JOURNAL OF THE AMERICAN COLLEGE OF TOXICOLOGY
Volume 7, Number 5, 1988
Mary Ann Liebert, Inc., Publishers

Screening Toxic Chemicals: How Accurate Must Tests Be?

L.B. LAVE[1] and G.S. OMENN[2]

ABSTRACT

A decision analysis framework is used to explore the value of screening tests for carcinogenicity. Whether a test lowers the social cost of screening depends on the test's sensitivity, specificity, and cost and the social cost of misclassification (exonerating carcinogenic chemicals or condemning noncarcinogenic chemicals). The model shows that the best screening test need not be either the most accurate or the least expensive.

INTRODUCTION

MANAGING POTENTIALLY TOXIC CHEMICALS has become one of the most controversial and difficult tasks facing government, corporations, and society more generally. About 60,000 chemicals are in common use, and about 1000 new chemicals are introduced each year.[1,2] Only a small proportion of these chemicals has been subjected to thorough toxicological examination.[1] New chemicals often have been approved by the Environmental Protection Agency despite very limited data.[3] However, during 1986, EPA began to take a more aggressive stance. The agency began using Section 5e of the Toxic Substances Control Act to require labeling as a "potential carcinogen" any new chemical from a class containing known carcinogens. Companies are permitted to market these new chemicals, but the chemicals must carry this label until adequate testing has demonstrated noncarcinogenicity.

The public has displayed intense concern for health and environmental effects from potentially toxic chemicals, particularly carcinogenic effects.[4] The manifestation of this concern includes such federal legislation as the Toxic Substances Control Act, the Federal Insecticide, Fungicide, and Rodenticide Act, and the recent expansion of Superfund. Even more stringent legislation, Proposition 65, was enacted in California in November 1986 as a result of a citizens' initiative.

The management of potentially toxic chemicals is a broad topic treated elsewhere.[5-8] (For a discussion of the underlying issues, see Refs. 9–12.) Here, we examine the role of toxicology in determining the carcinogenicity of chemicals. Specifically, we examine the value of information from rodent and short-term bioassays.

[1]Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, Pennsylvania.
[2]School of Public Health and Community Medicine, University of Washington, Seattle, Washington.
Presented at the Seventh Annual Meeting of the American College of Toxicology, November 1986, Philadelphia, Pennsylvania.

Since only a small fraction of the 60,000 commonly used chemicals have been tested and since these chemicals continue to be commonly used, society implicitly assumes either that the untested chemicals are innocuous or that their risks are acceptable. Either assumption might be a serious mistake.

Since screening tests are less than perfect, some of the chemicals screened will be classified incorrectly. Some carcinogens would test positive, would be considered carcinogens, and thus be true positives (TP). However, some would test negative, be considered noncarcinogens, and thus be false negatives (FN). Similarly, some noncarcinogens would test negative, be considered noncarcinogens, and be true negatives (TN), and some would test positive, be considered carcinogens, and be false positives (FP). The social cost of a screening program includes both the direct testing expenditures and the consequences of misclassification (both FN and FP).

IDENTIFYING CARCINOGENS

At present, uncertainty dominates the classification of chemicals. Every approach has major limitations. Epidemiology fundamentally studies past social ignorance and human tragedy. For epidemiology to be informative, large numbers of people must have been exposed to a toxic chemical, with many suffering measurable harm. Society wants to prevent people from being exposed to harmful levels of toxic chemicals and to prevent the development of statistically significant increases in cancer incidence rather than to wait to measure those that have occurred. Thus, epidemiology cannot serve as the principal tool for informing decision makers. Early identification without harm to people requires relying on toxicology. Biological markers, such as DNA adducts and preneoplastic lesions in certain tissues, hold the promise of detecting significant exposures and possible precursors to serious disease at an early enough stage that the disease process can be reversed. If so, a hybrid of epidemiology and toxicology may become a powerful approach.[13-16]

For carcinogenicity testing, toxicologists have focused on the long-term feeding study. This procedure is expensive (over $1 million for the NTP protocol), time consuming (at least 3 years), has many difficulties of interpretation (such as deciding how to treat benign tumors or tumors in organs that humans do not have, such as the rat's zymbal gland), and inevitably has been compromised by problems in execution (such as disease, heat or cold, or being given the wrong dose). With the standard protocol of 50 animals per group, long-term bioassays are not as powerful statistically as would be desired in classifying carcinogens. In addition, there is the difficulty of generalizing from rodents at high exposure to humans at low exposure.[17] For an anxious public, the principal drawback of long-term feeding studies is that it is not economically or physically possible to test all the chemicals in common use (the cost would be $60 billion) or even all new chemicals.

Short-term in vivo and in vitro tests offer the prospect of a superior alternative. A battery of screening tests can be quick (days to a few weeks) and cheap ($5000–$10,000) to perform. These tests can provide information on mechanisms of action – an advantage over black box tests, such as long-term feeding studies.[18] Their largest drawback is that most involve cultured cells or nonmammalian systems, rather than an intact animal with its toxicokinetic and immune systems. Certain classes of chemicals, such as mineral fibers, arsenic, and heavy metals, generally test negative in short-term tests, even though they are human carcinogens. Chemicals that may act as tumor promoters may be hard to detect with current assays.[19,20] Thus, it is important to recognize that some potential carcinogens will be missed by present short-term tests. The point is that no current test correctly classifies the toxicity of all chemicals; epidemiology is limited by the small numbers of exposed people, low exposures, and confounding factors.

Toxicologists have searched for a single short-term test that would identify carcinogens.[21] However, since there are several mechanisms for damaging DNA, as well as carcinogens that do not act directly on DNA, this search for a single test is simplistic. Rather, one needs to construct a battery of tests, at least one for each major mechanism.[22-28] Each of the proposed test batteries appears capable of classifying possible carcinogens with reasonable accuracy.[29]

WHEN IS SCREENING WORTHWHILE?

Suppose there are s chemicals, of which c is the proportion that are carcinogens and $1 - c$ is the proportion of noncarcinogens. Suppose that screening (either a single test or a battery) correctly classifies the proportion p of carcinogens and the proportion q of noncarcinogens. If so $(1 - p)cs$ will be the number of false negatives, and $(1 - q)(1 - c)s$ will be false positives. If there were no screening, all cs of the carcinogenic chemicals would be unrecognized.

Now suppose that the average social cost of each false negative or unrecognized carcinogen is n and the average social cost of each false positive is m. Also suppose that the cost of running the screening battery on each chemical is t. Then the cost of not screening is csn. The cost of screening all chemicals is the testing cost for each chemical (t) multiplied by the number of chemicals tested (st) plus the number of FN (and unrecognized carcinogens) times the cost of each plus the number of FP times the cost of each.

$$st + (1 - p)csn + (1 - q)(1 - c)sm$$

If screening is worthwhile, the cost of screening plus the costs of false negatives and false positives must be less than the social costs of failing to identify and control all carcinogenic chemicals.

$$[st + (1 - p)csn + (1 - q)(1 - c)sm] < csn \quad [1]$$

Dividing both sides of the inequality by csn yields

$$[t/(cn) + (1 - p) + (1 - q)(1 - c)m/(cn)] < 1 \quad [2]$$

The crucial parameters are t/n and m/n. The former represents the ratio of the cost of testing each chemical to the cost of a false negative; if that ratio is small, it indicates that testing is a small fraction of the value of discovering that a chemical is a carcinogen. The latter is the ratio of the cost of a false positive to that of a false negative; since FN can be reduced only at the cost of more FP (as elaborated subsequently), this ratio is essential for classification.[29]

SIMULATION EXERCISES

It is helpful to assume values for these parameters and to display their implications. Let us assume that the universe of chemicals in common use is 50,000 and that either 2, 10, or 50% of the chemicals are carcinogens. We simply do not know what proportion of the commonly used chemicals are carcinogens. The first two estimates are judgments based on the assumption that carcinogens are largely known already. The last assumption comes from the NTP experience in long-term bioassays. We suppose that the cost of testing a chemical is 1/10 the average cost of an FN using a long-term test and that the cost is 1/1000 using a battery of short-term tests. Furthermore, suppose that an FP (condemning a benign chemical) costs society only 10% of the cost of an FN (identifying a carcinogenic chemical as benign).

If 2% of chemicals are carcinogens ($c = 0.02$), this analysis can be applied to the long-term test by substituting the assumed values into Equation 2.

$$[5 + (1 - p) + (1 - q)(1 + 0.98(5))] < 1$$

This expression requires that both p and q be greater than 1.68, yet it is impossible to have a classification accuracy greater than 1.0. Thus, under the assumptions that a screening test costs 10% as much as the social cost of an FN and that only 2% of chemicals are carcinogens, screening cannot be better than doing no screening even if the test is completely accurate.

More generally, for screening to be better than not screening, $t/cn < 1$. Thus, if a test costs 10% as much as the cost of an FN ($t/n = 0.1$), more than 10% of chemicals must be carcinogens ($c >$ 0.1 for the test to be worthwhile, assuming the test is completely accurate. Since we regard it as unlikely that more than 10% of chemicals are carcinogens,[29] we conclude that a test is too expensive to be used as a screening test if it costs 0.1 as much as the social value of discovering a carcinogen. Thus, the long-term test would be inappropriate for screening even if it has 100% classification accuracy.

For short-term tests, assuming $p = q$ and $c = 0.02$, classification accuracy of carcinogens and noncarcinogens ($p = q$) must exceed 0.84. The larger the proportion of chemicals that are carcinogens, the less accurate need be the test (p). If 10% of chemicals were carcinogens, classification accuracy need be only 0.48, since the cost of not testing (of missing all carcinogens) is five times higher. Thus, this logical analysis illustrates the potential value of short-term test batteries as screening devices.

Since the large majority of chemicals are assumed to be noncarcinogenic, the reliability of a test battery depends heavily on accurate classification of noncarcinogens (specificity). For example, with 10% carcinogens ($c = 0.02$), if classification accuracy p for carcinogens is 0.9 instead of 0.84, classification accuracy (q) still must be 0.83 for noncarcinogens. Even for $p = 1.0$, q is reduced only from 0.84 to 0.81. In contrast, if the classification accuracy for noncarcinogens were 0.9, p need only be 0.54 for t/cn to be less than 1. If 10% of chemicals are carcinogens, the classification accuracy for carcinogens becomes more important.

Equation 2 identifies the tradeoff between cost of a screening test and its accuracy. A costly test is of no value for screening, even if it is completely accurate; an inexpensive test is not worth using unless its accuracy is fairly high. Even if a test were free, its classification accuracy would have to be at least 83% in order to be useful, assuming $c = 0.02$. If 10% of chemicals were carcinogens, classification accuracy would still have to exceed 82%.

If 50% of chemicals were carcinogens, using the lifetime bioassay for screening makes good sense; the assay need have accuracy ($p = q$) of only 27% in order to be worth using. Not surprisingly, identifying rare problem chemicals only makes sense with an inexpensive test, whereas identifying chemicals that comprise half the population can be achieved efficiently with an expensive test.

When 50% of chemicals are carcinogens, the less expensive test need only have an accuracy ($p = q$) of 9% for screening to make a contribution. Even if the accuracy in classifying noncarcinogens is zero, the accuracy for carcinogens need only be slightly greater than 10%. More generally, classification accuracies can be much lower when carcinogens are 50% of all chemicals.

Results are summarized in Table 1. The required p and q values are shown first for the case when 2% of chemicals are carcinogens, then for the 10% and 50% cases.

CONSTRUCTING A SCREENING TEST BATTERY

Suppose that adding a test to a battery that would increase classification accuracy (with $p = q$) from 0.84 to 0.85 (i.e., it increases accuracy by 0.01 or decreases $1 - p$ from 0.16 to 0.15). How much could this test increase the cost of the battery and still be worthwhile? According to the equations given, for $c = 0.02$, the answer is that it could more than double the cost of testing each chemical (increase the ratio of t/n from 1/1000 to 2.2/1000).

The net social value of testing is the social value of detecting all true carcinogens less the cost of testing plus the social cost of misclassification if all chemicals are tested.

$$v = csn - ts - (1 - p)csn - (1 - q)(1 - c)sm \quad [3]$$

The formula is simplified by dividing by s.

$$V/s = cn - t - (1 - p)cn - (1 - q)(1 - c)m \quad [4]$$

TABLE 1. SIMULATION RESULTS FOR SCREENING TESTS[a]

% Carcinogens	Classification accuracy for Carcinogen		Noncarcinogen	Relative screening cost	Characterization of each case
c	p	$p = q$	q	t/n	
2		1.68		1/10	NTP assay for screening
2		0.84		1/1000	Short-term screening tests
2	0.9		0.83	1/1000	Increased sensitivity
2	1.0		0.81	1/1000	Perfect sensitivity
2	0.54		0.9	1/1000	Increased specificity
2	0.05		1.0	1/1000	Perfect specificity
2		0.83		0	Zero cost for screening
10		1.0		1/10	NTP assay for screening
10		0.48		1/1000	Short-term screening tests
10	0.91		0	1/1000	Zero specificity
10	0.1		0.9	1/1000	Enhanced specificity
10	0		1.0	1/1000	Zero sensitivity, perfect specificity
10		0.005		0	Zero cost for screening
50		0.27		1/10	NTP assay for screening
50		0.09		1/1000	Short-term screening tests
50	0.10		0	1/1000	Zero specificity
50	0.002		1	1/1000	Perfect specificity

[a]The calculations are independent of the number of chemicals to be tested. The cost of an FN relative to that of an FP (n/m) is assumed to be 10.

Now, suppose classification accuracy can be improved from p to p^* and from q to q^*. Then the increase in social value is

$$V^*/s - V/s = (p^* - p)cn + (q^* - q)(1 - c)m$$

If we set $p^* - p = q^* - q = 0.01$, take $c = 0.02$, $n = \$10{,}000{,}000$, and $m = \$1{,}000{,}000$, we get

$$(V^*/s - V/s) = 0.01(20 + 98)10{,}000 = 11{,}800$$

or

$$V^* - V = 11{,}800s$$

If the classification accuracy for carcinogens alone improved by 0.01 ($p^* - p = 0.01$), the cost of testing could increase \$2000 for each chemical. If the classification accuracy for noncarcinogens alone improved by 0.01 ($q^* - q = 0.01$), the cost of testing could increase \$9800 per chemical. Thus, as modeled above, an improvement in accuracy is worth almost five times more for noncarcinogens than for carcinogens (for $c = 0.02$).

If 10% of chemicals are carcinogens, a 0.01 increase in classification accuracy for both carcinogens and noncarcinogens is worth an additional \$19,000 in testing cost. An 0.01 increase in the classification accuracy for carcinogens alone would be worth $10,000 per chemicals; the same increase for noncarcinogens alone would be worth $9000 per chemical. Thus, when 10% of chemicals are carcinogens, an increase in classification accuracy is worth about the same amount for carcinogens as for noncarcinogens.

The main conclusion from this analysis is that a small increase (0.01) in classification accuracy is worth a large increase (118% for $c = 0.02$) in testing cost. Since short-term tests are relatively inexpensive to perform (often a few thousand dollars), it appears to be worthwhile to add additional tests that give any measurable increase in accuracy.

STATISTICAL ISSUES

When a single test is used, the interpretation of test results is obvious: if the test is positive, the chemical should be regarded as a carcinogen; if the test is negative, the chemical should be exonerated, although it might still be suspect. However, the interpretation of test data is far from simple.[30] Neither the testing protocols nor the criteria for interpretation of the data have been standardized adequately, much less validated extensively across laboratories.

When a test battery is used, the interpretation is more complicated. If any of the tests is positive, should the chemical be classified as a carcinogen? Do some of the tests have greater predictive value than others? Should congruence of test results according to underlying mechanism be considered? Required?

Two steps are required to translate test results into a carcinogenicity classification. (For a different approach, see Refs. 31–33.) The first is to construct an index number (or weighted average) of the test results. Under particular conditions, this index number can be interpreted as the probability that a chemical is a carcinogen (given these test results). Several statistical techniques are available for fitting qualitative (0, 1) dependent variables, including logit and probit analysis.[34]

We fit a logit regression to data from the International Collaborative Programme for the Evaluation of Short-Term Tests for Carcinogenicity.[29] The resulting estimates provided valuable information, including the relative importance of each test and the formula for estimating the probability that each chemical is a carcinogen. The resulting logit scores span the full range from zero to one.

The second step is deciding how large a probability is required to classify a chemical as a carcinogen. If society were determined that no chemical would ever falsely receive a clean bill of health, the cut point would have to be set at zero, i.e., all chemicals would have to be classified as carcinogens even if all tests were negative. If so, the vast majority of chemicals would, inappropriately, be classified as carcinogens. Of course, society has not banned, regulated, or even tested all chemicals, so there is no indication that the cost of an unrecognized carcinogen is regarded as so high that all chemicals are treated as carcinogens. The higher the value of the cut point, the greater the number of FN, but the lower the number of FP. Thus choosing a cut point requires a decision analysis of the social cost of misclassification.[9,10,12,29] How much does society lose for each FP? Each FN?

These concepts are illustrated in Figure 1. Assume that a logit equation for a test battery has been fitted to data on chemicals whose carcinogenicity is known. The estimated logit score is the horizontal axis in the graph, and the vertical axis is the frequency of chemicals having this logit score. The scores for all known noncarcinogens are shown as the solid curve. Many of these chemicals would have a logit score of zero or near zero, and only a few would have scores near 1. The scores for all chemicals known to be carcinogens are shown as the dotted curve. Few of these chemicals would have scores near zero and most would have scores near 1. Suppose the cut point is chosen as 0.5. The solid curve to the right of the cut point would represent noncarcinogens falsely classified as carcinogens (noncarcinogens with a score greater than the cut point). The dotted curve to the left of the cut point would represent carcinogens falsely given a clean bill of health (carcinogens with a score lower than the cut point). A perfect battery would be one that gave scores to all carcinogens greater than a cut point and scores to all noncarcinogens less than the cut point. However, for currently available tests and batteries, there must always be some FP and FN.

We have applied this method to data from an International Collaborative Program,[29,35] as shown in Figure 2. The 38 chemicals graphed in Figure 2 consisted of 12 noncarcinogens (shown as 0) and 26 carcinogens (shown as X). Each chemical was tested with up to 30 short-term assays, several of which were performed in more than one laboratory. Using logit analysis, we fit each chemical (0 for noncarcinogens and 1 for carcinogens) to its individual results on three assays (*Salmonella* plate assay, differential killing of *Bacillus subtilis*, and mammalian cell transformation). Having estimated the coefficients for each assay (the weights to compute a weighted average), we computed the logit score, which is interpreted as the probability that this chemical is a carcinogen. These scores are graphed in Figure 2. As expected, the noncarcinogens are concentrated near zero (all but one has a probability of 0.41 or less), whereas the carcinogens are concentrated near 1 (all but three have a probability of 0.81 or more). Shown on the graph are two cut points: alpha = 1 lies between 0.41 and 0.81 and corresponds to the assumption that the cost of an FN equals that of an FP; alpha = 10 lies below 0.05 and corresponds to the assumption that the cost of an FN is 10 times that of an FP. For the 38 chemicals tested, several of which were chosen because they were carcinogens known to be negative on the Ames test, classification accuracy is very high (81%). For cut point alpha = 1, there are three FN and one FP.

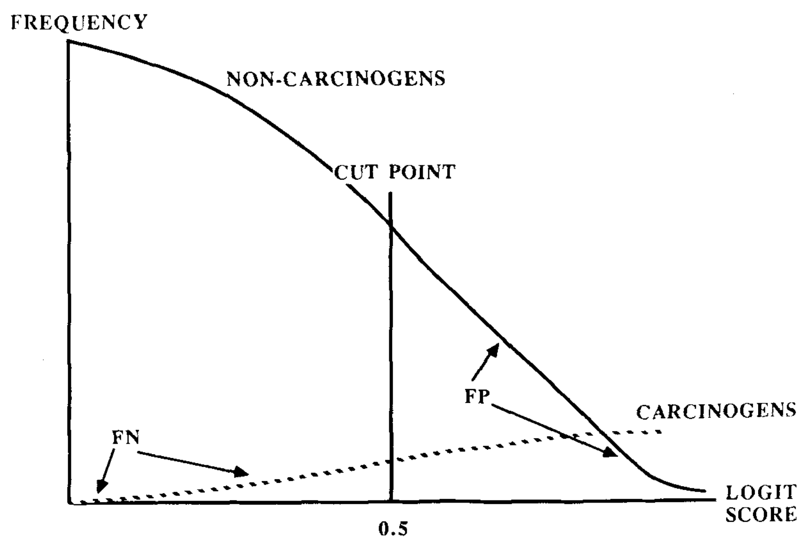

FIG. 1. Classifying chemicals using logit scores.

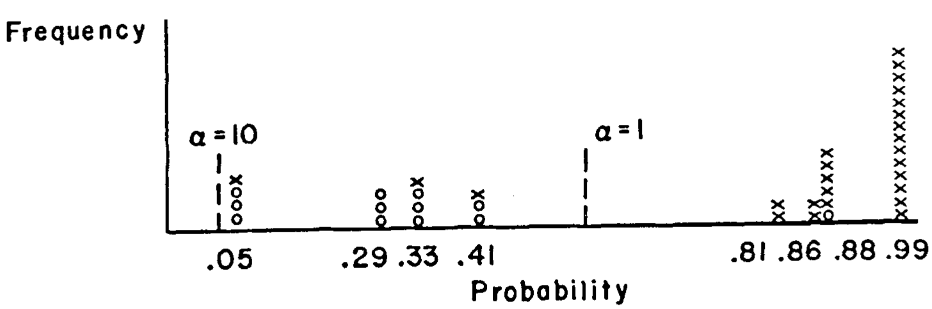

FIG. 2. Predicted probability of carcinogenicity.

The statistical analysis also can estimate the cost-effectiveness of each test, the amount of classification accuracy it adds. The most desirable points are those in the lower right corner of Figure 2, since they have the greatest classification accuracy and the least cost. Shown in Figure 3 is a graph of the cost of certain tests plotted against classification accuracy. In our analysis of the International Collaborative Study data, test 13 (*Salmonella* plate assay) was both more expensive and less accurate than test 21 (*B. subtilis* repair). Using both tests 13 and 54 (mammalian cell transformation) gave more information (but at greater expense) than any single test, but this pair of tests is less cost-effective than a battery of tests 13, 21, and 54. These results are tied to the particular set of chemicals tested, of course. With a different set of chemicals, test 13 might give better classification accuracy than test 21.

MANAGING TOXIC CHEMICALS

The public, Congress, regulators, and toxicologists have underestimated the difficulties associated with managing 60,000 potentially toxic chemicals. A major impediment has been the identification of carcinogens, where progress has been slow because of the expense and time required to perform long-term bioassays. Our decision analytic framework isolates the circumstances under which screening tests can be useful as a function of test cost, sensitivity, specificity, and the proportion of chemicals that are carcinogens. The most accurate test is not necessarily the most desirable one to use, just as the least-cost test need not be most desirable. Short-term, in vivo and in vitro, tests have a major role to play in screening. Understanding the role of testing in managing risks from carcinogenic chemicals contributes to setting research priorities in toxicology and in designing testing protocols.

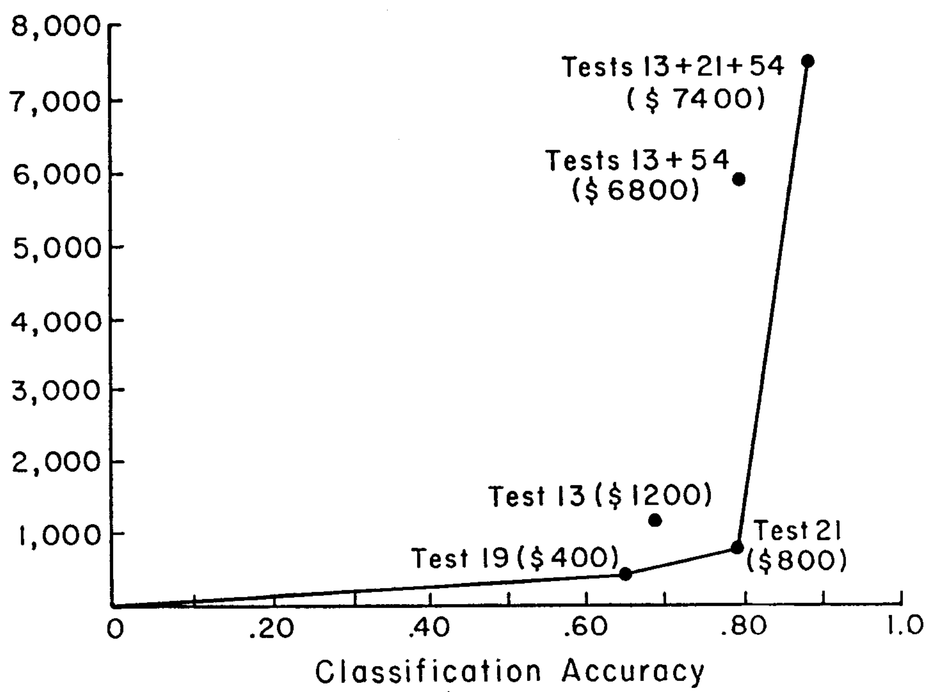

FIG. 3. Classification accuracy at least cost (38 chemical data set).

REFERENCES

1. NATIONAL RESEARCH COUNCIL. (1984). *Toxicity Testing: Strategies to Determine Needs and Priorities.* Washington, DC: National Academy Press.
2. LAVE, L.B., and UPTON, A.C. (eds.). (1987). *Toxic Substances, Health, and the Environment.* Baltimore: Johns Hopkins University Press.
3. US GENERAL ACCOUNTING OFFICE. (June 15, 1984). *Assessment of New Chemical Regulation Under the Toxic Substances Control Act*, GAO/RCED 84-84. Washington: US General Accounting Office.
4. LAVE, L.B. (1987). Health and Safety Risk Analyses. Science **236,** 291–295.
5. LAVE, L.B. (1985). Improving the management of environmental health. Environ. Health Perspect. **62,** 359–363.
6. RUCKELSHAUS, W. (1985). Risk, Science, and Democracy. Issues Sci. Technol. **1,** 19–38.
7. COVELLO, V.T., MENKES, J., and MUMPOWER, J. (1986). *Risk Analysis and management.* New York: Plenum.
8. OMENN, G.S. (1986). A framework for risk assessment for environmental chemicals. Washington Public Health **6,** 2–6.
9. PAGE, T. (1978). A generic view of toxic chemicals and similar risks. Ecology Law Q. **7,** 207–244.
10. WEINSTEIN, M.C. (1979). Decision making for toxic substances control: Cost-effective information development for the control of environmental carcinogens. Public Policy **27,** 333–343.
11. BOGLE, G., and PAGE, T. (1982). Priority setting for testing chemicals. In: *Safety Evaluation and Regulation of Chemicals.* F. Homburger (ed.). Boston: S. Karger.
12. LAVE, L.B., OMENN, G.S., HEFFERNAN, K.D., and DRANOFF, G. (1983). A model for selecting short-term tests of carcinogenicity. J. Am. Coll. Toxicol. **2,** 125–130.
13. WEINSTEIN, I.B. (1983). The monitoring of DNA adducts as an approach to carcinogen detection. Annu. Rev. Public Health **4,** 409–413.
14. HOEL, D.G., MERRILL, R.A., and PERERA, F.P. (1985). *Risk Quantification and Regulatory Policy.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, Banbury Report 19.
15. HATTIS, D.B. (1986). The promise of molecular epidemiology for quantitative risk assessment. Risk Analysis **6,** 181–194.
16. PERERA, F. (1986). New approaches in risk assessment in carcinogens. Risk Analysis **6,** 195–202.
17. US OFFICE OF SCIENCE AND TECHNOLOGY POLICY. (1985). Federal Register **50,** 10371–10442.
18. OMENN, G.S. (1984). Characterizing risks: Utilizing knowledge about mechanisms. In: *Molecular and Cellular Approaches to Understanding Mechanisms of Toxicology.* A.H. Tashjian, Jr. (ed.). Boston: Harvard School of Public Health, pp. 224–245.
19. ASHBY, J., DE SERRES, F.J., DRAPER, M.H., ISHIDATE, M., Jr., MARGOLIN, B.H., MATTER, B.E., and SHELBY, M. (eds.). (1985). *Evaluation of Short-Term Tests for Carcinogens: Report of the International Programme on Chemical Safety Collaborative Study on in Vitro Assays. Progress in Mutation Research.* Amsterdam: Elsevier Science Publishers, Vol. 5.
20. WORLD HEALTH ORGANIZATION. (1985). *Summary Report on the Evaluation of Short-Term Tests for Carcinogens (Collaborative Study on in Vitro Tests).* Environmental Health Criteria 47, Geneva: WHO.
21. AMES, B.N., McCANN, J., and YAMASAKI, E. (1975). Methods for detecting carcinogens and mutagens with the *Salmonella*/mammalian-microsome mutagenicity test. Mutat. Res. **31,** 347–364.
22. BRIDGES, B.S. (1976). In: *Screening Tests in Chemical Carcinogenesis.* R. Montesano, H. Bartsch, and L. Romatis (eds.). Lyon, France: International Agency for Research on Cancer.
23. MARCUS, W.L. (1978). In: *Carcinogenesis: A Comprehensive Survey, Vol. 3: Polynuclear Aromatic Hydrocarbons.* P.W. Jones and R.I. Fruendenthal (eds.). New York: Raven.
24. NATIONAL TOXICOLOGY PROGRAM. (1980). Technical Bulletin **1,** 2.
25. PURCHASE, I.F. (1982). International commission for protection against environmental mutagens and carcinogens. Mutat. Res. **99,** 53–71.
26. WEISBURGER, J.H., and WILLIAMS, G.M. (1981). Carcinogen testing: Current problems and new approaches. Science **214,** 401–407.
27. WILLIAMS, G.M. (1980). Batteries of short-term tests for carcinogen screening. In: *The Predictive Values of Short-Term Screening Tests in Carcinogenicity Evaluation.* G.M. Williams et al. (eds.). Amsterdam: Elsevier.
28. WILLIAMS, G.D., LASPIA, M.F., and DUNKEL, V.C. (1982). Reliability of the hepatocyte primary culture/DNA repair test and testing of coded carcinogens and non-carcinogens. Mutat. Res. **97,** 359–370.

29. LAVE, L.B., and OMENN, G.S. (1986). Cost-effectiveness of short-term tests for carcinogenicity. Nature **324**, 29–35.
30. BRESLOW, N., and KALDOR, J. (1985). Statistical approaches to the estimation and testing of mutagenic potency with data from in vitro tests. IARC Workshop paper.
31. CHANKONG, V., HAIMES, Y.Y., ROSENKRANZ, H., and PET-EDWARDS, J. (1985). The carcinogenicity prediction and battery selection (CPBS) method: A bayesian approach. Mutat. Res. **153**, 135–166.
32. PET-EDWARDS, J., ROSENKRANZ, H.S., CHANKONG, V., and HAIMES, Y.Y. (1985). Cluster analysis and predicting the carcinogenicity of chemicals using short-term assays. Mutat. Res. **153**, 167–185.
33. PET-EDWARDS, J., CHANKONG, V., ROSENKRANZ, H.S., and HAIMES, Y.Y. (1985). Application of the carcinogenicity prediction and battery selection (CPBS) method to the gene-tox data base. Mutat. Res. **153**, 187–200.
34. GOLDSTEIN, M., and DILLON, W.R. (1978). *Discrete Discriminant Analysis.* New York: Wiley.
35. DeSERRES, F.J., and ASHBY, J. (eds.). (1981). *Evaluation of Short-Term Tests for Carcinogens: Report of the International Collaborative Program.* New York: Elsevier.

Address reprint requests to:
Lester B. Lave, Ph.D.
Professor of Economics
Graduate School of Industrial Administration
Carnegie-Mellon University
Pittsburgh, PA 15213