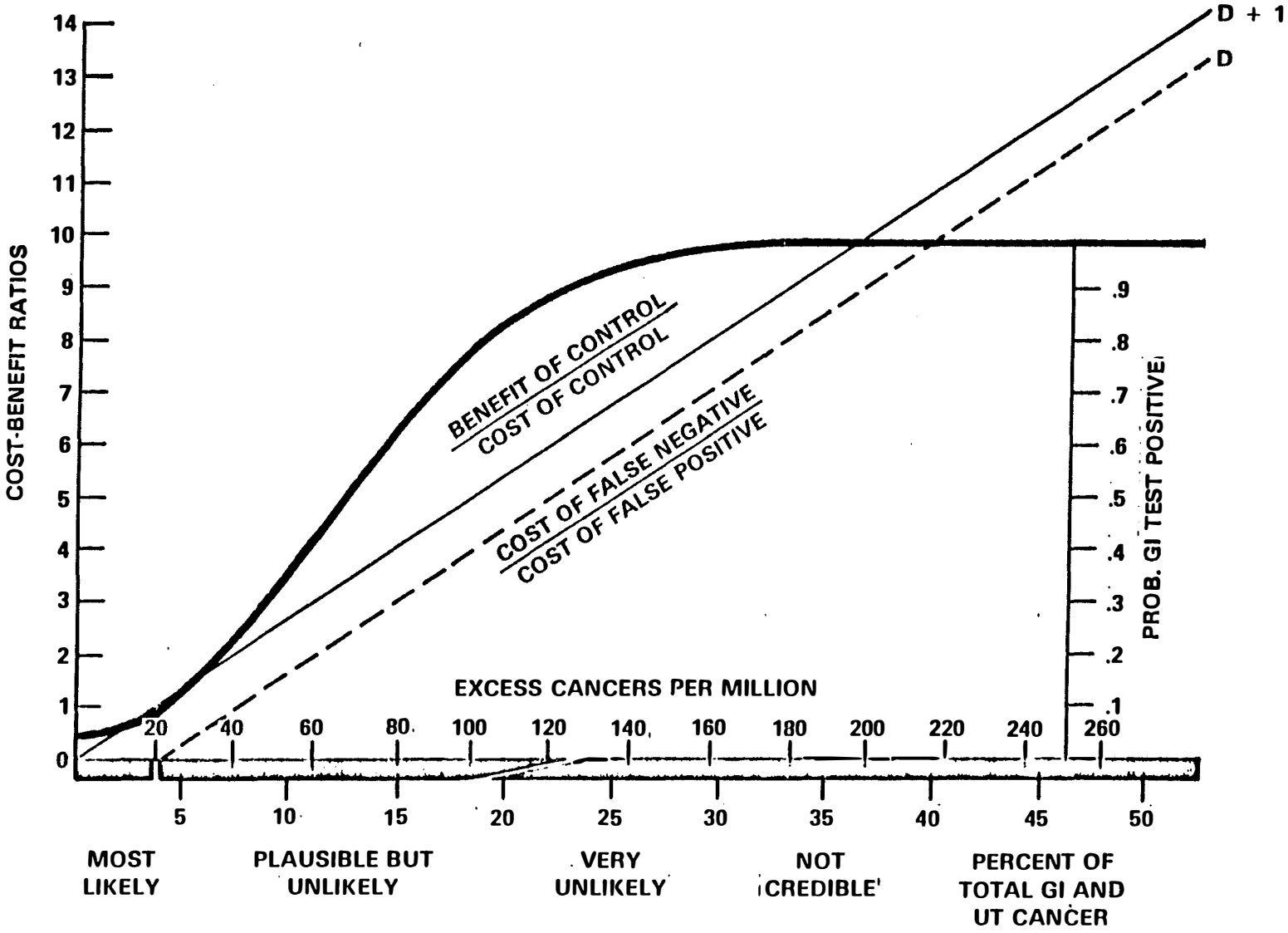
Figure 2

STATISTICAL POWER
(AS A FUNCTION OF BINOMIAL VARIANCE)



SOURCE: HARRIS, PAGE, AND REICHES, IN *ORIGINS OF HUMAN CANCER*, COLD SPRING HARBOR, N.Y., 1977

BENEFITS OF GAC AS A FUNCTION OF EXCESS CANCER

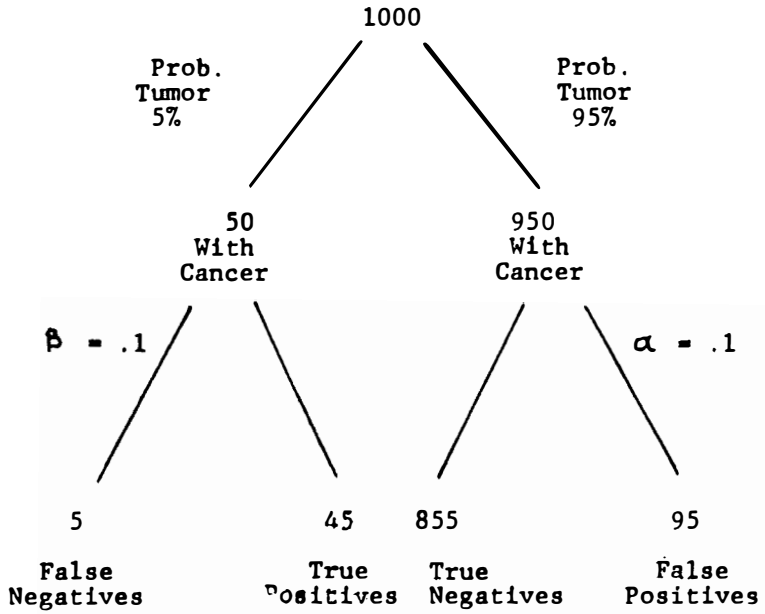


ANNUAL GAC COSTS $10.3 MILLION, FOR CITY OF 1,193,000; THM 250 ppb; GAC 90% EFFECTIVE REMOVAL; BENEFIT PER CANCER REMOVED $500,000; STEADY STATE COMPARISON

SOURCE: PAGE, AND BRUSER, IN *SCIENTIFIC BASIS FOR HEALTH AND SAFETY REGULATION*, BROOKINGS, FORTHCOMING.

Figure 4

Expected Values for 1000 Patients



$$P(C|pos) = \frac{\text{True Positives}}{\text{Total Positives}} = \frac{45}{45+95} = 32\%$$

$$\text{Mistake Ratio} = \frac{\text{False Positives}}{\text{False Negatives}} = \frac{95}{5} = 19$$

Figure 5

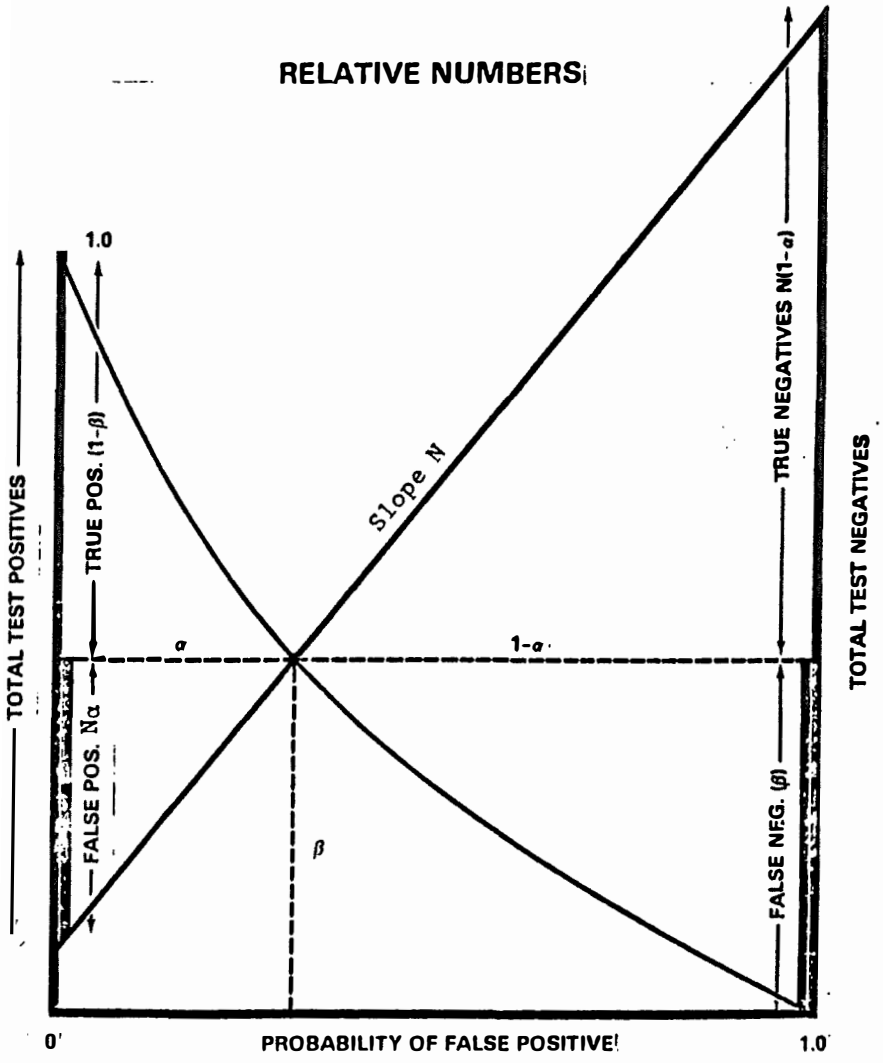
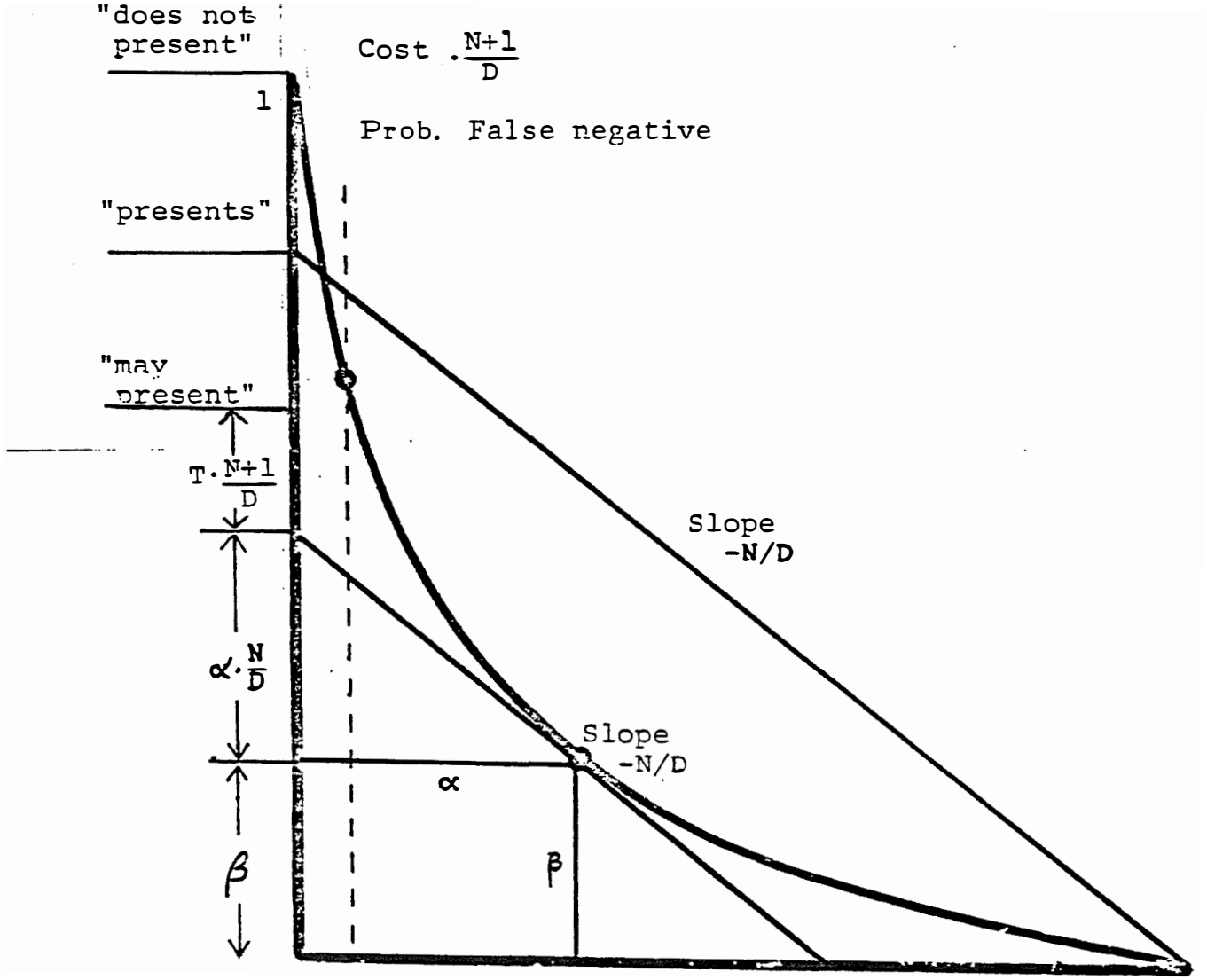


Figure 6

EXPECTED COSTS



PROBABILITY OF FALSE POSITIVE

"Does not present an unreasonable risk" corresponds to the strategy no test, no control.

"Presents an unreasonable risk" corresponds to the strategy no test, precautionary control.

"May present an unreasonable risk" corresponds to the strategy test and act depending on the result of the test.

Footnotes

1/ The House Report states:

Because the determination of unreasonable risk involves a consideration of probability, severity, and similar factors which cannot be defined in precise terms and is not a factual determination but rather requires the exercise of judgment on the part of the person making it, the Committee did not attempt a definition of such risk. In general, a determination that a risk associated with a chemical substance or mixture is unreasonable involves balancing the probability that harm will occur and the magnitude and severity of that harm against the effect of proposed regulatory action on the availability to society of the benefits of the substance or mixture, taking into account the availability of substitutes for the substance or mixture which do not require regulation, and other adverse effects which such proposed action may have on society.

The balancing process described above does not require a formal benefit cost analysis under which a monetary value is assigned to the risks associated with a substance and to the cost to society of proposed regulatory action on the availability of such benefits. Because a monetary value often cannot be assigned to a benefit or cost, such an analysis would not be very useful.

As noted above, the Committee recognizes that risk is measured not solely by the probability of harm, but instead includes elements both of probability of harm and severity of harm and those elements may vary in relation to each other. Thus, the Administrator may properly find that health or the environment are exposed to an unreasonable risk by a lesser probability of a greater harm as well as by a greater probability of a lesser harm.
(H. Rept. No. 94-1341, 94th Cong., 2d Sess., 7/14/76, at 13-14, Legis. Hist. 421-22, footnote omitted.)

2/ United States vs. Carroll Towing Co. 159 F. 2d 169 (2nd Cir. 1947).

3/ For further discussion of the balancing concept of unreasonable risk in tort law see Harold Green, "The Role of Law in Determining Acceptability of Risk," paper presented at the New York Academy of Sciences workshop "The Management of Assessed Risk for Carcinogens," March 17-19, 1980. For application of qualitative expected cost minimization for the analysis of contract law, see The Economics of Contract Law, Anthony Kronman, Richard Posner, Little Brown and Company, Boston, 1979

4/ Talbot Page, "A Generic View of Toxic Chemicals and Similar Risks," Ecology Law Quarterly, Vol. 7, No. 2, 1978

5/ The relationship is simply $P = 1/(1+N)$. Later on the "odds" notation is slightly cleaner, and we shall use it from time to time.

6/ There is already an enormous and growing literature on the assessment of risk. One of the best introductions to risk assessment and its place in decision making is Howard Raiffa, Decision Analysis: Introductory Lectures on Choices. For an insightful critique of the expected utility approach see Daniel Kahneman and Amos Tversky, "Prospect Theory: An Analysis of Decision Under Risk," Econometrica, Vol. 47, No. 2, March 1979, pp. 263-91.

There is also a literature on the evaluation of the accuracy of assessment. See, for example Tversky and Kahneman, Judgment under Uncertainty: Heuristics and Biases" 183 Science 1124, 1974; Slovic, Kunreuther, and White, "Decision Processes, Rationality, and Adjustment to Natural Hazards:" in Natural Hazards: Local, National and Global 187 (G. White ed. 1974); Baruch Fischhoff, "The Perception of Risk and Its Influence on Decision Making," N.Y. Academy of Sciences Conference on Management of Assessed Risk for Carcinogens, March 17-9, 1980; David Grether, "Bayes Rule as a Descriptive Model: The Representativeness Heuristic," Caltech Social Science Working Paper 245, Jan. 1979, Pasadena, Ca.; Talbot Page, Keeping Score: Actuarial Approach to Zero-Infinity Dilemmas" in Energy Risk Management (eds. G.T. Goodman and W.D. Rowe) Academic Press, New York, 1979.

For a spectacularly incorrect assessment of risk by one of the great statisticians see R.A. Fisher, Smoking -- The Cancer Controversy: Some Attempts to Assess the Evidence, Oliver and Boyd, London 1959.

7/ Straightforward but messy calculation shows that in a one-tailed Fisher exact test with (nominal) α equal to 5 percent, when there is a 2 percent background risk of cancer in the controls, for a particular site, and a five fold increase in the background incidence for that site due to the carcinogen, β is 54%.

8/ T.R. Fears, R.E. Tarone, and K.C. Chu, "Error Rates for Carcinogenicity Screens," Cancer Research, Vol. 37, 1941-5, July 1977. The tradeoff between α and β depends critically on the background rate of cancer in the controls, the magnitude of the effect of concern, and the decision rule. For a two dose experiment, for two sexes and two species, if our decision rule is to find an effect when the test is positive at any site for either dose for either sex, then the true α can be greatly elevated.

9/ Robert Harris, T. Page, and N. Reiches, "Carcinogenic Hazards of Organic Chemicals in Drinking Water," in Origins of Human Cancer, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1977; Talbot Page, R.H. Harris, and J. Bruser, "Removal of Carcinogens from Drinking Water; A Cost-Benefit Analysis," in Scientific Basis for Health and Safety Regulation, Brookings Institution, Washington, D.C., forthcoming; Talbot Page R.H. Harris, and S.S. Epstein, "Drinking Water and Cancer Mortality in Louisana, "Science, Vol. 193, 2 July 1976; Talbot Page and R.H. Harris, "Statistical Analysis of Cancer Mortality in Louisiana" in "The Implications of Cancer-Causing Substances in Mississippi River Water, "The Environmental Defense Fund, Washington, D.C., Nov. 1974.

10/ The results of a power analysis can be represented as a surface in three dimensional space with coordinates α, β, and the magnitude of effect. Figure 1 shows a slice of the surface with the magnitude of effect held constant, and Figure 2 shows another slice with α held constant.

11/ With the definition of D, we can say that a zero-infinity dilemma is a decision problem where both N and D are "high."

12/ For a discussion of how the underlying rarity dilutes the power of a test see Donald Weiner et al., "Correlations Among History of Angina," The New England Journal of Medicine, Vol. 301 No. 5, and the editorial in the same issue, which makes the application of Bayes Theorem. See also, D.R. Calkins, R.L. Dixon, C.R. Gerber, D. Zarin, and G.S. Omenn, "Identification Characterization, and Control of Potential Human Carcinogens: A Framework for Federal Decision-Making," JNCI, Vol. 64, No. 1 Jan. 1980, p. 172.

13/ Cited from the Wall Street Journal, Thursday, Jan. 19, p. 38.

14/ Letter from Carmen Guarino, to the editor of the American City Magazine, March 7, 1975.

15/ Calkins et al., "Identification," 1980.

16/ For a discussion of what is required to meet the judicial requirement of "substantial evidence in the rulemaking record taken as a whole," see the recent cotton dust decision, AFL-CIO v. Marshall, No. 78-1736; Cotton Warehouse Assn. v. Marshall, No. 78-1736 (Bazelon opinion).

17/ One could read, for example, the $21 million risk assessment of the risk of ozone depletion without a clear idea of what the scientists judged to be the level of suspicion for the principle effects of concern. U.S. Dept. of Transportation, Climatic Impact Assessment Program (1975) series of six monographs, available from National Technical Information Service, Springfield, Va.