# DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

# CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

PRIORITY SETTING FOR TESTING CHEMICALS

Gib Bogle and Talbot Page
California Institute of Technology

**SOCIAL SCIENCE WORKING PAPER 468**

February 1983

ABSTRACT

It is necessary to set priorities for testing chemicals because there are many thousand chemicals in commercial use, only a few thousand of which have been tested adequately, and limited resources with which to test them. Moreover, there are dozens of effects to test for, (e.g. carcinogenesis, mutagenesis, fetotoxicity) and hundreds of tests to choose from. The priority problem is to choose which chemicals to test first and to choose which tests to employ.

The approach of this paper is to design an optimal priority process (for a given year) by maximizing the value of information obtained from testing, subject to a budget constraint.

# Priority Setting for Testing Chemicals

Gib Bogle and Talbot Page[1]

This paper addresses the problem of prioritizing chemicals for testing. To be concrete, consider the problem faced by the National Toxicology Program. This program has a budget of about $60 million to test chemicals for toxic effects. There is a universe of about 70,000 commercial chemicals to choose from, and this universe might be expanded to include natural chemicals and inadvertent by-products. There are dozens of effects which might be tested for, including carcinogenesis, mutagenesis, teratogenesis, fetotoxicity, reproductive failure, nervous system disorder, and pulmonary disorder among others. And finally there are hundreds of tests and test systems, ranging in cost from about $100 to $750,000.

Loosely stated, the objective of prioritization is to choose chemicals to be tested and tests to be performed in a way to maximize the value of information obtained from testing, subject to the budget constraint. In our approach to this problem we combine ideas from decision theory and statistical estimation theory. From decision theory we take a decision tree approach, with judgemental probabilities and sequential analysis. From statistical estimation theory we take our criterion — a criterion analogous to minimizing an estimated variance.

The approach differs from others to priority setting in several ways. Our approach is to optimize the choice or design of the priority process (or to choose the best decision tree among a large number of possible decision trees). In Weinstein (5) the optimization is over a decision whether to test a chemical for a given priority process. In Weinstein's approach, which is probably the closest to ours, the criterion is to minimize the expected cost of final actions. This means that expected regulatory actions and inactions, along with the costs of chemical benefits foregone, must be estimated as a part of the priority process. By using a criterion of misclassification cost, the costs of final actions, which are expensive to estimate for many chemicals and highly uncertain at the time of prioritization, play a minor and indirect role. The justification for our approach is that gathering more than rudimentary information on the costs and benefits of final (regulatory) actions for 70,000 chemicals at the first stage of prioritization would not be efficient in terms of the expected value of this information.

To our knowledge, other existing approaches to prioritization for testing (1,2,3,4,6) do not attempt to optimize.

I. **Definition**. A priority process can be summarized as a decision tree, as follows.

$$S = \{s_1, s_2, \ldots, s_N\}$$ 
Set of N categories of exposure and toxicity to which a chemical may belong.

$$T = \{t_1, t_2, \ldots, t_K\}$$ 
Set of tests.[1] (This set includes the null test, which costs nothing and always has the same outcome.)

$$R_k = \{r_{k1}, r_{k2}, \ldots, r_{k,m(k)}\}$$ 
Set of m(k) possible results of test k.

An L stage priority process $\pi$ is defined by specifying a test for stage 1; a test at stage $\ell$ for each possible sequence of outcomes from the preceding $\ell - 1$ stages; and a recommended test for each outcome of the L stage priority process.

II. **Priority Processes Viewed as Classification Devices**. A test k is characterized by an m(k) by N array of probabilities, called the performance characteristics:

$$P(r_{ki}|s_j) = \text{probability of test k having outcome } r_{ki} \text{ when the true category of a chemical tested is } s_j.$$

---

[1] A test may be a battery of tests, a data element, or a battery of data elements.

The current state of knowledge about any universe (or subuniverse) of chemicals is represented by a probability distribution over the N chemical types or categories:

$P(s_j)$ = probability that a chemical drawn randomly from the universe of chemicals will fall into category $s_j$. $P(s_j)$ is the prevalence of exposure and toxicity of type j.

The probability that a chemical chosen at random from the initial collection (the universe) being put into the $i^{th}$ result category by the $k^{th}$ test is

$$\sum_{j=1}^{N} P(r_{ki}|s_j)P(s_j) .$$

A test k transforms a prior distribution $P(s_j)$ into a set of posterior distributions $P(s_j|r_{ki})$, $i = 1,...m(k)$.

The increase in our level of information about the chemicals is manifested in the increased localization (decreased standard deviation) of each of the posterior distributions in comparison to the prior distribution. As a diagrammatic illustration, consider a continuous distribution f(x) and a test with two possible results, as shown in Figure 1. The sum of the two posterior distributions, weighted by the total probabilities of each result, is the prior distribution.
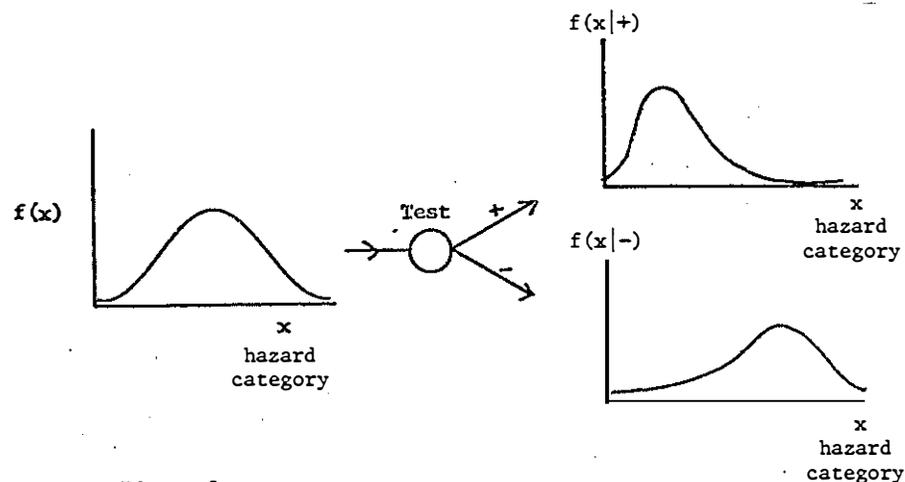
Figure 1. A simple test.

As long as tests yield information, repeated tests will produce a proliferation of outcome categories of chemicals with ever sharpening probability distributions. Eventually, if each test conveys information, each of the multitude of distributions would have zero standard deviation, in other words the true type of every chemical would be known with certainty. The existence of testing costs and a budget constraint means that such a perfect sorting cannot be achieved. The aim of a prioritization and testing program is to sort, as well as possible for a given budget, the universe of chemicals under consideration into the N types of true classification.

III. <u>Comparability and Misclassification Cost</u>. Different priority processes will misclassify chemicals in different ways. Without a specification of the severity or cost of a misclassification there is no way to compare priority processes and to recommend one over another.

To achieve comparability among alternative priority processes, a concept of misclassification cost is now developed. A misclassification cost (as distinct from a monetary cost) is incurred when a chemical is finally classified as being of a certain type, which is not the true type of the chemical. The final classification of a chemical occurs when, on the basis of the probability distribution associated with the chemical (which reflects the results of the tests to which the chemical has been subjected), the decision is made to terminate testing on it and to label it as a certain type. Write $\hat{s}_i$ to mean that a chemical is classified, by the process, as in category $s_i$. The misclassification of a chemical classified as $\hat{s}_i$ when its true type is $s_j$ is written $C(\hat{s}_i|s_j)$. This cost is zero when $i = j$.

IV. <u>Budget Cost</u>.

A second requirement for comparability is that the budget cost of a priority process be defined. The expected cost (per chemical) is specified as follows. A priority process $\pi$ is characterized by a set of paths $\pi_i$; $i = 1, M^\pi$ where $M^\pi$ is the number of paths. Each path corresponds to a sequence of tests and test results.
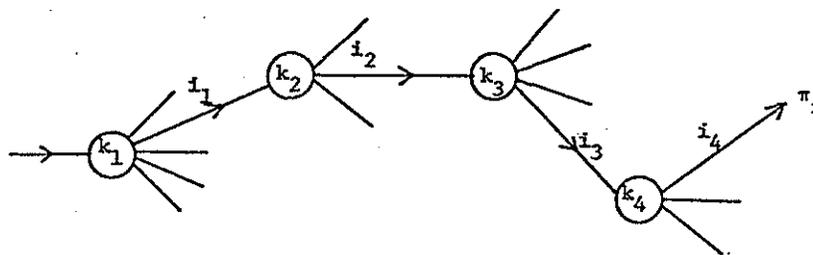
Figure 2. A path with 4 tests.

In the example shown in Fig. 2, the path $\pi_i$ is the sequence of (test, result) pairs $(k_1,i_1)$, $(k_2,i_2)$, $(k_3,i_3)$, $(k_4,i_4)$.

Considering the test scheme to be one big test, the paths $\left\{\pi_i; \ i = 1 \ldots M^\pi\right\}$ can be viewed as defining the $M^\pi$ test outcomes. As in the case of a simple test, the scheme $\pi$ is characterized by the conditional probabilities: $P(\pi_i|s_j)$.

If the path $\pi_i$ is the sequence $(k_1,i_1)$, $(k_2,i_2),\ldots,(k_n,i_n)$, of tests and results, then

$$P(\pi_i|s_j) = P(r_{k_1 i_1}|s_j)P(r_{k_2 i_1}|s_j)\ldots P(r_{k_n 1_n}|s_j)$$

The dollar testing cost $C_{\pi_i}$ associated with a path $\pi_i$ is simply the sum of the costs of the tests along the path, that is

$$C_{\pi_i} = CT_{k_1} + CT_{k_2} + \ldots + CT_{k_n}, \quad \text{where } CT_k \text{ is the cost of test k.}$$

Rather than a single testing cost, the scheme $\pi$ has a vector of expected costs $\left\{E_j^\pi; \quad j = 1\ldots N\right\}$, since chemicals of different types follow the various paths with different probabilities:

$$E_j^\pi = \sum_{i=1}^{M^\pi} P(\pi_i|S_j)C_{\pi_i} = \sum_{i=1}^{M^\pi} P(\pi_i|S_j)(CT_{k_1} + CT_{k_2} + \ldots CT_{k_n}).$$

The total expected cost ($TCT^\pi$) of process $\pi_i$ per chemical is

$$TCT^\pi = \sum_j^N P(S_j)E_j^\pi$$

## V. Minimizing Misclassification Cost.

For an initial probability distribution $P(s_j)$ the probabilities associated with outcome categories of a process $\pi$ are

$$P(\pi_i|s_j)P(s_j); \quad i = 1,\ldots M^\pi, \quad j = 1,\ldots N.$$

The best final classification of chemicals in outcome category i is that which minimizes expected misclassification cost. The minimum misclassification cost for path $\pi_i$ is defined as

$$CM_i^\pi = \underset{t=1\ldots N}{Min} \sum_j P(\pi_i|s_j)P(s_j)C(\hat{s}_t|s_j)$$

and $t^o$ is the choice of t which minimizes this cost.

The total expected misclassification cost for policy $\pi$ is

$$TCM^\pi = \sum_{i,j} CM_i^\pi P(\pi_i|s_j)P(s_j)$$

per chemical.

The foregoing shows how a priority process $\pi$ acting on an initial collection of chemicals described by a probability distribution $P(s_j)$ incurs a misclassification cost $TCM^\pi$ and a testing cost $TCT^\pi$. Defining the budget limit as B, we now state the priority problem in closed form.

The optimal process is that which minimizes $TCM^\pi$ subject to $TCT^\pi \leq B$.

The above statement of the priority problem has been made operational in terms of a computer program, which chooses the optimal priority process for a given set of parameters. Thus parameters are: (1) estimates of the distribution of toxicity and exposure over the universe of chemicals $P(s_j)$; (2) the performance characteristics of tests $P(r_{ki}|s_j)$; (3) dollar costs of tests $CT_k$; and (4) misclassification costs $C(\hat{s}_i|s_j)$.

1.  Astill, B.D., H.B. Lockhart, Jr., J.B. Moses, A.N.M. Nasr, R.L. Raleigh, and C.J. Terhaar. 1981. Sequential testing for chemical risk assessment. In R. Conway, Ed. Environmental Risk Analysis of Chemicals. second International Congress of Toxicology, Brussels, 1978. New York: Van Nostrand-Reinhold Company. (in press)

2.  Cramer, G.M., R.A. Ford, and R.L. Hall. 1978. Estimation of toxic hazard - A decision tree approach. Food Cosmet. Toxicol. 16: 255-276.

3.  Gori, G.B. 1977. Ranking of environmental contaminants for bioassay priority, pp. 99-111. In U. Mohr, D. Schmahl, L. Tomatis, and W. Davis, Eds. Air Pollution and Cancer in Man. IARC Scientific Publications No. 16. Lyon: International Agency for Research on Cancer.

4.  Nisbet, I.C.T. 1979. Ranking Chemicals for Testing: A priority-setting exercise under the toxic substances control act, pp. B-41--B-54. In Scoring Chemicals for Health and Ecological Effects Testing. TSCA-ITC Workshop. Rockville, Md.: Enviro Control, Inc.

5.  Weinstein, M.C. 1979. Decision making for toxic substances control. Public Policy 27(3), 333-383.

6.  Wilhelm, S. 1981. Selecting Priorities from Large Sets of Alternatives: The Case of Toxic Substances Regulation. Dissertation. Providence, R.I.: Brown University, Department of Chemistry. 259 pp.