

PROBLEMS IN VALIDATING THE AMES ASSAY  
AS A PREDICTOR OF HUMAN ENVIRONMENTAL CARCINOGENIC RISK

BY



CLYDE HERTZMAN, B.Sc., M.D.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Master of Science

McMaster University

April 1981

PROBLEMS IN VALIDATING THE AMES ASSAY  
AS A PREDICTOR OF HUMAN ENVIRONMENTAL CARCINOGENIC RISK

MASTER OF SCIENCE (1981)  
(Medical Sciences)

MCMASTER UNIVERSITY

TITLE: Problems in Validating the Ames Assay as a  
Predictor of Human Environmental Carcinogenic  
Risk

AUTHOR: Clyde Hertzman, B.Sc., M.D.  
McMaster University  
Hamilton, Ontario  
Canada

SUPERVISOR: Professor R.S. Roberts

NUMBER OF PAGES: 140

## ABSTRACT

The Ames Assay is a battery of tests which measures the mutagenicity of chemical substances. In order for its results to be relevant to human health, Ames Assay mutagenicity must relate to human cancer in a biologically credible and comprehensive way. It must also correlate with other measures and attributes of carcinogenicity. The evidence presented demonstrates its biologic credibility, but illustrates the improbability of its being a comprehensive measure of all attributes of carcinogenesis. Agreement between the Ames Assay and animal cancer studies, other short-term bioassays, human carcinogens, and chemical structures predictive of carcinogens is presented. The range of possible predictive accuracy for the Ames Assay is calculated from these agreement evaluations. A hypothetical model is developed to test whether the Ames Assay is a sufficiently valid predictor of human carcinogenesis regardless of the extremes of this range. It is not sufficiently valid in the lower half of the range. So more work must be done to define precisely its accuracy.

The problems of applying the range of accuracy of a laboratory test of pure substances to biologic and environmental samples are explored. Unsolved problems in collection and analysis of airborne samples are identified. A study design is presented which minimizes the chance of bias in collecting airborne samples. The limitations of the laboratory procedure and the problems of sample collection are brought together in two continuous flow schemes. Together these demonstrate the unsolved biologic and biostatistical problems in validating the Ames Assay.

## ACKNOWLEDGEMENTS

A special thanks goes to my supervisor, Robin Roberts, for giving me a great deal of discretionary power in writing this thesis. That gave me a chance to stand or fall on my own analysis.

Dr. McCalla, who got me started working with the Ames Assay, and to whose data I am privileged to have access, also gave excellent support for this thesis. I hope my contributions to his projects will serve as some repayment.

Drs. Brian Haynes and Harry Shannon were very helpful in their perceptive and efficient reading and criticizing of my rough drafts. Many thanks, again, for the support they showed.

To Martha, Amos, and Emily, I hope that putting up with thesis neurosis was no worse than medical school neurosis, intern neurosis, or fellowship neurosis. Thank you for smiling through them one at a time.

TABLE OF CONTENTS

	PAGE
Abstract .....	i
Acknowledgements .....	ii
Table of Contents .....	iii
List of Tables .....	v
List of Figures .....	vi
CHAPTER I Introduction .....	1
CHAPTER II What is the Ames Assay? .....	7
CHAPTER III Criteria for a Valid Predictive Test of Occupational Cancer Risk .....	13
Biologic Credibility .....	15
Comprehensiveness .....	18
A. The Multi-stage Model .....	18
B. Synergism and Antagonism of Envir- mental Influences .....	21
C. Host Factors .....	23
CHAPTER IV Prognostic Accuracy and Substitution .....	35
A. Substitution .....	36
B. Chemical Properties of Animal Carcinogens .....	37
C. Other Short-term Bioassays .....	40
D. Populations with Expected Risk of Cancer .....	46
E. Sensitivity of Ames Assay to Known Human Carcinogens .....	55
F. Agreement Between the Ames Assay and Animal Studies .....	57

	PAGE
G. Summary .....	60
CHAPTER V Prospects for Practical Application and the Range of Possible Error Rates of the Ames Assay .....	62
CHAPTER VI What do Colony Counts Really Mean? .....	75
A. Reproducibility and Precision .....	75
B. Does Ames Assay Mutagenic Potency Predict Carcinogenic Potency .....	80
CHAPTER VII Issues in Environmental Sampling for Ames Assay Mutagens .....	87
A. Choice of Sampling Universe .....	88
B. Choice of Specific Sampling Material .....	97
C. Choice of Sampling Strategy .....	101
I. Dose-response Models .....	102
II. Models of the Distribution of Airborne Mutagens .....	104
III. Between Worksite Comparisons of Exposure .....	106
CHAPTER VIII Recommendations for Use of the Ames Assay as an Exposure Variable in Historical Prospective Occupational Cancer Studies .....	116
1. Defining Discrete Work Sites .....	117
2. Defining a Sampling Strategy .....	117
3. Laboratory Analysis .....	120
4. Analysis and Application of Data to Workers Records .....	125
CHAPTER IX Conclusions .....	132
REFERENCES .....	135

LIST OF TABLES

TABLE		PAGE
4-1	Agreement Between Ames Assay and Other Short Term Tests .....	44
4-2	Agreement Between Animal Studies and the Ames Assay .....	59
6-1	Problems with Designing a Between-Laboratory Agreement Study .....	79
9-1	Overall Progress in Validating the Ames Assay .....	133



LIST OF FIGURES

FIGURE		PAGE
1-1	Qualitative Performance of the Ames Assay .....	4
1-2	Quantitative Performance of Ames Assay .....	5
2-1	Flow-of-Events for Ames Assay .....	9
2-2	Characteristics of Salmonella Tester Strains in Widespread Current Use .....	10
2-3	A Full Test Procedure for a Given Substance .....	10
3-1	Flow-of-Events for Environmental Carcinogenesis .....	14
3-2	Effect of Host Defenses on the Ability to Detect Environmental Carcinogenic Stimuli .....	29
3-3	Possible Events in Malignant Transformation of a Human Cell and Expression of Cancer .....	31
3-4	Contingency Chart of the Plausible Components of Human Carcinogenesis, Ordered by the Possibility that the Ames Assay can Tap Them .....	32
4-1	Substitutions Available for Prognostic Accuracy .....	38
4-2	Mutagenic Urine with Exposure to Coke Plant Emissions .....	50
4-3	Analysis of Variance: Urine Mutagenicity as a Function of Smoking and Exposure to Tire-making Emissions .....	52
4-4	Variation in Steel Foundry Lung Cancer Rates .....	54
4-5	Sensitivity of Ames Assay to Human Chemical Carcinogens .....	56
4-6	Range of Agreement Between Animal Studies and the Ames Assay .....	61
5-1	Requirements of Sampling for Hui and Walter's Technique .....	64
5-2	Relationship of Prevalence of Carcinogenic Exposures to Ames Assay Results	
	(a) and (b) .....	66
	(c) and (d) .....	67

FIGURE		PAGE
5-3	Change in the Relative Probability of a Carcinogenic Exposure with Changing Underlying Prevalence of Carcinogenic Exposure .....	69
5-4	Use of Ames Assay Exposure Data to Identify "At Risk" Groups .....	74
6-1	Factors Affecting Colony Counts .....	76
6-2	The Relationship Between Colony Counts and Carcinogenic Potency .....	81
6-3	Contingencies in Validating Ames Assay Mutagenic Potency on Environmental Samples .....	83
6-4	Contingencies in Validating Ames Assay Mutagenic Potency on Environmental Samples .....	84
7-1	Possible Results of Comparing Air and Biological Samples Before and After Industrial Hygiene Intervention .....	91
7-2	Effects of True and False Confounders on the Use of Air and Biological Samples and the Ames Assay .....	96
7-3	The Alveolar Dust Curve .....	98
7-4	Contingency Chart of Possible Airborne Mutagen Distributions with Time and Mutagen Dose/Cancer Response Relationships .....	103
7-5	Model of Factors Affecting Cumulative Risk	
	(a) .....	107
	(b) .....	108
7-6	Risk Assessment when Dose-Response Curve is Linear and Passes Through Origin .....	109
7-7	Risk Assessment with a Threshold Dose-Response Relationship When the Underlying Mutagen Distributions at Different Work Sites Have Unequal Variance .....	111
7-8	Risk Assessment and a Threshold (or Plateau) Dose-Response Relationship .....	112
7-9	Risk Assessment with Between Work-site Differences in the Mutagenic Dose/Carcinogenic Response Relationship .....	113

FIGURE		PAGE
7-10	Flow Scheme Relating the Limitations of Airborne Samples to the Comprehensive Validity of the In Vitro Procedure .....	115
8-1	General Form of the Analysis .....	126
8-2	Translating Milligrams to Standardized Volumes of Air Exposure .....	126
8-3	Flow of Events for Ames Assay Exposure Variable and Occupational Cohort Study .....	130

## CHAPTER I

### INTRODUCTION

The Ames Assay is the most popular of a series of short-term in vitro tests designed to evaluate indirectly the carcinogenic potential of pure chemicals and environmental samples<sup>1</sup>. These tests have been developed at a time when there is a perceived need throughout the industrial nations for primary preventive measures against environmental carcinogens.

Primary prevention involves identifying a risk of morbidity or mortality before its physical onset and taking steps to eliminate the risk. By definition, cohort studies of exposed groups or case-control studies of those with a target outcome cannot be exercises in primary prevention since they depend upon the fact of morbidity or mortality for a positive result.

In the past the primary method of evaluation of substances for carcinogenic potential has been the animal experiment. Ideally, the animal experiment is the perfect randomized, blind, controlled trial with few ethics or compliance problems to interfere with the purity of the experiment. It covers the entire life span of the animal, allows various routes of entry to be tested (though not at once), and may give insights into the non-carcinogenic toxicity of the test substance. Significant outcomes may include the total number of tumours identified in the test population, loss of life span due to cancers, or the total number of test animals with tumours.

Criteria for a credible clinical trial should be met in an animal experiment, including random allocation and prognostic stratification of animals, precisely defined manoeuvre and outcome criteria, prevention of contamination and co-intervention, and guarantees of total outcome reporting? Particularly, the experiment should be carried out in a "specific pathogen free" environment to avoid co-morbidity due to infection. Depending on the study species, such experiments may take one to three years to complete. They are expensive (several hundred thousand dollars per experiment) and few centres have the facilities to carry them out, even without the constraint of a specific pathogen free environment. Although consistent results across species and with different routes of administration may strengthen the conclusions drawn, animal experiments have not been readily accepted as a criterion standard for carcinogenic potential in humans. Generalizability is the main issue:

(1) The dose levels necessary to induce cancer in an animal with a 2 to 3 year life span are greater than for a species with a 70 year life span. Moreover, the dose levels must be high enough to ensure adequate statistical power among a relatively small animal cohort. This gives rise to uncertainty regarding applicability of the effects of high-dose, short-term exposure in test species to low-dose, long-term exposure in humans.

(2) Inhalation, the primary route of entry for environmental carcinogens, is often difficult to simulate with animals. Do differences in site-specific bioavailability of test substances affect the validity of results?

(3) Metabolic differences between humans and animals may cause very different end-products to be presented to target tissues.

(4) Test animals are usually exposed to pure compounds whereas humans are exposed to environmental mixtures. If a mixture contains an animal carcinogen, how do the other components affect carcinogenic potential?

By contrast, the Ames Assay (described in detail in the next chapter) is cheap, quick, relatively easy to perform, and adaptable to a wide variety of test situations. This thesis will identify criteria necessary for its validation and examine the strength of the evidence currently available, against these. It will focus on the problems involved with validation and present a general design of a data collection system for Ames Assay studies of the occupational environment.

The central research question is, "Is the Ames Assay a valid predictive test for increased risk of environmentally-induced cancer among exposed populations?" The concern expressed is with collective, not individual risk, contrary to clinical tests. The question can be broken down into qualitative and quantitative components. The qualitative question concerns the performance of the Assay as a dichotomous test for carcinogenic potential. It can be answered by determining the values of the four major parameters (sensitivity, specificity, positive and negative predictive value) in the hypothetical 4-fold table in figure 1-1. The quantitative question concerns the performance of the Assay as a measure of relative potency of a carcinogen. It can be answered by determining the nature of the curve on the hypothetical, two-dimensional graph in figure 1-2.

Figure 1-1: Qualitative Performance of the Ames Assay

Human Cancer Risk

	Increased	Not Increased	
Cohorts Exposed to Ames Assay + environment	A	B	A + B
Cohorts Exposed to Ames Assay - environment	C	D	C + D
	A + C	B + D	N

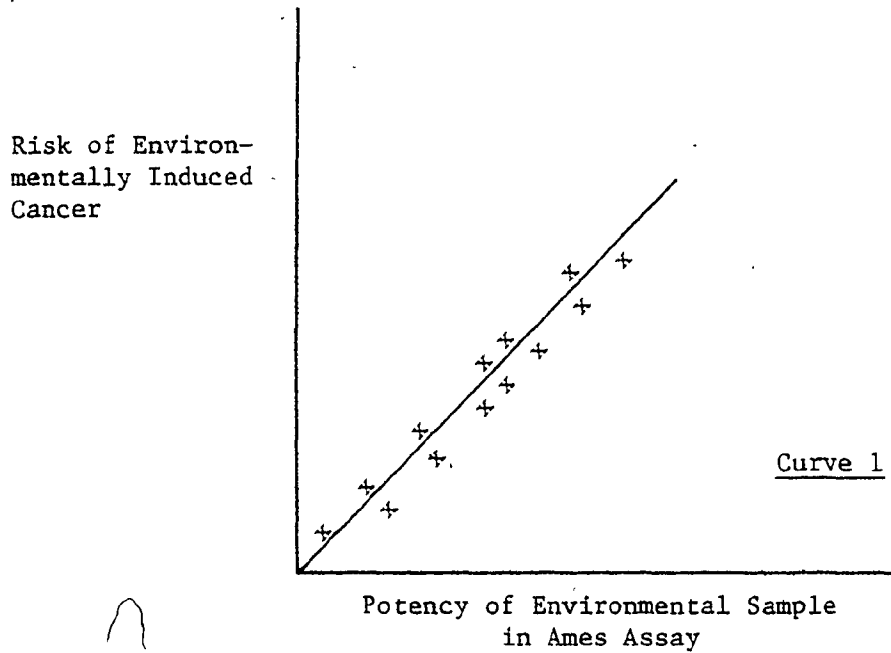
Sensitivity =  $\frac{A}{A+C}$ ; the probability that an environment which truly carries with it an increased cancer risk will be Ames Assay positive.

Specificity =  $\frac{D}{B+D}$ ; the probability that an environment which truly carries with it no increased cancer risk will be Ames Assay negative.

Positive Predictive Value =  $\frac{A}{A+B}$  (if  $\frac{A+C}{B+D}$  is approximately equal to the ratio of risk to no risk environments in the universe of interest); the probability that an Ames Assay positive environment poses human cancer risk.

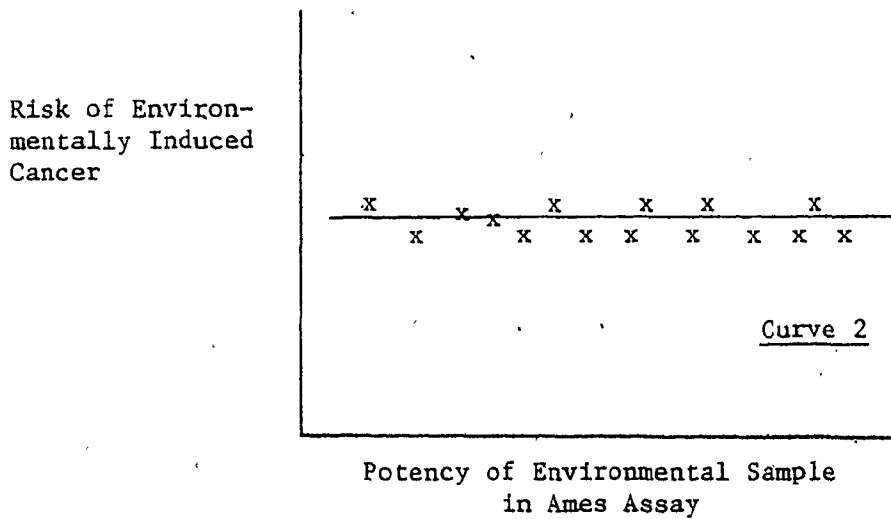
Negative Predictive Value =  $\frac{D}{C+D}$  (same constraint as above); the probability that an Ames Assay negative environment poses no human cancer risk.

Figure 1-2: Quantitative Performance of Ames Assay



Curve 1 - Monotonic increasing risk with increasing exposure-validity of Ames Assay would be maximum.

Curve 2 - No relationship between risk and exposure. Ames Assay would be invalid as a quantitative measure of cancer risk.





Because of the paucity of direct evidence validating or invalidating the Ames Assay, this thesis will have a theoretical orientation. It will focus on the problems in assessing the validity of the Ames Assay in relation to airborne contaminants in the industrial environment.

## CHAPTER II

### WHAT IS THE AMES ASSAY?

The Ames Assay measures the ability of a test substance to induce mutations in DNA.<sup>3</sup> Specifically, it measures the ability of a substance to cause a mutation at a particular location on the chromosome of special tester strains of *Salmonella typhimurium* bacteria. In order for this to come about, the tester strains were originally changed from the "wild type" in two ways:

(1) The "excision repair system", responsible for identifying and repairing damage to the DNA, was deleted, so that mutations would be irreversible.<sup>3</sup>

(2) Each (of several) tester strains had a different mutation induced in the "histidine operon" (that stretch of DNA responsible for directing the synthesis of the amino acid histidine) making it incapable of synthesizing histidine. This resulted in dependency on external supplies of histidine for normal growth and cell division. But each strain was capable of synthesizing its own histidine if a substance were to enter the bacterium and cause a "back mutation" (back to the wild type) at the proper site on the histidine operon.<sup>3</sup>

How is the test performed? The test bacteria are spread on an agar plate which is deficient in histidine. The test substance is added to the plate and it is incubated for 48-72 hours, much like a "culture and sensitivity" test. If the test substance is a mutagen, it may penetrate the bacterial cell membrane and cause the necessary mutation (of

the histidine operon) among a proportion of the bacteria. These will continue to divide normally and produce macroscopic colonies. If the substance is not a mutagen, only spontaneous revertants will grow. So the Ames Assay is based on the difference in the number of macroscopic colonies present on the test plate, compared to the control plates (figure 2-1)<sup>1</sup>

In order to increase the sensitivity of the original strains, two more modifications of the bacteria were introduced later.

(1) Mutations were introduced that eliminated, to differing extents, the lipopolysaccharide coating of the bacterial surface, making the bacteria more permeable to potential carcinogens and completely nonpathogenic<sup>4</sup>

(2) An intracellular inclusion of genetic material (known as the R-factor plasmid) was added, which enhanced the sensitivity of the tester strains for unknown reasons<sup>5</sup>

It was known that many animal carcinogens required in vivo metabolic activation from pre-carcinogen to ultimate carcinogen. The best studied and most important activation system was the NADPH - dependent microsomal fraction in the liver. So the test procedure was further modified to include a duplicate set of test plates wherein rat liver homogenate (known as the "S<sub>9</sub> fraction") was added to the top agar (figure 2-1). This modification, combined with the addition of the R-factor plasmid, markedly increased the chemical range of sensitivity of the test bacteria<sup>6</sup>

Figure 2-1: Flow-of-Events for Ames Assay

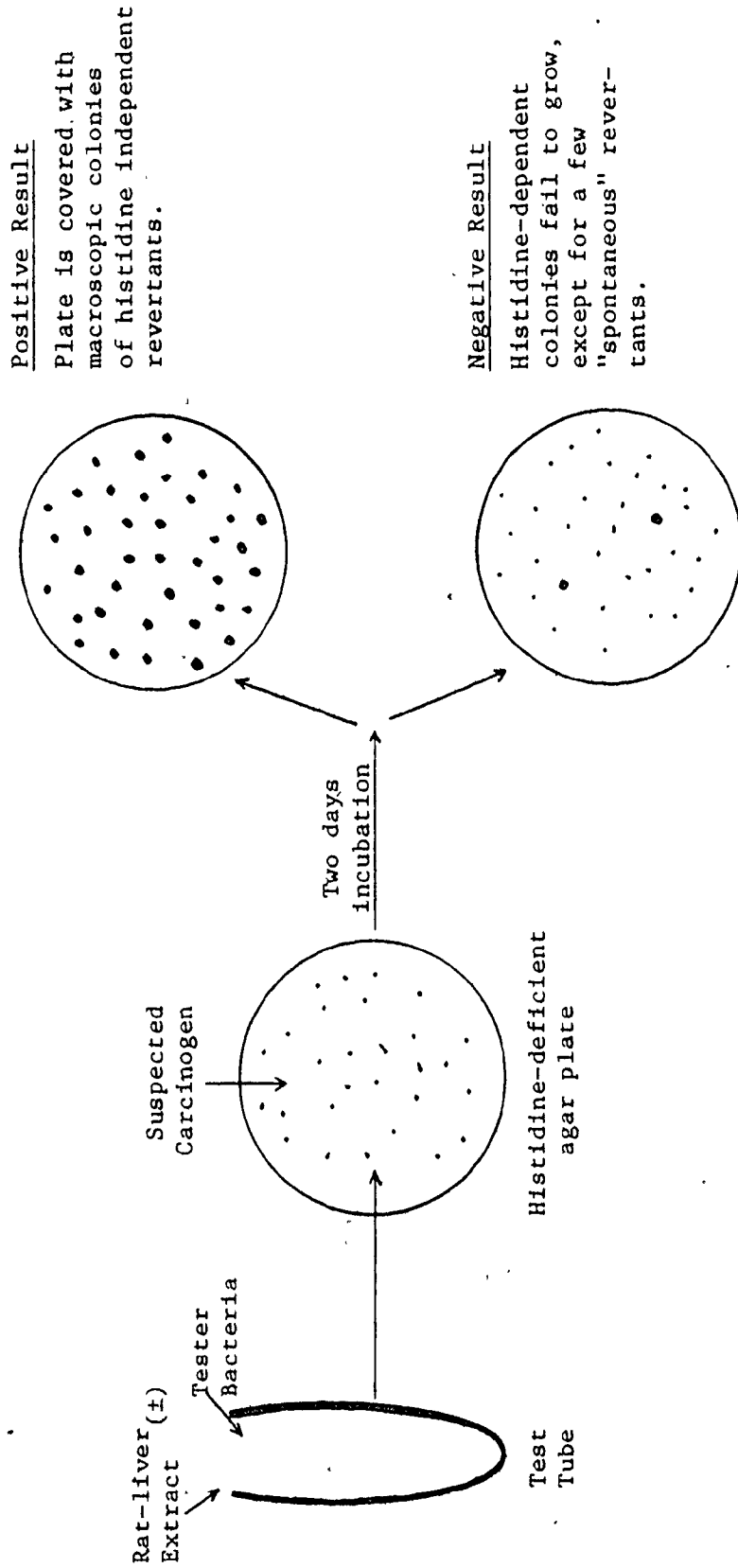


Figure 2-2: Characteristics of Salmonella Tester Strains in Widespread Current Use

Mutation Sensitivity			
	Base-pair Substitution	Frameshift	
<u>Plasmid</u>	<u>Present</u>	TA 100	TA 98
	<u>Absent</u>	TA 1535	TA 1536, 1537, 1538

Figure 2-3: A Full Test Procedure for a Given Substance

<u>Trial</u>	<u>Active Ingredients</u>			<u>Results</u>		<u>Decision</u>
1.	Substance X	+	TA 100	+	S <sub>9</sub>	
2.	Substance X	+	TA 100	(no S <sub>9</sub> )		+ / -
3.	Substance X	+	TA 98	+	S <sub>9</sub>	+ / -
4.	Substance X	+	TA 98	(no S <sub>9</sub> )		+ / -

The different tester strains were sensitive to one of two different types of mutagenic event, depending upon the nature of the original mutation induced in the histidine operon.

(1) Base-pair substitutions: The mutagen causes replacement of the original base-pair at a certain point on the double-stranded DNA by another pair.<sup>5</sup>

(2) Frameshifts: The mutagen causes base-pairs to be added to or subtracted from the DNA strands<sup>5</sup>

In the original tester strains, substances tended to test positive as one sort of mutagen while not as the other. The addition of the plasmid increased the number of compounds positive in both frameshift and base-pair substitution sensitive strains. Figure 2-2 gives the characteristics of Salmonella tester strains in widespread current use.<sup>5</sup> A full test procedure generally involves both R-factor strains, with and without the Sq fraction. That is, four sets of plates must be set up for each substance tested (see figure 2-3).

Because of its simplicity, cheapness, and adaptability to a wide variety of samples, a large number of potential uses have been suggested for the Ames Assay<sup>7</sup>

(1) Mass pre-screening of commercial chemicals before they reach the market.

(2) Identifying mutagenic impurities in otherwise non-mutagenic products.

(3) Surveying air, water, and soil samples for mutagenic activity, and in conjunction with analytic chemistry techniques, identifying

the mutagenic components.

(4) Testing of urine, feces, and other biologic samples in relation to dietary, environmental, or medical exposure to suspect substances.

(5) Screening substances to decide which should be tested in animal bioassays.

These applications depend critically on the predictive validity of the Ames Assay for human cancer risk, since they may lead to a series of public health interventions to control or eliminate exposures to mutagens. It is important to note at the outset that the Ames Assay is only applicable to non-metallic substances whose potential for carcinogenic potency derives from chemical properties. Substances such as asbestos, whose carcinogenicity derives from its physical characteristics, cannot be detected by the Ames Assay<sup>7</sup>.

## CHAPTER III

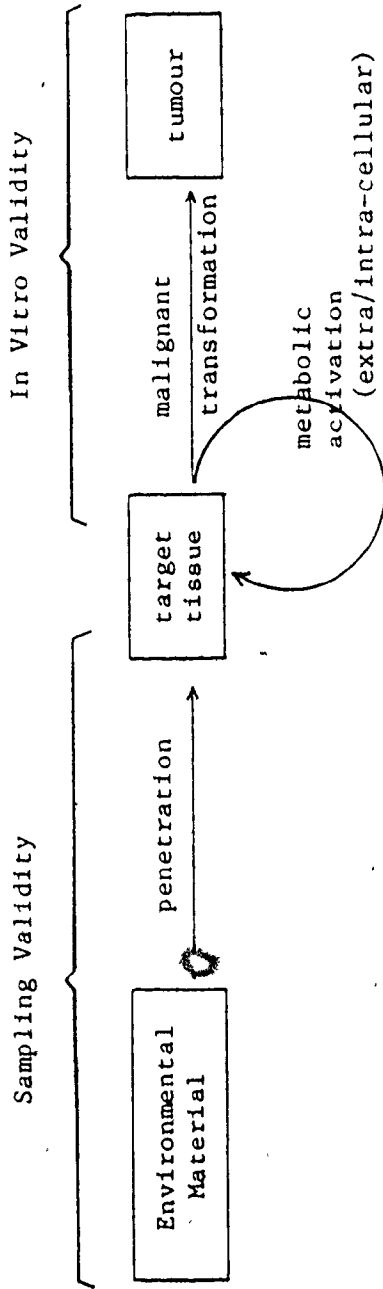
### CRITERIA FOR A VALID PREDICTIVE TEST OF OCCUPATIONAL CANCER RISK

The basic flow-of-events for environmental carcinogenesis is given in figure 3-1. The diagram shows that in order for environmental substances to cause cancer they must first contact a target tissue and then initiate, stimulate, or participate in a "carcinogenic event". The laboratory procedure known as the Ames Assay attempts to simulate the carcinogenic event. The former aspect depends on the nature of the samples presented to the test bacteria. So the overall validity of the Ames Assay as a measure of environmentally-induced cancer risk depends on the validity of the sampling procedure (sampling validity) and the validity of the laboratory procedure (in vitro validity). This chapter will explore the validity of the laboratory procedure. Discussion of the sampling strategy will follow later in the text.

The validity of a predictive test concerning a physical attribute may be divided into four categories:<sup>8</sup> biologic credibility, comprehensiveness, prognostic accuracy, and (failing the latter) substitution by attribute, manifestation, and measurement. Biologic credibility concerns the appearance of validity. "Does the Ames Assay measure something that should be of predictive importance in carcinogenesis?" Comprehensiveness is evaluated by answering the question: "Does the Ames Assay measure all aspects of carcinogenesis while excluding from influence all non-carcinogenic effects?" Prognostic accuracy is the criterion



Figure 3-1: Flow-of-Events for Environmental Carcinogenesis



Test Validity = Sampling Validity + In Vitro Validity

standard for validity of a predictive test. "Do Ames Assay results collected prospectively correctly predict those groups of people with increased risk of cancer mortality or morbidity?" Finally, if prognostic accuracy is unmeasurable, "Does the Ames Assay correlate with a theoretical attribute or manifestation of carcinogenesis or with other measures of carcinogenic risk?" This is known as validation by substitution. These criteria will be examined in order.

### Biologic Credibility

Since the Ames Assay attempts to directly measure the mutagenic potency of substances, the central credibility question is, "Is somatic mutation a necessary step in carcinogenesis?" The theory that somatic mutation is the fundamental initiating step in carcinogenesis is the working hypothesis of much contemporary cancer research. Four separate lines of evidence support this.

(1) In animal systems and in vitro, it has been demonstrated that exposure to carcinogens confers on cells a transmissible potential for neoplastic transformation. Since the chromosomes contain the transmissible information, this implies a genetic effect.<sup>9</sup>

(2) Certain rare congenital diseases involving known genetic defects carry with them a striking increase in the risk of developing cancer.<sup>10</sup> One example is xeroderma pigmentosum, an inherited defect in the enzymes that repair DNA damaged by ultraviolet light, which leads to multiple skin cancers. Others include Bloom's syndrome and Fanconi's anemia, retinoblastoma, polyposis coli, and ataxia telangiectasia.

(3) Most known animal carcinogens either react directly with DNA or indirectly, after metabolic activation.<sup>11</sup>

(4) Most animal carcinogens have been demonstrated to have mutagenic potency in systems other than the Ames Assay.<sup>12</sup>

However, several experimental results challenge the assumption somatic mutation is always fundamental to carcinogenesis.

(1) In the mouse, the nucleus of a cancerous cell has been injected into a normal blastula, producing a normal, genetically mosaic mouse. In the frog, nuclei from kidney carcinomas have been transplanted into enucleated eggs and have produced normal tadpoles. So under some circumstances the obvious genetic changes within a malignant nucleus may not be expressed in cellular behaviour.<sup>13</sup>

(2) In one series of experiments, tumours were induced in rats simply by inserting a solid sheet of a chemically inert material into tissue. Porous or fibrous material did not produce tumours. Such results cannot be explained by a genetic effect of the material and imply a physical mechanism.<sup>13</sup>

(3) Other experiments have demonstrated that transplantation of some endocrine tissues into the spleen within the original host causes them to become neoplastic. Although mutagenic events may accompany such disruptions, a more plausible hypothesis would involve loss of co-ordination of body tissues.<sup>13</sup>

Taken together, the supporting and challenging lines of experimental evidence do not provide conclusive evidence that mutagenesis is fundamental to carcinogenesis. They suggest that, in animals, mutagenesis is not always fundamental. Since the experimental evidence has not

been replicated in humans, it is difficult to assess the relative importance of the supporting and non-supporting evidence. In terms of Dr. Sackett's nine diagnostic tests of causation,<sup>2</sup> evidence of "analogy" is given in the xeroderma pigmentosum example. Taken together, the other positive and negative evidence suggests that mutagenesis is associated with carcinogenesis in a biologically sensible but non-specific way. The non-specificity concerns both the pathology of cancer (there may be more than one pathologic process leading to malignancy) and the action of mutagens (their mutagenic activity may not be the same as their carcinogenic activity). Because direct evidence of the biologic role of mutagenesis must be found by laboratory experiment, the major diagnostic tests of causation cannot be addressed. The strongest possible evidence would be the demonstration of a step-by-step series of intracellular reactions leading from mutagenesis to tumour formation. Such evidence is not available.

Here, the concerns of basic science and health surveillance diverge. It is the purpose of basic science to generate a conception of carcinogenesis that simultaneously explains all genuine experimental results. Yet it is possible for the Ames Assay to measure something that is not always fundamental to carcinogenesis and still be a valid test. Persons with high blood pressure do not all have strokes and not all persons with strokes have high blood pressure. Nonetheless, blood pressure measurement is a valid predictor of the risk of stroke because those with high blood pressure are more at risk of stroke than those without.

In conclusion, the evidence available does not sufficiently challenge the biologic credibility of the Ames Assay to negate its ultimate validity.

### Comprehensiveness

Comprehensiveness is a two-part criterion. The first part asks, "Does the Ames Assay measure all aspects of carcinogenesis?" The basis of the test is mutation of the DNA after extracellular and intracellular metabolic activation. So as the "true" process of carcinogenesis diverges from this model, the comprehensive validity of the Ames Assay will diminish. Evaluation requires knowledge of the influences brought to bear on a human cell which becomes malignant. Comprehensiveness also concerns exclusivity: Is the Ames Assay sensitive to factors not important in carcinogenesis? This section will identify processes thought to be important in human carcinogenesis. A contingency table, summarizing these processes, will be generated to illustrate the range of possible mechanisms and their implication for the validity of the Ames Assay.

#### A. The Multi-stage Model

Armitage and Doll developed a mathematical model relating age to cancer death rates based on the concept of carcinogenesis as a multi-stage process.<sup>14</sup> The model made the following assumptions:

1. Carcinogenesis consists of a series of improbable events, separated by months and years, each defining a stage of malignant transformation.

2. The various events take place in a specific order.

3. The rate of occurrence of the  $i^{\text{th}}$  event is the sum of a constant background rate plus the rate brought about by environmental influences.

Therefore,  $D(t) \propto P_1 P_2 \dots P_k (t)^{k-1}$ ,

where  $D(t)$  = age-specific death rate for a particular cancer

$P_i$  = probability of occurrence of a specific carcinogenic event over a lifetime

$k$  = number of events necessary for tumour formation

$t$  = time (age) of death from cancer,

So  $\text{Log } D(t) = (k-1) \text{Log}(t) + \text{constant}$

If the model is correct, then log-log plots of site-specific cancer death rates versus age at death should give straight line relationships. The investigators checked this prediction using the age-specific death rates for 17 common cancers in England and Wales, 1950-51.<sup>14</sup>

Nine of seventeen cancers (esophagus in males; pancreas, colon, stomach, and rectum in both sexes) gave straight line relationships with slopes of 4.97 to 6.48, implying a six ( $4.97 + 1 = 5.97$ ) or seven ( $6.48 + 1 = 7.48$ ) stage process, depending on site. The confidence intervals around the slopes did not always include a whole number. Moreover, the data showed between-sex disagreement in two of four cancers as to whether there was a six or seven stage process.

Eight of seventeen cancers (lung, both sexes; bladder and prostate, males; breast, ovary, uterus, and cervix, females) deviated from

linearity. This observation could be taken as evidence against the multi-stage model. However, there is an alternative explanation. In the cases of lung and bladder cancer the prevalence of risk factors such as smoking and high exposure occupations have changed over the years. Since the mortality data used by Doll and Armitage were cross-sectional, the historic period of exposure varied with age at death. If the prevalence of risk factors has increased with time, then the age-specific death rates should deviate negatively from the linear at higher ages. This is because the  $P_i(t)$ 's would vary with time, causing the "constant" term in the equation to vary. This is observed with the curves for lung cancer among males and females, consistent with the historic increase in the prevalence of risk factors.

Breast, ovary, cervix, and uterus cancer also showed negative deviations from the linear in older age groups. These are all cancers of reproductive organs, with age-dependent endocrine influences on the host organ. Since the endocrine influences decrease after menopause, the age-dependency should operate in a manner analogous to increasing prevalence of environmental risk factors with time; causing the  $P_i(t)$ 's to decrease with increasing age.

The model is consistent with five important observations:

- (1) The rapid increase in mortality with age observed in cancer of some sites.
- (2) The irregularity in the increase in cancer of some other sites.

(3) The long latent period observed after exposure to a carcinogen before a tumour develops.

(4) The experimental finding that cancer incidence tends to be proportional to the concentration of the applied carcinogen.

(5) The existence of pathologically observable pre-malignant transitional states for some tumours!<sup>4</sup>

The assumption that the events must take place in a specific order carries with it the implication that early-stage events will have more impact on outcome if they occur early in life than if they occur later. The reverse would be true for late stage events. The model was adapted to include a weighting scheme which predicted the effect of changes in the period probability of early and late-stage events on age-specific mortality rates!<sup>5</sup> This adaption made the model consistent with patterns of cancer mortality among those exposed to more than one known carcinogen, such as cigarette smoke and asbestos. It was further extended to fit the concept of initiation and promotion and to correct for the time between tumour onset and death!<sup>6,17</sup>

#### B. Synergism and Antagonism of Environmental Influences

In 1941, Berenblum carried out a series of experiments wherein animals were exposed to carcinogenic benzo ( $\alpha$ ) pyrene or non-carcinogenic croton oil or both, at various doses!<sup>8</sup> By comparing tumour yields among the doubly exposed group to each of the singly exposed groups, three important effects were identified.

(1) Adding croton oil to the benzo ( $\alpha$ ) pyrene increased the



tumour yield for a given benzo ( $\alpha$ ) pyrene dose.

(2) The dose of benzo ( $\alpha$ ) pyrene needed to reach a statistically significant increase in tumour incidence was decreased with the addition of croton oil.

(3) The latent period between first exposure to benzo ( $\alpha$ ) pyrene and tumour incidence was decreased with the addition of croton oil.

The name given to a non-carcinogen which could enhance the effects of a carcinogen was "promoter". The carcinogen itself was called an "initiator". A family of substances called the "phorbol esters" were isolated from croton oil.<sup>19</sup> Their promoting properties have been replicated and extensively analyzed in animal and in vitro systems. Since then, other promoters have been identified including long-chain paraffins, fatty acids, esters, irritants, and amino acids.<sup>20</sup> One substance, 1 - dodecanol, was experimentally demonstrated to decrease the observable "threshold dose" of benzo ( $\alpha$ ) pyrene one-thousand fold.<sup>20</sup> Several mechanisms have been suggested for these effects.

(1) Promoters may enhance absorption of carcinogens across cell membranes.

(2) They may induce enzymes which cause metabolic activation of carcinogens.<sup>20</sup>

(3) They may mimic a growth stimulus, possibly by decreasing contact inhibition between cells or inducing hyperplasia.<sup>9</sup> If the experimental conclusion that "the transformation frequency depends on the number of times after exposure to a chemical carcinogen that cells within a colony can be kept dividing before they contact other cells within

the same colony..." can be applied in vitro, this mechanism would be consistent with the observation that some promoting agents are effective when administered after the initiator has been fully metabolized.

A wide range of experimental inhibitors of carcinogenesis have been identified. Wattenburg et.al.<sup>21</sup> suggested possible mechanisms of inhibition, including:

- (1) Decreased metabolic activation of the carcinogen.
- (2) Increased metabolic detoxification of the carcinogen.
- (3) Scavenging of active molecular species of carcinogens to prevent their reaching critical target sites in the cell.

#### C. Host Factors

Three host factors have been considered as important modulators of the carcinogenic response in animals and humans: immune surveillance, hormone effects, and metabolic variability. How might each of these affect the carcinogenic process?

##### I. Immune Surveillance

Proponents of immune surveillance claim that the immune system has an important role to play in tumour identification and elimination at the pre-malignant stage.<sup>22</sup> Three lines of evidence are offered to support this contention:

- (1) Immuno-suppressed patients and renal transplant patients (who are therapeutically immuno-suppressed), tend to have increased rates of cancer. However, the increased cancer rates are confined to the immune tissues. This suggests that the carcinogenic effect is due to in-

jury to the immune system and not a breakdown in immune surveillance.<sup>22</sup>

(2) Some chemical carcinogens are immuno suppressive in animals. However, the immuno suppressive response usually takes place at higher doses than the carcinogenic response. By titrating the dose, the carcinogenic and immuno suppressive responses have been separated experimentally in animals.<sup>22</sup>

(3) The immune surveillance concept is based on the model that carcinogenic events are relatively frequently-occurring, producing many potentially malignant cells which produce antigenic markers. Those are detected by the immune surveillance mechanism, which causes destruction of the premalignant cell. Biologic senescence of the immune system will allow malignant cells to be missed and tumours to develop more frequently with advancing age. However, not all tumours have antigenic markers and relatively few tumours have strong ones. In animal experiments the immune response to premalignant lesions is very inefficient. The notion that carcinogenic events are frequently occurring conflicts directly with the assumptions of the multi-stage model. It is uncertain whether a rational model of immune surveillance failures with increasing age could be developed to fit the age distribution of site-specific cancer incidence.<sup>22</sup>

## II. Hormones

Interest in a hormonal role in carcinogenesis began at the turn of the century when it was observed that removal of the ovaries often resulted in regression of breast cancer in women.<sup>23</sup> Research has centred on the influence of hormones on endocrine organs or their target tissues.

There is evidence that non-physiologic hormones and hormone-like substances may be carcinogenic. The best example is diethylstilbestrol, while the possible carcinogenicity of post-menopausal estrogens is still in question. The evidence relating physiologic hormones to human cancer is confined to the discovery of estrogen receptors on some breast tumours and the correlations found between fertility patterns and breast cancer in women. Outside of breast, ovary, and uterine cancers, no correlations with hormones have been found.<sup>24</sup>

### III. Metabolic Variability

Metabolic variability is a catch-all category which brings in all the between-host differences in the activation of carcinogens, and host repair of genetic damage by carcinogens. Several aspects of the area have been addressed by various investigators. These include:

(1) whether or not individual variability in the activation of pre-carcinogens to carcinogens follows a unimodal or a polymodal distribution. Conflict centres around aryl hydro carbon hydroxylase, one of two important enzymes converting benzo (a) pyrene to its carcinogenic form in the human lung. One investigator produced data showing a trimodal distribution of AHH activity,<sup>25</sup> while another showed a broad unimodal range.<sup>26</sup> This conflict is unsettled and is equally important for other enzyme systems not as well explored as the AHH system.

(2) the patterns of variability in DNA repair systems. These may also follow a polymodal or unimodal distribution. They are important since DNA repair is considered the primary host defense against mutagenic activity.<sup>27</sup>

(3) the role of other host factors such as nutritional variability. The systemic effects of nutritional variability are virtually unknown.

How might the various possible influences on carcinogenesis affect the validity of the Ames Assay? To answer this question, the environmental and host influences will be discussed separately.

According to the research mentioned above, environmental influences on carcinogenesis may include initiation, promotion/ inhibition, and "multiple events". In this discussion, initiation means a single event or the first of a "multiple event" carcinogenic process. Multiple events are taken to mean necessary events, while promotion is taken to mean a non-essential influence. In the ideal assay system, initiating and multi stage stimuli would directly reflect themselves in the result. Promoting and inhibiting influences would only increase and decrease results in the presence of necessary carcinogenic stimuli. They would not demonstrate mutagenic potency on their own. However, experimental evidence suggests that promoters can increase the effect of a previous carcinogenic stimulus. So if promoters never test positive on their own, their true impact on risk will not be reflected in vitro in certain cases. If this model of promotion is correct, the Ames Assay would not be able to completely reflect the biologic process. On the other hand, if another model of promotion were true, it might be consistent with the performance of the Ames Assay. Nevertheless, the biologic mechanism of promotion would have to be simulatable in the Ames Assay to be expressed at all. A mechanism such as increasing carcinogen absorption or induc-

ing enzymes which convert pre-carcinogens to carcinogens may be simulatable in the Ames Assay. Mechanisms such as influencing between-cell contact inhibition will not be simulatable.

If any necessary environmentally-influenced events were non-mutagenic, they would show up as false negatives in the Ames Assay and reduce its validity as a qualitative measure of risk. The more non-mutagenic events, the greater the erosion of validity. The extreme example would be a thoroughly non-mutagenic process, wherein the Ames Assay would be useless in identifying risk.

If carcinogenesis were a single-stage mutagenic process, Ames Assay revertant counts could be a quantitative measure of risk, subject to the limitations of between species variability. If it were a multi-stage mutagenic process, the Ames Assay would be unable to separate the various mutagenic stages represented in the samples presented to it. This would add a second level of uncertainty into any quantitative measure of risk.

If the multi-stage model is correct and a specific sequence of environmentally-induced events is required for carcinogenesis, the age of the exposed person would interact with the stage of the event to strongly influence its impact on risk. Late stage events would have optimum influence among "older" people and early stage events would have optimum influence among younger people. A strongly mutagenic sample related only to early stage events may not be associated with any observable increase in risk in an aging population. But a weakly mutagenic

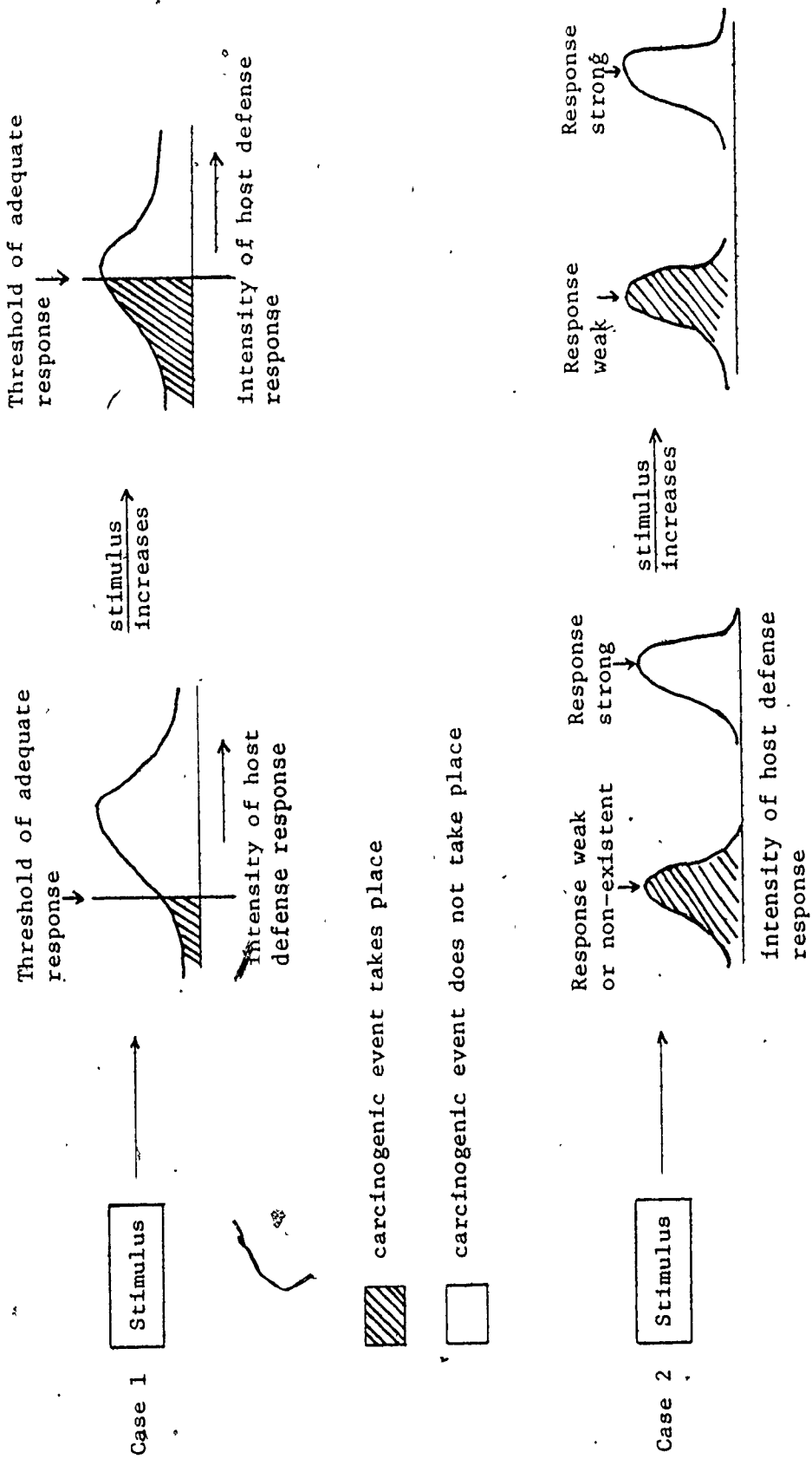
sample related to a late stage event in the same population may be associated with observable risk.

Under what conditions could promoting or inhibiting influence be simulated by the Ames Assay? Three hypothetical pathways were outlined above: changing membrane transport, altering metabolic activation, and releasing cells from contact inhibition. Because membrane transport follows the same principles regardless of species, this is the most simulatable of the three proposed mechanisms. It is possible that some metabolism-altering activities could be simulated. But between-species variability makes it less likely. Since contact inhibition is a function of multi-cellular organisms, it is not simulatable in the Ames Assay system.

Do the host factors previously mentioned affect the validity of the Ames Assay? It is axiomatic that the Ames Assay cannot simulate long-term responses in the human host. The important issue is whether or not human variability determines which environmental influences are sufficient for carcinogenesis. Two models of variability cover the plausible range: a unimodal distribution of a host response (on a plot of intensity versus population) or discrete responses in various subpopulations (figure 3-2).

In the former case, the sufficiency of a stimulus will vary quantitatively but not qualitatively between subjects. At any given level of exposure, a certain proportion of the exposed population will experience the carcinogenic event. This proportion will increase with in-

Figure 3-2: Effect of host defenses on the ability to detect environmental carcinogenic stimuli





creasing exposure and decrease with decreasing exposure. So the predictive value of the Ames Assay will depend upon the representativeness of any exposed "study" population in terms of the host factors involved, as well as whether the impact of the mutagenic event(s) are great enough to be detectable in the study population.

If the discrete subgroups model were correct, a stimulus would only be sufficient conditional on the subpopulation of the individual host. Despite wide fluctuations in exposure level, one or more subgroups may remain refractory to the carcinogenic stimulus while other subgroups remain susceptible. Once again, the predictive value of the Ames Assay will depend upon the representativeness of the exposed population. If the grouping factor were a predictor of exposure (eg. sex, race), representativeness would be precluded. If the susceptible subgroups were small, the mutagenic impact may not be great enough to be measureable in a representative population.

Figure 3-3 sets out the six influences of carcinogenesis which have been discussed. The alternate mechanisms plausible for each influence have been resolved into extremes. These have been arranged in columns in figure 3-4 in ascending order of their theoretical contribution to the predictive validity of the Ames Assay. The columns are arranged in rough biologic sequence (although metabolic variation may be important before or after initiation). So the collapsed contingency table formed allows the identification of several hundred possible combinations of mechanisms in carcinogenesis.

Figure 3-3: Possible Events in Malignant Transformation of a Human Cell and Expression of Cancer

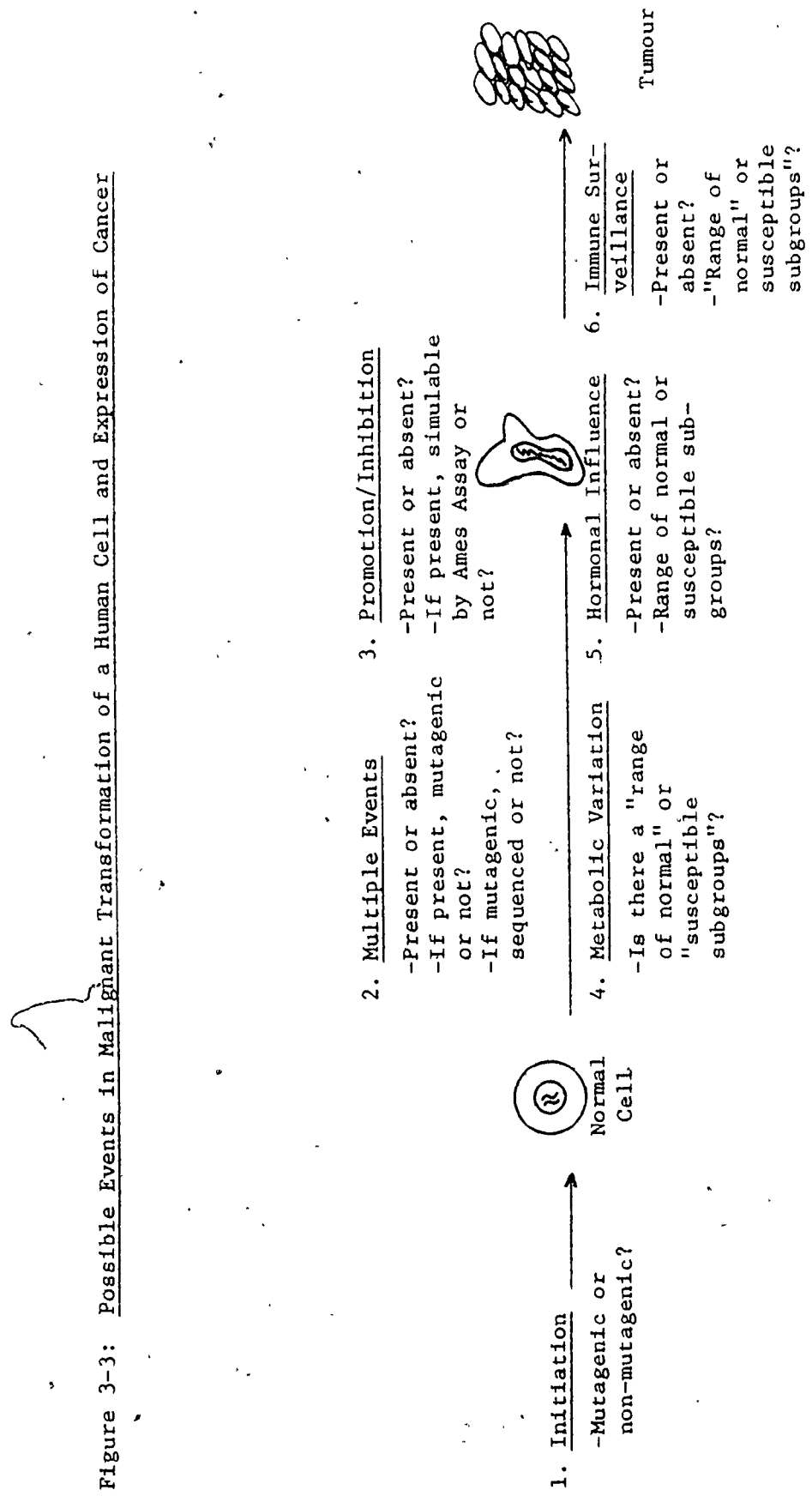
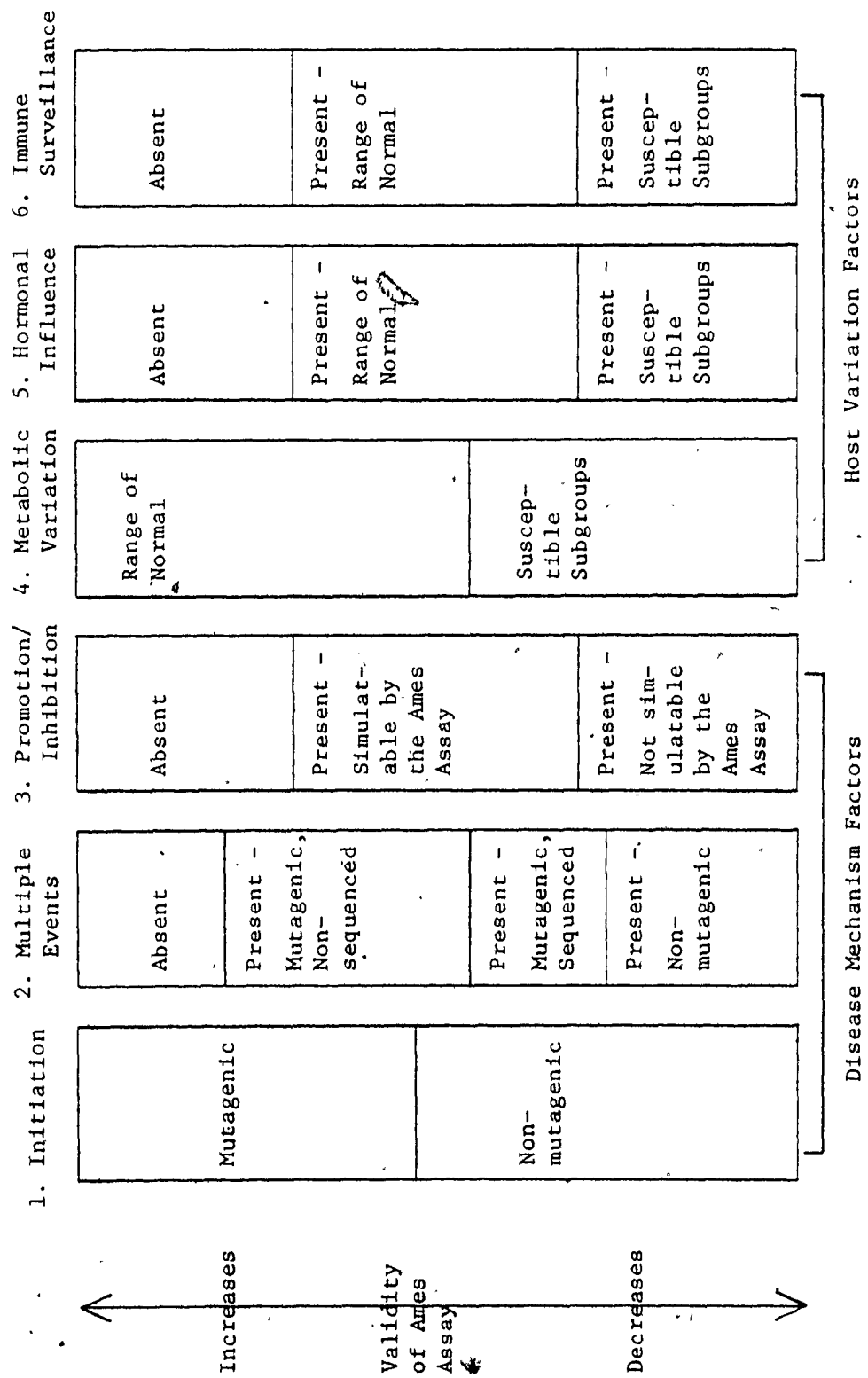


Figure 3-4: Contingency Chart of the Plausible Components of Human Carcinogenesis, Ordered by the Possibility that the Ames Assay can Tap Them



How should the table be interpreted? By making selection from each column, a model of carcinogenesis can be built. It can be compared against the performance characteristics of the Ames Assay to answer the question: If this model were true, for a specific cancer, would the Ames Assay tap all its particular biologic attributes? The range of possible selections begins at:

Initiation : Mutagenic  
Multi-events : Absent  
Promotion/Inhibition : Absent  
Metabolic Variation : Unimodal  
Hormonal Influences : Absent  
Immune Surveillance : Absent

in which case the Ames Test would tap all the attributes of the hypothetical cancer. It ends at;

Initiation : Non-mutagenic  
Multi-events : Present and Non-mutagenic  
Promotion/Inhibition : Present, non-simulatable  
Metabolic Variation : Polymodal (discrete subgroups)  
Hormonal Influences : " ( " " )  
Immune Surveillance : " ( " " )

in which case the Ames Assay would measure none of the attributes of the cancer.

There are relatively few combinations of factors which allow complete comprehensive validity. If maximum and minimum probability estimates were given to each factor based on a thorough search of the re-

search literature, one could generate summary estimates of the probability that a truly carcinogenic exposure would be biologically amenable to the Ames Assay. This would be a long and difficult task, and numerous assumptions would have to be made given current knowledge. Moreover, the probabilities would have to be altered depending on the histological type of tumour and its site. The task would nonetheless be important to carry out in order to structure the nature of important gaps in our basic science understanding of carcinogenesis.

## CHAPTER IV

### PROGNOSTIC ACCURACY AND SUBSTITUTION

The prognostic accuracy of a predictive test may be ascertained when the outcome it predicts is measurable and when the test data are practicable to collect. In the case of environmentally-induced cancer, the outcome (cancer mortality and/or morbidity) is concrete and measurable, but great difficulties exist in the collection of Ames Assay exposure data. To evaluate whether the difficulties are insurmountable, this section will explore some of the requirements for a methodologically acceptable study of the Ames Assay's prognostic accuracy. )

What would be an acceptable study environment?

(1) In order to ensure an inception cohort, the study would have to focus on individuals whose exposure to the environment of interest began after the beginning of the sampling period. In the industrial environment this would probably involve a work place where a large scale hiring program coincided with the initiation of exposure monitoring.

(2) There would have to be significant between-worksites variation in exposure so that internal controls could be generated, or else an acceptable control group whose environment could be monitored on an ongoing basis.

(3) The hiring program would have to be "large enough" that differences in cancer mortality 20 or more years later would be demonstrable, if they existed, when between-worksites comparisons were made.

Moreover, there would have to be some guarantee that a "sufficient" number of workers would have long-term exposure.

(4) The cohort would have to be "old enough" for a "sufficient" number to reach the critical age range for adult onset of cancer (approximately 45 to 65 years).

(5) There would have to be some sort of agreement to not modify work processes and exposures even in areas with consistently high mutagenic activity.

Taken together, these conditions would seldom be met. Even if they were, the benefits of collecting twenty years of exposure data would be limited. The results would only be generalizable to similar exposures elsewhere and not to all mutagenic exposures. So a direct measurement of prognostic accuracy is sufficiently impractical to make it unascertainable. In its place, a series of substitutions can be made which collectively address the predictive validity of the Ames Assay.

#### A. Substitution

Figure 4-1 illustrates the central question concerning the qualitative performance of the Ames Assay: What are the sensitivity, specificity, and predictive value (positive and negative) of the Ames Assay in identifying observable increased risk of cancer among an exposed population? <sup>28</sup> In order to generate the values which belong in each cell of the four-fold table, a series of comparisons would have to be made (in true prospective fashion) between Ames Positive exposed and non-exposed

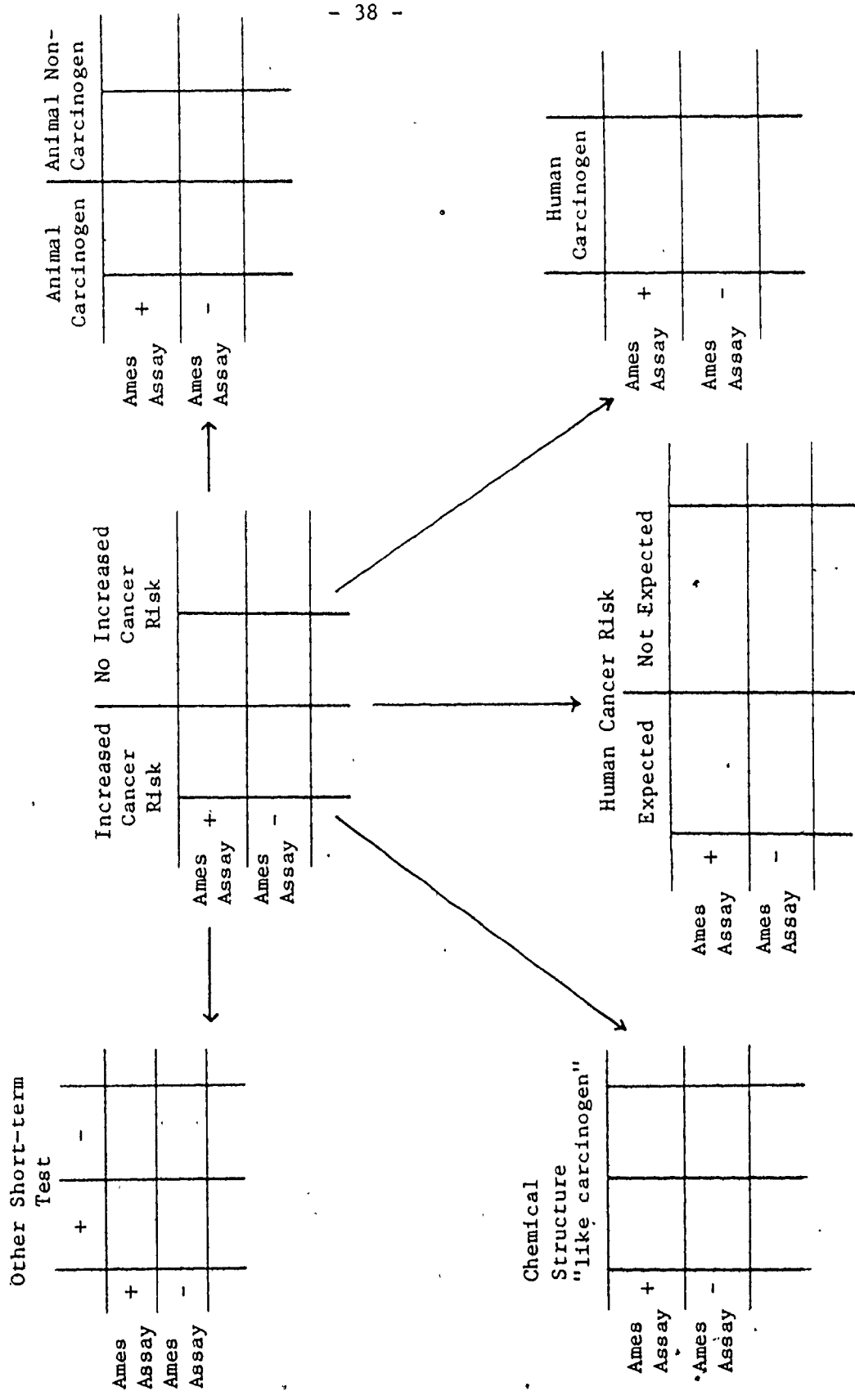
populations. Each data point in the table would represent a unique comparison from a complete prospective study or one of a series of sub-populations from a multi-group study such as the INCO study.<sup>2,9</sup> Surrounding the central question in figure 4-1 are a series of substitute outcomes which have been recognized in the literature. These include: agreement with other short-term bioassays, Ames Assay sensitivity to chemical classes of carcinogens, sensitivity to known human carcinogens, and accuracy in identifying situations where human cancer risk is or is not suspected, due to the results of previous epidemiologic studies. This chapter presents those substitutions. Chapter 5 describes a method to estimate indirectly the four major parameters of the Ames Assay using substitution data. Chapter 6 explores the question of the Ames Assay's validity as a quantitative measure of carcinogenic risk.

#### B. Chemical Properties of Animal Carcinogens

The generally-accepted conclusion among basic scientists studying chemical carcinogens is that, with few exceptions, they are all strong electrophilic reactants.<sup>11</sup> This means that "the known ultimate carcinogens contain relatively electron-deficient atoms that seek to react with nucleophilic sites, i.e., atoms that have easily shared electrons". These sites are abundant on the major intra-cellular macromolecules: DNA, RNA, and proteins. Correlations between the quantities of macromolecule-bound "suspect carcinogens" and the likelihood of tumour development have been consistently identified in animal studies. The great majority of chemical carcinogens are active only after metabolism to an electrophilic form; except when the agents are already



Figure 4-1: Substitutions Available for Prognostic Accuracy



strong electrophiles.

Rinkus and Legator<sup>30</sup> divided 465 substances with evidence of animal carcinogenicity into 39 categories, according to the nature of their electrophilic centres (if present). A sufficient number of chemicals from 26 categories had been previously Ames tested to evaluate its responsiveness to the given category. At least one member from 24 to 26 categories demonstrated mutagenicity; but the proportion of Ames Assay mutagens within each category ranged from 100% to 0%. Several chemical mechanisms were identified which may explain the wide variations in within-category sensitivity:

1. The compound may be metabolized to the electrophilic form in the intestine (but not by  $S_9$ ).
2. The compound may require intra-cellular enzyme systems other than the  $S_9$  fraction to be activated to the ultimate carcinogen.
3. The compound itself is toxic to the test bacteria.
4. The compound operates through a hormonal mechanism.

One of the two categories of substances with 0% Ames responsiveness was the "steroid" group, in line with exception four. The other was the "phenyl" group whose mechanism of activation has not been satisfactorily elucidated.

Is the sensitivity of the Ames Assay to at least one member of each category of animal carcinogen, an important substitute outcome? The principles of organic chemistry which govern the reactions of carcinogens with intra-cellular macromolecules are constant

across all life forms. But the mechanisms of chemical activation, and therefore, the range of ultimate carcinogens, may vary widely between species. As our focus narrows from the universe of electrophiles to specific subgroups to specific compounds, knowledge of activating mechanisms, derived from animal and in vitro experimentation, become more precise as the generalizability of this information to human carcinogenicity decreases. So there is an apparent trade-off between accuracy and precision if evidence of animal carcinogenicity and animal metabolic pathways are taken as the criterion standard for evaluating the Ames Assay. The conclusion of this author is that the sensitivity of the Ames Assay to at least one member of a category of animal carcinogens is an important finding in its own right. It accepts the notion that the common factor among a group of animal carcinogens is its most important attribute. The performance of the Ames Assay with respect to animal carcinogens and non-carcinogens is evaluated later in the chapter.

### C. Other Short-term Bioassays

A literature review uncovered eight studies of agreement between the Ames Assay and other short term tests, where the range of substances tested was wide enough to consider them general studies of agreement. These encompassed seven tests by four different groups of investigators. This section will briefly describe the tests and discuss the level of agreement with the Ames Assay.

#### 1. Mammalian Cell Transformation Test<sup>12</sup>

The mammalian cell transformation test operates on the same

principle as the Ames Assay. Cells exposed to the tester substance (with or without the S<sub>9</sub> fraction) may be "transformed" to malignant cells by the exposure. A positive result occurs when the transformation frequency surpasses the expected spontaneous, cell-type specific frequency.

2. Degranulation Test<sup>12</sup>

The degranulation test is based on the observation that degranulation of rough endoplasmic reticulum and a corresponding increase in smooth endoplasmic reticulum are early changes seen in the secretory organs of animals treated with some experimental carcinogens. Liver rough endoplasmic reticulum, containing radiolabelled RNA, is isolated and the RNA to protein ratio is measured. A statistically significant increase in degranulation of the rough endoplasmic reticulum by the test compared with negative controls is taken as a positive result.

3. Sebaceous Gland Suppression Test<sup>12</sup>

The sebaceous gland suppression test is based on the observation that carcinogens may produce regression of a test mouse's sebaceous glands when applied topically. The test outcome is the change in ratio of sebaceous glands to hair follicles after a total of three days of treatment and four days post-treatment.

4. Tetrazolium Reduction Test<sup>12</sup>

Tetrazolium reduction takes place in the presence of certain dehydrogenase enzymes. A good experimental correlation has been found between the carcinogenic potency of certain animal carcinogens and the amount of tetrazolium reduction in certain mice tissue in vivo. The

test outcome involves the quantity of tetrazolium reduction taking place in a 2 day period after skin-painting the test solution on the dorso-lumbar region of a mouse.

5. Implant Test<sup>12</sup>

Subcutaneous implantation of solid materials in experimental animals has repeatedly produced results invalid for inference to humans. Inert materials have caused tumours, and characteristics such as shape of material, presence or absence of holes, and powdered versus solid form have governed their carcinogenicity. To correct for these problems the implant test delivers the test substance on a porous material whose characteristics make it non-carcinogenic to the test species. Short-term histological changes surrounding the implant are scored based on previous evidence of their predictive value for future tumour development.

6. DNA - Synthesis Inhibition Test<sup>27</sup>

Painter demonstrated that damage to DNA causes temporary inhibition of DNA synthesis in mammalian cells. This is measured by determining the rate of inclusion of radiolabelled thymidine into the DNA of HeLa cells after exposure to a tester substance (with or without the inclusion of the S<sub>q</sub> fraction). A positive result occurs when thymidine inclusion is reduced in the test cells relative to the controls.

7. Micronucleus Test<sup>31,32</sup>

The micronucleus test measures chromosomal breakage by detecting chromosome aberrations in actively-dividing bone marrow cells. Some chromosome fragments, which lack a centromere, may be left in the cytoplasm when daughter nuclei are formed. These fragments become micro-

nuclei. A positive result occurs when there is a statistically significant increase in the frequency of micronuclei in the presence of a tester substance compared to a control (by small sample t-test).

Table 4-1 presents the agreement found between these tests and the Ames Assay. This author used the raw data presented in the published papers to compute the Kappa statistic for 2 x 2 tables of agreement.<sup>33</sup> Interpretation of the Kappa values depends on the methodologic calibre of the studies and the nature of the attribute being tested in each bioassay.

The data generated by Purchase et.al.<sup>12</sup> involved laboratory personnel blinded to the identity of the test chemical and the paper described in explicit terms how a substance was defined positive or negative. The other papers did not mention blind laboratory analysis and were less explicit as to method. None of the papers stated that the test substances had been randomly selected from a universe of candidate substances. Most important, the Ames Assay results reported by Jenssen et.al.<sup>31</sup> were not generated concurrently with the micro-nucleus test results. They were taken from the literature. On the other hand, Heddle et.al.<sup>32</sup> generated their own Ames Assay results. The Kappa statistic suggests slight to moderate agreement with the Jenssen data, but no agreement with the Heddle data. This discrepancy may be explained by a selection bias involving fortuitous choices of tester substances by Jenssen et.al.

With these limitations in mind, the data suggest a high degree

Table 4-1: Agreement Between Ames Assay and Other Short Term Tests

Investigators	Test	Kappa	95% Confidence Interval
Purchase et.al. <sup>12</sup>	Cell Transformation	.799	.977 — .621
Purchase et.al. <sup>12</sup>	Degranulation	.361	.541 — .181
Purchase et.al. <sup>12</sup>	Sebaceous Gland Suppression	.328	.507 — .149
Purchase et.al. <sup>12</sup>	Tetrazolium Reduction	.035	.208 — .138
Purchase et.al. <sup>12</sup>	Implant	.319	.478 — .160
Painter et.al. <sup>27</sup>	Synthesis Inhibition	1.00	1.00 — .511
Jenssen et.al. <sup>31</sup>	Micronucleus	.352	.528 — .146
Heddle et.al. <sup>32</sup>	Micronucleus	.123	.362 — -.116

of agreement between the Ames Assay and the Mammalian Cell Transformation and DNA synthesis Inhibition Tests. Moderate to slight agreement is observed with the Degranulation, Sebaceous Gland Suppression, and Implant Tests. No agreement is observed with the Tetrazolium Reduction and Micronucleus Tests. If correct, the high level of agreement with the DNA synthesis Inhibition Test provides evidence that the Ames Assay is measuring what it claims to measure since inhibition of DNA synthesis, like Ames Assay revertant counts, is an indirect measure of mutagenic activity. The investigators found a quantitative relationship between the number of revertant colony counts and the degree of DNA synthesis inhibition when between-substance comparisons were made. If the high level of agreement with the Cell Transformation Test is correct, it provides evidence that the attribute being tested is at least a correlate of a process which can transform mammalian cells to malignant forms in vitro. This bears primarily upon the stages of initiation and promotion in the flow-chart of carcinogenesis and suggests that the Ames Assay is measuring an attribute of carcinogenesis, not just mutagenesis.

The Implant Test, Degranulation, Sebaceous Gland Suppression, and Tetrazolium Reduction are all tests of an empirical correlate of animal carcinogenesis. Of these the Tetrazolium Reduction Test is the most specific in that it assays for a particular enzyme system. The other tests measure a morphological change. The issue in interpretation here is similar to that dealt with in the section concerning chemical structure: the trade-off between precision and accuracy. Moderate to



slight agreement was found with the more non-specific tests of morphological change. No agreement was found with the more specific test. Does this mean that the accuracy of a bioassay is decreased as the attribute it measures becomes narrower (from morphological change to enzyme-system-based assay)? Or does it mean that between-species variability is a practical limiting factor in the accuracy of short-term tests?

The Micronucleus Test is important since it measures a phenomenon which should correlate with mutagenesis: chromosome breaks. In fact, both agreement tables showed a marked bias in the direction of discordant results, the Ames Positive/Micronucleus Negative pair outnumbering the Ames Negative/Micronucleus Positive pair by more than two to one in both. This observation is consistent with the concept that chromosome breakage is a less subtle phenomenon than mutagenesis. So it is reasonable that mutagenesis would be found in the absence of chromosome breakage more frequently than the reverse. The significance of this observation for the validity of either test is uncertain.

#### D. Populations with Expected Risk of Cancer

##### 1. Smokers' Urine

Yamasaki et.al.<sup>34</sup> examined the urine of 10 smokers and 21 non-smokers for mutagenicity in the Ames Assay. No description was given regarding the selection of the study subjects or the blindness of the laboratory personnel to their identity. None of the 37 samples from the 21 non-smokers showed mutagenicity in the urine. Among the 10 smokers, 7 had mutagenic urine. The mutagenic potency of the urine cor-

related well with the reported number of cigarettes smoked per day and their tar content. Among the 3 smokers with non-mutagenic urine, 2 claimed to not inhale and the third was smoking low-tar cigarettes on the day of urine collection. A similar experiment was carried out locally with the laboratory personnel blind to the identity of the urine donors. The results agreed with Yamasaki et.al. However, only 2 smokers and 2 non-smokers were tested.

2. Exposure to Cytostatic Drugs

Falck et.al.<sup>35</sup> assayed the urine of cancer patients taking a mixture of cyclophosphamide and various other cytostatic drugs, the nurses who mixed the various agents, and an unexposed control group of office clerks and psychologists. The urine samples were collected from all study subjects at the same time on the same day. The group of exposed nurses also gave urine samples after a weekend off work. All the subjects were non-smokers having normal dietary habits. No information was given as to randomness of subject selection or blindness of subjects and laboratory analysis.

The investigators reported the data in terms of the ratio of revertant colonies found on solvent-only control plates to plates with urine samples. The  $R_{\text{test}}/R_{\text{control}}$  ratio was 1 for the unexposed control group, 1.5 for the nurses after a weekend off, 2 for the nurses during the work week and 11 for the chemotherapy patients. The investigators report that the means of the revertant colony counts were all statistically significantly different from one another in series of t-tests ( $p < .01$  in all cases). However, the analysis is flawed by discrepancies

between the methods section and the data reported. Two assays of mutagenesis were used. The second was based on *E. coli* bacteria. It is unclear from the paper which samples were analyzed which way. The statistical conclusions are ambiguous as to which test they refer to and nevertheless involve questionable statistical methods. These uncertainties make this paper uninterpretable.

Evidence to suggest that cyclophosphamide is carcinogenic to the urinary bladder comes from three case series of patients undergoing long-term therapy<sup>36,37</sup>. No controlled studies have been reported. Although the relationship is consistent across these series, fits the temporal sequence of cause preceding effect, and is biologically sensible, none of the other diagnostic tests of causation have been met. The role of the other drugs in carcinogenic risk is even more uncertain.

### 3. Exposure to Coke Plant Emissions

Increased risk of lung and bladder cancer among coke plant workers has been documented in three cohort studies. The association meets the 8 diagnostic tests of causation possible in sub-experimental research. Moller and Dybing<sup>38</sup> compared early morning and after work mutagenicity among four groups: 10 smoking coke plant workers, 10 non-smoking coke plant workers, 6 smoking office workers, and 4 non-smoking office workers. The exposed smokers tended to work in higher exposure areas than the exposed non-smokers and the non-exposed smokers tended to smoke more than the exposed. Because of these biases, and the fact that the data were analyzed with a series of t-tests, rather than by regression or a two-way ANOVA, only the descriptive out-

comes are meaningful. Figure 4-2 illustrates the results. Urine mutagenicity appears to be zero among non-exposed non-smokers, slight, but measureable among exposed non-smokers, higher amongst non-exposed smokers, and highest amongst exposed smokers. The increase in urine mutagenicity during the day is identifiable in both groups of smokers and the exposed non-smokers.

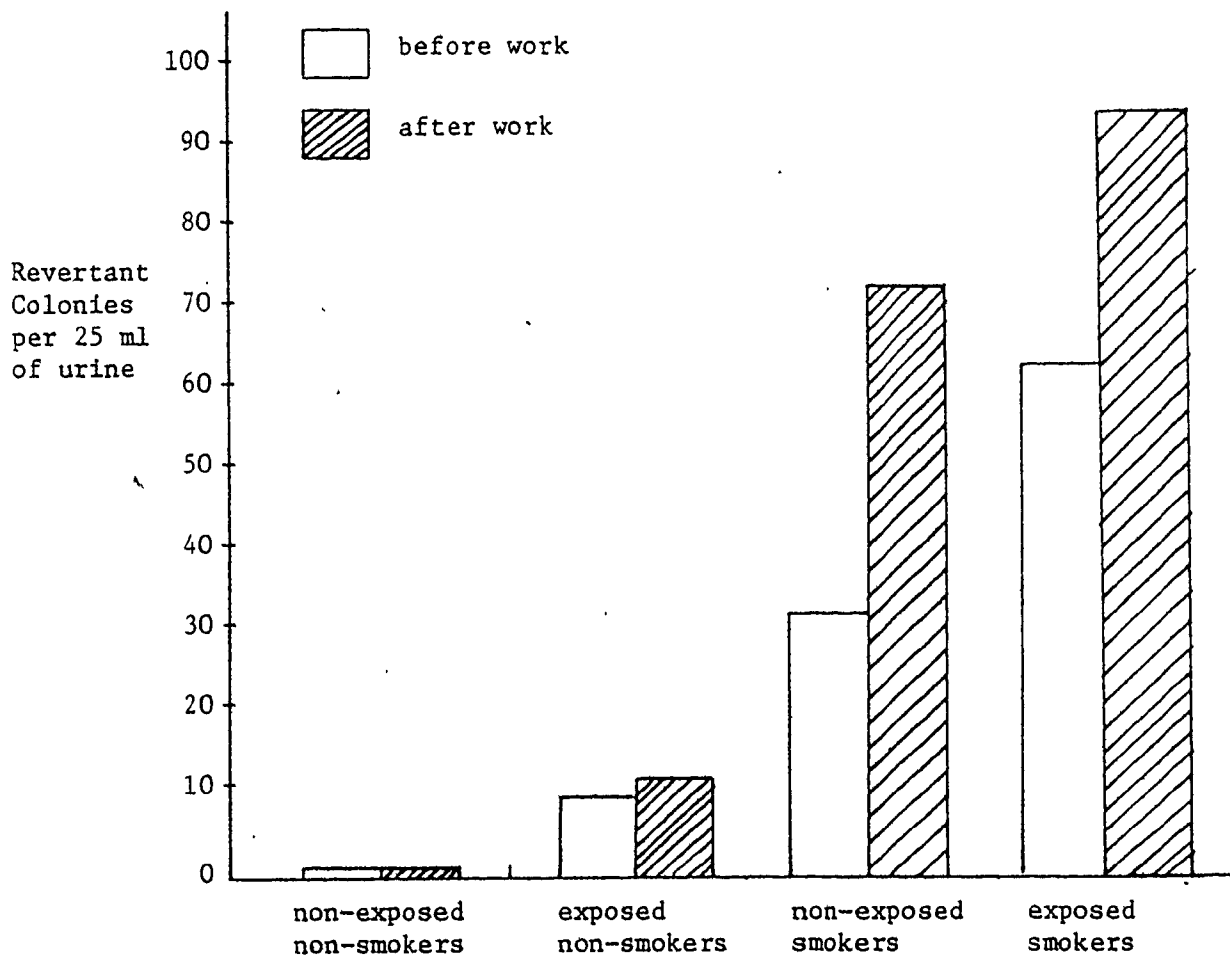
#### 4. Exposures in the Rubber Industry

Several mortality studies have been carried out among workers in the rubber industry.<sup>39, 40</sup> A wide variety of associations have been found between several cancer sites and several different job classes within the rubber industry. These associations tend to be inconsistent across studies. However, the manufacture of tires: building, curing, and assembling, shows up consistently as an "at risk" job class. The malignancies reported to show excess mortality include lung, bladder, skin, brain, and stomach cancer, as well as leukemias and lymphomas. Falck et al.<sup>41</sup> assayed the urine of smoking and non-smoking tire makers and compared them with unexposed smoking and non-smoking office clerks. All study subjects were males. The range of cigarette habit was implied to be comparable between the exposed and non-exposed smokers. All urines were taken the same Thursday afternoon and the data were standardized by creatinine clearance using the following formula:

$$\text{Mutagenic Activity} = \frac{\text{Revertants in sample (Rs)} - \text{Revertants in solvent control (Rc)}}{\text{creatinine (n moles) in urine sample}}$$

The data were originally analyzed by means of all possible combinations of t-tests. Re-analysis using a two-way ANOVA approach demonstrated a

Figure 4-2: Mutagenic Urine with Exposure to Coke Plant Emissions



highly significant smoking effect and a highly significant effect due to exposure to tire-making (see figure 4-3). These results are consistent with the expectation of increased cancer risk due to both cigarette smoking and working in the rubber industry.

#### 5. Ecological Variations in Fecal Mutagens

The age-adjusted colon cancer incidence rates for white males in the United States was 28.5 per 100,000 in the mid-1970's, but was only 5.6 per 100,000 in rural Finland (Kuopio)<sup>4,2</sup>. Among North American Seventh-Day Adventists (who consume little or no meat), the incidence rate is approximately 60% of the general population ( 17 per 100,000)<sup>4,2</sup>. Reddy et.al.<sup>4,2</sup> collected feces samples from 15 middle-aged healthy male volunteers from Kuopio, 11 healthy male Seventh-Day Adventist vegetarians from New York City, and 18 healthy non-vegetarian males from New York City. Despite the unequal numbers of subjects in each group, the investigators claim they "matched" for age. Two-day fecal samples were pooled, extracted, and plated using both TA 98 and TA 100, with and without S<sub>9</sub>.

The results were reported in an awkward manner, but the following general observations can be made:

(1) Fecal mutagenesis was consistently above control values and rose with increasing fecal mass plated for the Non-Seventh-Day Adventists TA 98/-S<sub>9</sub>, TA 100/-S<sub>9</sub>, and TA 100/+S<sub>9</sub> assays. Mutagenesis was generally not above control values with TA 98/+S<sub>9</sub>.

(2) Fecal mutagenesis was above control values and rose with fecal mass plated for the rural Finlanders in TA 98/+S<sub>9</sub> assays, but

Figure 4-3: Analysis of Variance: Urine Mutagenicity as a Function of Smoking and Exposure to Tire-making Emissions

<u>Source of Variation</u>	<u>Sum of Squares</u>	<u>DF</u>	<u>Mean Square</u>	<u>F</u>	<u>p(F)</u>
Smoking	1473083.500	1	1473083.500	44.998	<.001
Exposure to Tiremaking	599042.125	1	599042.125	18.299	<.001
Interaction of Smoking and Exposure	163719.750	1	163719.750	5.001	.032
Residual	1047564.000	32	32736.375		
Total	3379599.500	35	96559.984		

none of the rest.

(3) In general, fecal mutagenesis was not above control values amongst the Seventh-Day Adventists.

These data, if reliable, are in partial agreement with the expectation of risk of colon cancer. But they are inconsistent with the difference in risk between Seventh-Day Adventists and rural Finlanders.

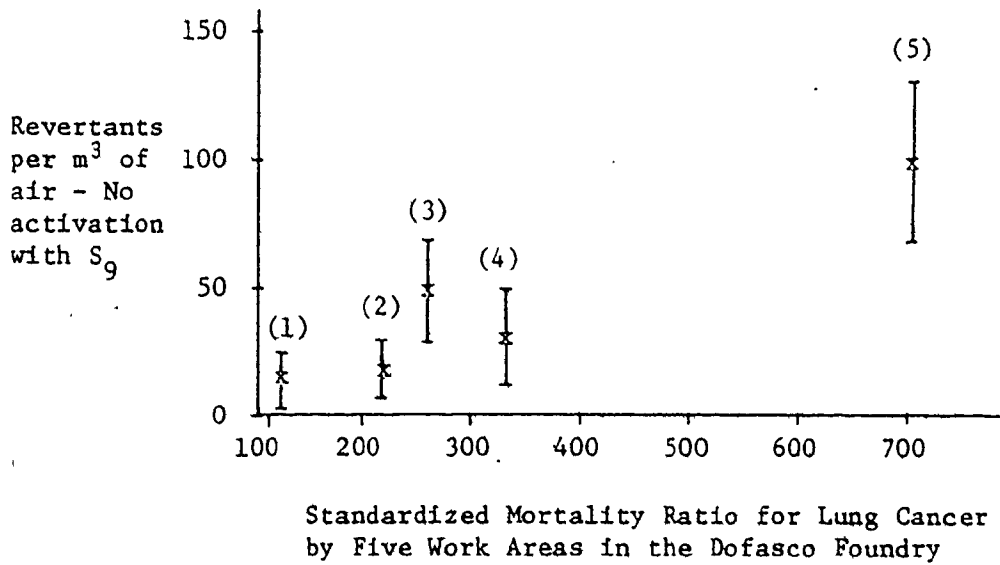
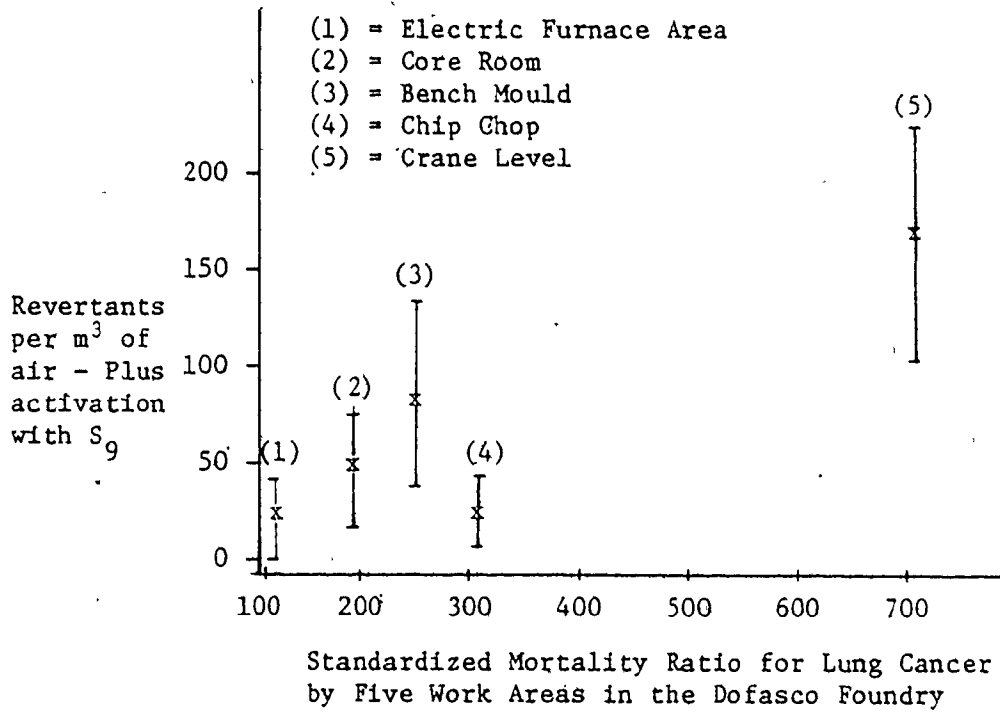
#### 6. Exposure in a Steel Foundry

Gibson et.al.<sup>43</sup> studied lung cancer mortality within a cohort of steel foundry workers at Dofasco's Hamilton foundry. The cohort, 1,542 strong, included those aged 45 years and over in 1967. Lung cancer mortality was ascertained over the 10 year period 1967-1976 and the mortality experience at 5 job sites within the foundry was compared with that of the non-foundry Dofasco population. Standardized mortality ratios ranged from 114 for electric furnace/open hearth workers to 714 for Crane Operators. Only the SMR for Crane Operators was statistically significantly different from 100 ( $p < 0.01$ ).

An industrial hygiene study carried out in 1976-77 failed to show a correlation between polycyclic aromatic hydrocarbon levels and the variation in SMR's.<sup>43</sup> Air samples were collected from the five work areas over a six-day period in November, 1979. The samples were chemically extracted and plated with TA 98, with and without  $S_9$ . The mean counts per cubic metre of air sampled are plotted in figure 4-4 against the SMR for lung cancer at each work site. The confidence limits surrounding each work site mean are based on day-to-day variability and not laboratory variability.



Figure 4-4: Variation in Steel Foundry Lung Cancer Rates



The SMR is plotted as the independent variable because it preceded the Assay results in time. So the question addressed is: Does epidemiologic risk predict Ames Assay counts? rather than: Do Ames Assay counts predict epidemiologic risk? Nevertheless, four of five work sites' revertant counts rise with the SMR for lung cancer, with the Crane Level highest. The only outlier is the Chip Shop area. This is the only work site which Gibson et.al. describe as having undergone a process change within the past 10 years.

Tests of significance have not been reported here because of difficulties with the assumptions of linear regression. Nonetheless, these results are promising for the validity of Ames Assays linked to air sampling.

E. Sensitivity of Ames Assay to Known Human Carcinogens

Although the accuracy of the Ames Assay in identifying known human carcinogens and non-carcinogens is an important characteristic for assessing its validity, several problems obstruct a direct assessment.

(1) Evidence of human carcinogenic risk usually comes from epidemiologic studies where exposure is to a mixture of substances. It is usually impossible to identify which substance(s) are the active carcinogens. However, most Ames Assay information comes from studies of pure compounds. So the number of cases where the exposure of interest has been replicated in the laboratory is limited.

(2) The number of "demonstrated" human carcinogens varies with

Figure 4-5: Sensitivity of Ames Assay to Human Chemical Carcinogens

Human Carcinogens Positive in  
Ames Assay

aflatoxin  
4-aminobiphenyl  
auramine dye mixture  
benzidine  
chlornaphazin  
bis (chloro methyl) ether  
chloroprene  
cigarette smoke condensates  
coal tar  
cyclophosphamide  
melphalan  
mustard gas  
B-naphthylamine  
4-nitro biphenyl  
soot  
vinyl chloride

Human Carcinogens Negative in  
Ames Assay

benzene  
diethylstilbestrol

the degree of critical appraisal given the evidence.

(3) There is no accepted definition of a non-carcinogen and no generally accepted list of "demonstrated" human non-carcinogens.

Using a list of 18 "generally accepted" human chemical carcinogens (excluding arsenic, chromates, nickel compounds, ionizing radiation, and asbestos),<sup>7,44</sup> the Ames Assay is sensitive to 16 (see figure 4-5). The exceptions include diethylstilbestrol and benzene. The former false negative is expected because its mechanism of action is unknown. The latter false negative is disturbing because of benzene's importance as an industrial carcinogen and because it is a parent molecule to which many other industrial compounds are related.

#### F. Agreement Between the Ames Assay and Animal Studies

Because Animal Studies have been considered the criterion standard for assessing carcinogenicity in substances, agreement between animal work and the Ames Assay is an important validity check. Five published studies have looked at a broad range of animal carcinogens and non-carcinogens in conjunction with independent Ames Assay screening. This section presents the results of those studies, re-analyzed to give Cohen's Kappa<sup>33</sup> for a four-fold agreement table. Each study has been given a Methodologic Score between zero and 11.5, based on the following criteria:

1. (a) Were the criteria for labelling a substance an Animal Carcinogen explicitly stated? (If yes, 1 point)

- (b) If yes to (a), were the criteria rational and/or accepted, or were they irrational and/or sufficiently peculiar to make the results non-generalizable? (1 point)
2. (a) Was the method of selection of test substances clearly stated? (If yes, 1 point)
- (b) Were they a: stratified (by chemical class) random sample? (2 points)  
simple random sample? (1.5 points)  
judgement/representative sample? (1 point)  
judgement/non-representative sample? (0 points)
3. Was laboratory testing stated to have been carried out blindly? (2 points)
4. (a) Was more than one class of bacteria used? (If yes, .5 points)  
Were R-factor plasmid strains used? (If yes, .5 points)
- (b) Was Sq activation/no activation done with each bacteria used for each substance? (If yes, 1 point)
- (c) Were negative controls used? (If yes, .5 points)
- (d) Were positive controls used? (If yes, .5 points)
- (e) Was bacteriotoxicity accounted for? (If yes, .5 points)
- (f) Were rules of evidence for +/- outcome given? (If yes, .5 points)  
Were these rules rational/accepted (i) or irrational/non-generalizable (ii)? (If (i), .5 points)

The results are presented in Table 4-2, along with the Kappa values for each study.

Table 4-2 shows that the level of agreement is largely independent of the studies' methodologic rigour. The "best" study in method-

Table 4-2: Agreement Between Animal Studies and the Ames Assay

Investigators	Methodologic Score	Cohen's Kappa	95% Confidence Interval
1. Purchase et.al.	9.5	.85	1.00 ↔ .67
2. Poirier	3.0	.74	1.00 ↔ .45
3. McCann et.al.			
a) excluding co-carcinogens and weak carcinogens	5.5	.66	.81 ↔ .51
b) including them	5.5	.55	.68 ↔ .41
4. Heddle and Bruce	5.5	.47	.72 ↔ .22
5. Odashima	3.0	.35	.60 ↔ .10

Kappa based on the following four-fold table:

	Animal Studies		
	+	-	
Ames Assay	+	A      B	A+B
	-	C      D	C+D
	A+C	B+D	A+B+C+D

$$K = \frac{(A+D) - \left[ \frac{(A+B)(A+C)}{(A+B+C+D)} + \frac{(B+D)(C+D)}{(A+B+C+D)} \right]}{1 - \left[ \frac{(A+B)(A+C)}{(A+B+C+D)} + \frac{(B+D)(C+D)}{(A+B+C+D)} \right]}$$

ological terms gave the highest level of agreement. The important differences between this study and the rest were:

1. The criteria for animal carcinogens were clearly and explicitly stated.

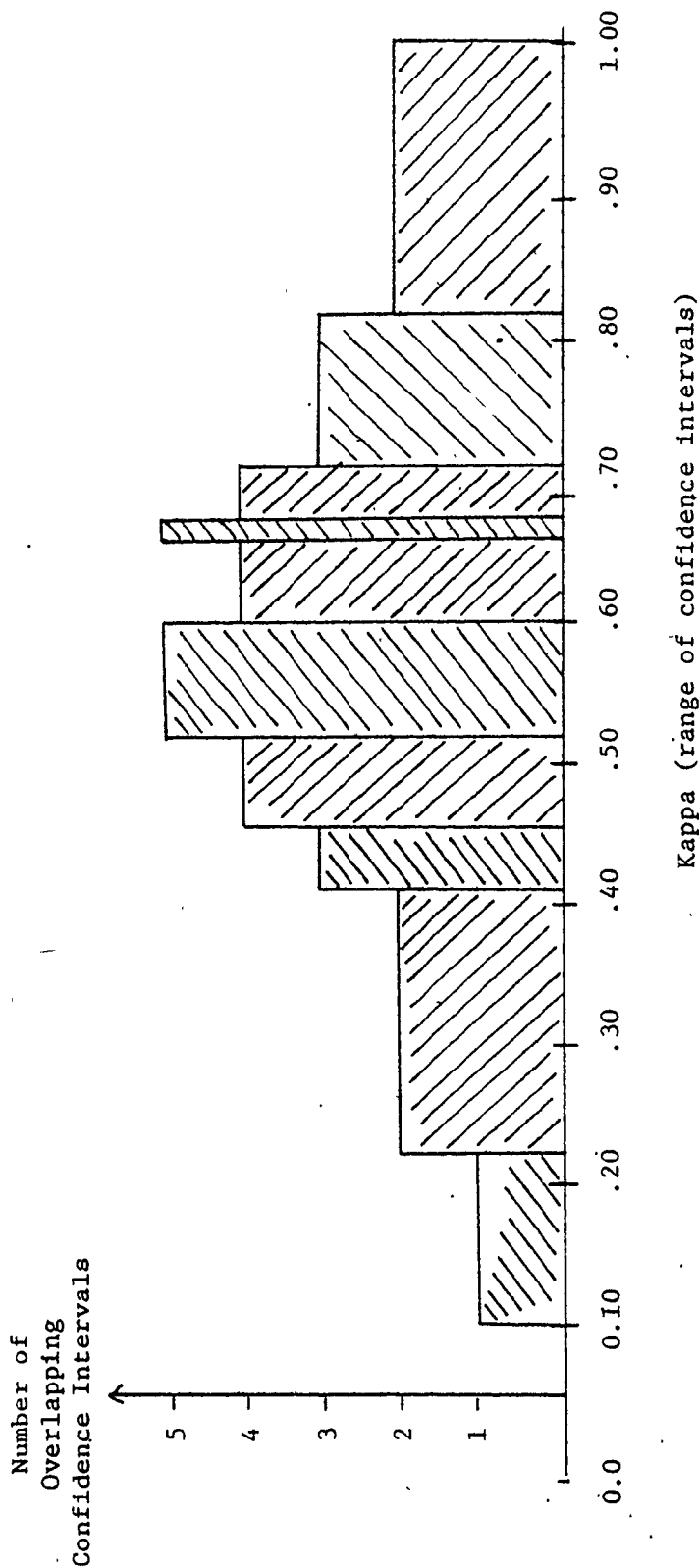
2. Laboratory analysis was carried out blind.

The range of Kappa values includes near perfect agreement and mild agreement. Figure 4-6 shows the overlap of confidence intervals across the range of kappa in Table 4-2. Modal values are found at Kappas of .51 to .60 and .67 to .68, wherein five of six confidence intervals overlap.

#### G. Summary

The results of studies of validity by substitution do not carry with them an objective means of combining evidence and coming up with a summary decision regarding validity. Moreover, many of the studies summarized in this chapter were seriously flawed in their methods or were designed for other purposes. However, in each category evidence exists that the Ames Assay taps the attributes of interest in human carcinogenesis. In the following chapters, especially chapter five, the underlying assumption is tested that the Ames Assay is a test, in that it is better than random selection in identifying human carcinogens. Moreover, it is assumed that the universe wherein Ames Assay and Animal results overlap is the universe of interest: human carcinogenesis (positive agreement) and non-carcinogenesis (negative agreement).

Figure 4-6: Range of Agreement Between Animal Studies and the Ames Assay





## CHAPTER V

### PROSPECTS FOR PRACTICAL APPLICATION AND THE RANGE OF POSSIBLE ERROR RATES OF THE AMES ASSAY

In the previous chapter a series of validation studies were reported which provide an impression of the performance of the Ames Assay. This chapter will explore the question "Can the Ames Assay identify a subset of environments where the cancer risk is truly greater than in other environments?" This begs the further question "What are the sensitivity and specificity of the Ames Assay?" The most useful information in dealing with this question is found in the agreement data between Animal Studies and the Ames Assay.

In assessing the sensitivity and specificity of the Ames Assay, the Animal Study can be seen in two ways; either as a criterion standard against which other tests must be compared or as just another test with quantifiable error rates. The latter approach is more attractive because the set of animal carcinogens is most certainly not precisely the same as the set of human carcinogens. Hui and Walter<sup>48</sup> developed a method of simultaneously estimating the error rates of two or more tests which measure the same condition, but no one of which is a criterion standard. The following conditions are necessary in order for the calculation to work:

- (1) At least two agreement tables are needed between the two tests
- (2) The true prevalence of the outcome of interest must be dif-

ferent in the different tables. (But the false positive and negative rates will be independent of prevalence).

(3) The tests should be based on different sorts of attributes, so that they are subject to independent errors.

(4) There must be six degrees of freedom between the two agreement tables so that six parameters may be estimated:  $\alpha$  and  $\beta$  errors for test 1,  $\alpha$  and  $\beta$  errors for test 2, and estimates of prevalence of the outcome of interest in both agreement tables.

This latter constraint made the method inoperative for the currently available data. Why? When naturally occurring populations are sampled, only the total number of subjects is fixed, leaving three degrees of freedom in the resulting 2 x 2 agreement table. Two independent tables would then give the required six degrees of freedom. In the agreement currently available, the investigators conditioned on the selection of animal carcinogens and non-carcinogens. This is equivalent to setting both column totals of each 2 x 2 table, and reducing the remaining degrees of freedom to four (see figure 5-1). So in order for the technique to be valid, the process determining the underlying prevalence of the outcome of interest must be naturally determined and not concocted by prior knowledge of animal carcinogenicity. Insufficient data of this kind exist at present. So the technique must be held in reserve until adequate samples are available.

If the animal studies are provisionally accepted as the gold standard for human carcinogens, then the six agreement studies reported give six separate estimates of the false positive and false negative error

Figure 5-1: Requirements of Sampling for Hui and Walter's Technique

Table 1

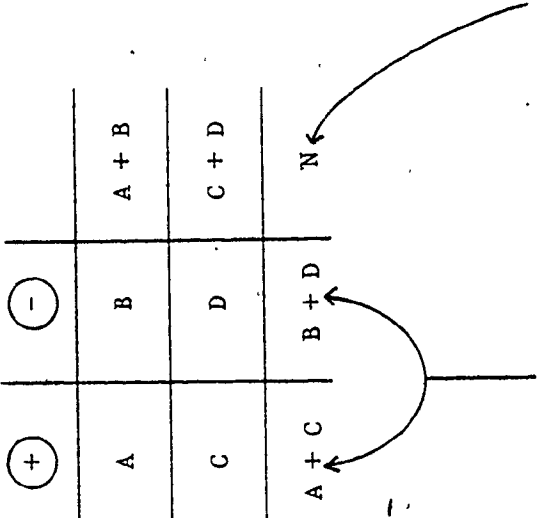
Ames		Assay		N	
(+)	(-)	A	B	A+B	
(-)	(+)	C	D	C+D	
		A+C	B+D		

Table 2

Ames		Assay		N	
(+)	(-)	A	B	A+B	
(-)	(+)	C	D	C+D	
		A+C	B+D		

Sampling direction  
gives 2 df x 2 tables  
= 4 df

Sampling direction  
gives 3 df x 2 tables  
= 6 df



rates of the Ames Assay. If we assume that the real error rates of the Ames Assay fall within the ranges set by these estimates, then we can use the maximum and minimum estimates of each error rate to explore the possible variations in the performance of the Ames Assay with plausible variations in its error rates. Calculations based on the six agreement studies reveal the following estimates:

1. false positive error rate ( $\alpha$  error)

(a) maximum = .21

(b) minimum = .06

2. false negative error rate ( $\beta$  error)

(a) maximum = .44

(b) minimum = .09

Figures 5-2 (a) to (d) show the way in which the positive predictive value of the Ames Assay increases with increasing prevalence of human carcinogens in a hypothetical collection of test samples, and how the four combinations of error rates affect this relationship. (For completeness, the curve showing the probability of an Ames Negative sample being truly carcinogenic has been included, too.) In each graph an identity line called the "Line of Random Agreement" has been included. This is equivalent to the performance of an unbiased coin in assessing carcinogenicity, wherein the false positive and negative error rates would each be .5. Data points on the "enrichment curves" above and below this line represent the subset of Ames Assay positive or negative samples with a greater or lesser proportion of human carcinogens than that found in the complete set of samples. By definition, when the true

Figure 5-2: Relationship of Prevalence of Carcinogenic Exposures to Ames Assay Results

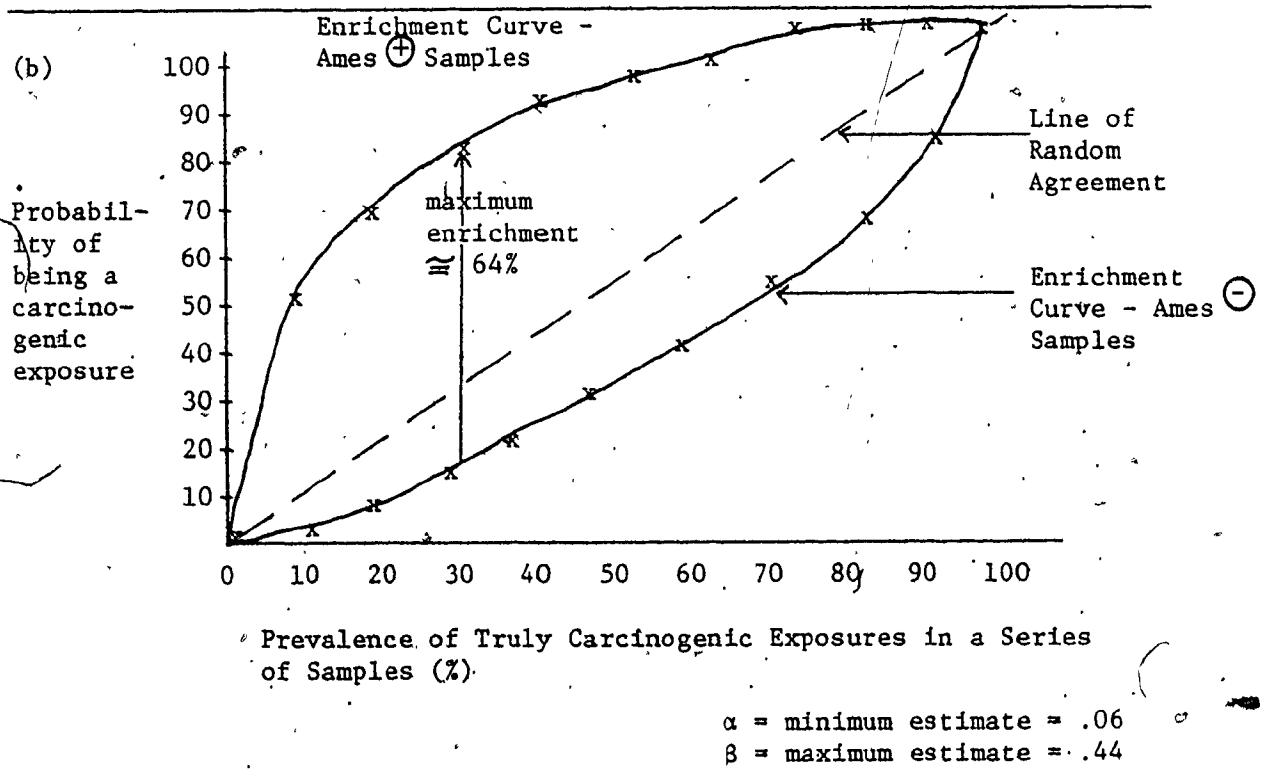
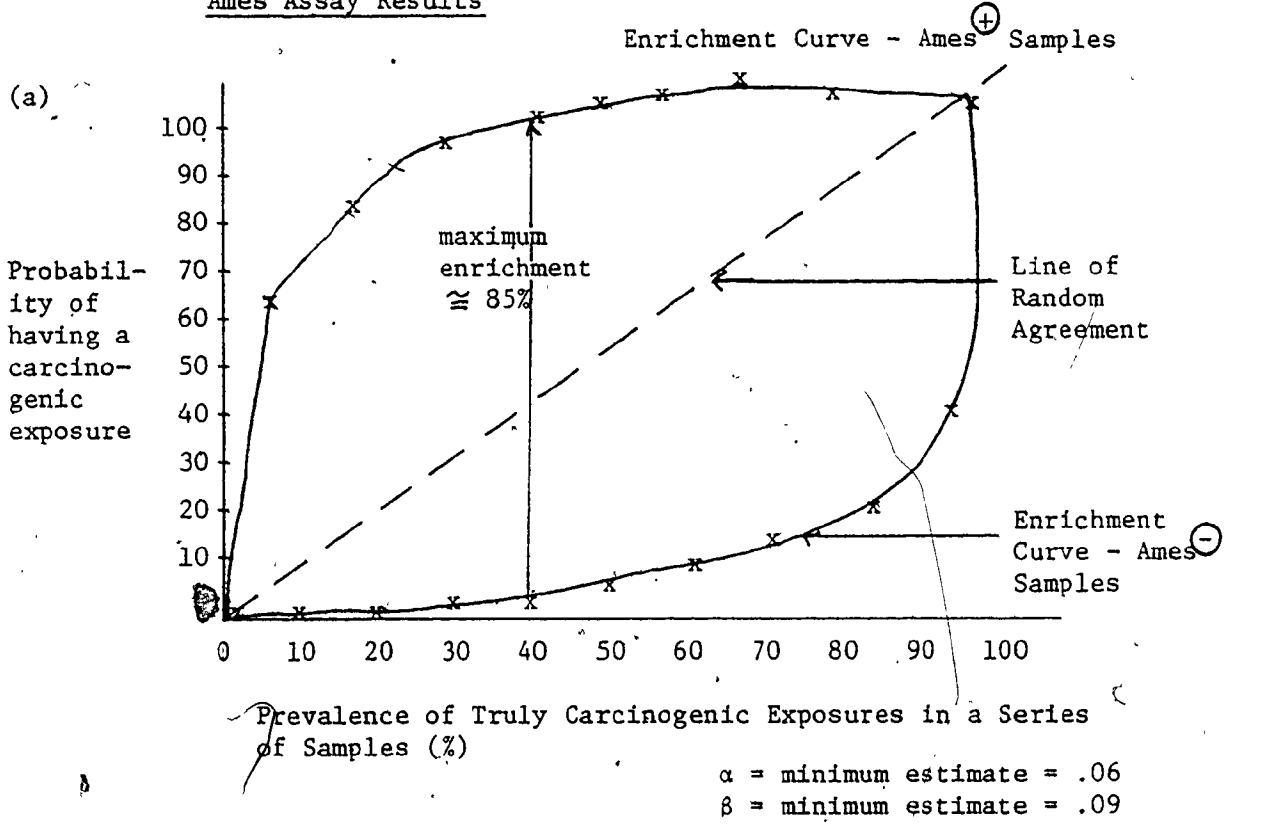
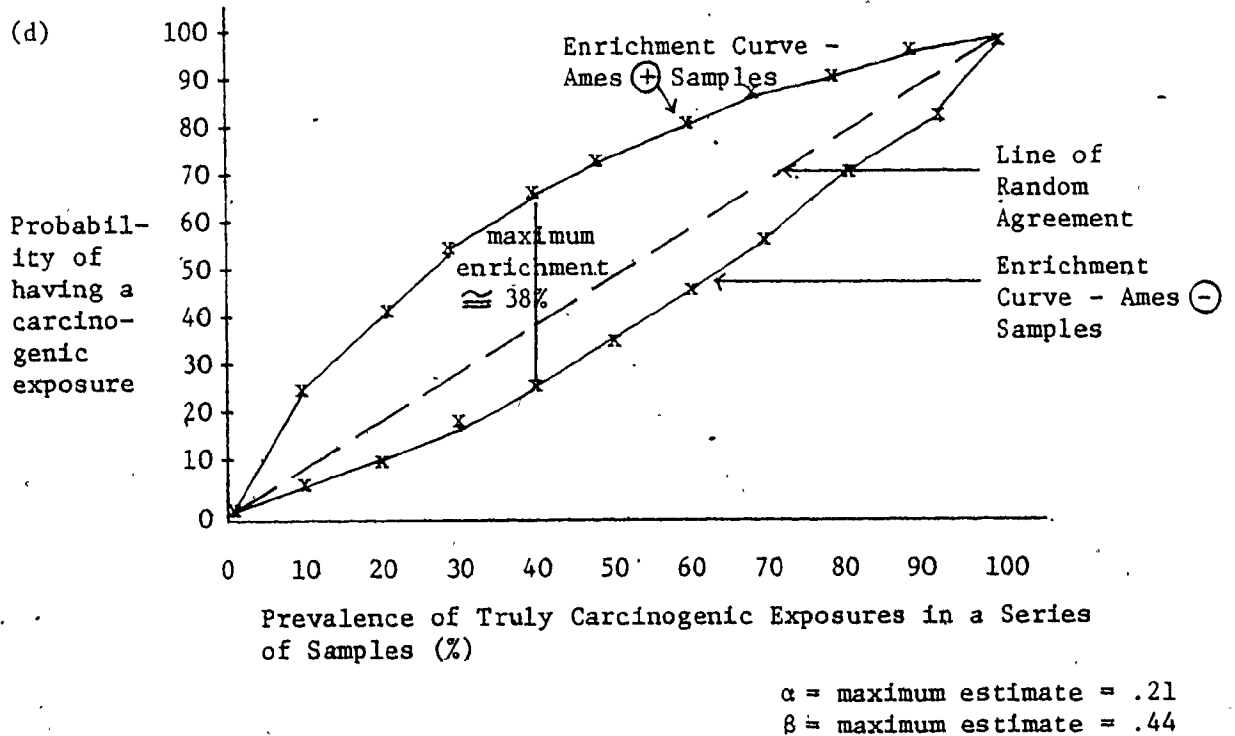
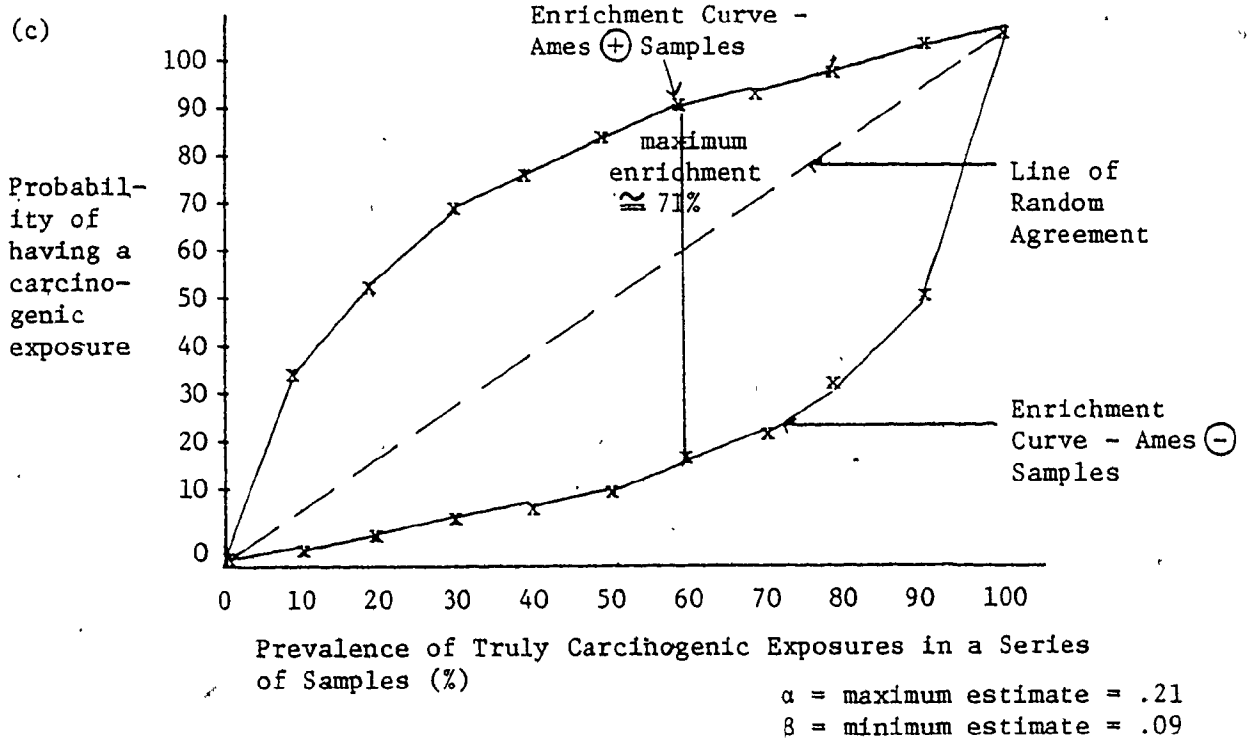


Figure 5-2 (continued)



prevalence of human carcinogens is zero or 100%, no enrichment can take place. This property is illustrated graphically by the way in which the "enrichment curves" meets the "line of Random Agreement" at the 0% and 100% prevalence points. In between, the degree of enrichment rises to a maximum value, defined as the point of greatest difference between the proportion of human carcinogens in the Ames positive subset and the Ames negative subset. The better the test, the greater the area surrounded by the two enrichment curves. Several observations may be made by comparing the four graphs.

1. Some enrichment is evident in each graph, regardless of the error rates. (Range of maximum enrichment: 38% to 85%.)

2. The degree of enrichment rises when either the false positive or the false negative error rate drops.

3. The greatest enrichment occurs when the prevalence of human carcinogens falls above 20% but below 80% of the test samples. Invariably, the degree of enrichment becomes small as the prevalence of human carcinogens rises above 80% of the test samples.

4. The point of maximum enrichment changes with changing error rates: (a) when  $\alpha = .06$  and  $\beta = .09$ , it is at the 40% prevalence point of human carcinogens in the test samples.

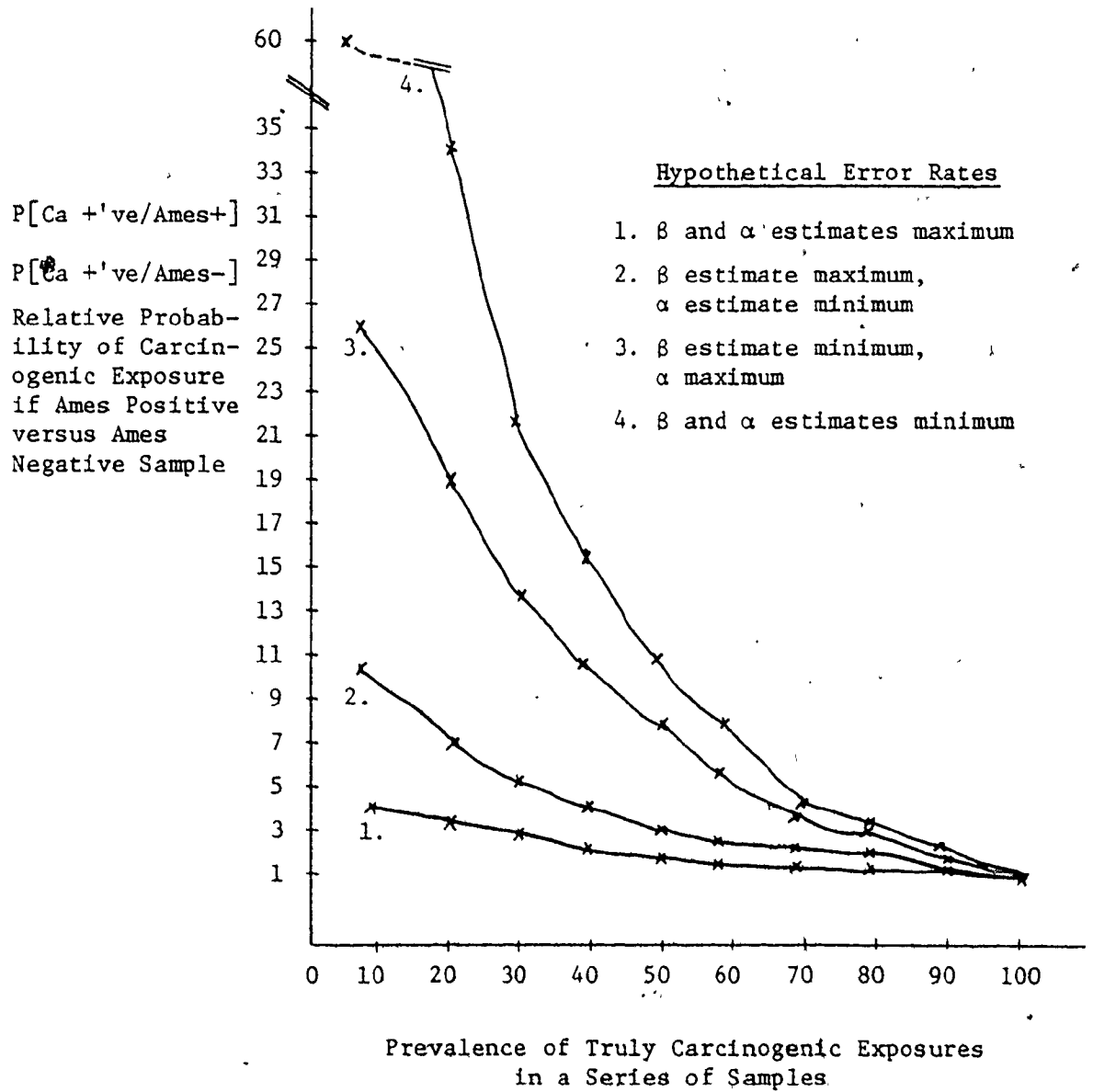
- (b) when  $\alpha = .06$  and  $\beta = .44$ , it is at the 30% prevalence point.

- (c) when  $\alpha = .21$  and  $\beta = .09$ , it is at the 60% point

- (d) when  $\alpha = .21$  and  $\beta = .44$ , it is at the 40% point.

Figure 5-3 gives a different perspective by showing how the relative probability of finding human carcinogens among Ames positive com-

Figure 5-3: Change in the Relative Probability of a Carcinogenic Exposure with Changing Underlying Prevalence of Carcinogenic Exposures





pared with the Ames negative samples varies with varying error rates and the true prevalence of human carcinogens in the test samples. From the series of four curves it can be seen that:

(1) Regardless of the estimates of  $\alpha$  and  $\beta$ , the relative probability drops to 1 as the prevalence of human carcinogens approaches 100%.

(2) As the prevalence of human carcinogens drops, the relative risk rises in all cases. The rate of rise is greatest when  $\alpha$  and  $\beta$  are smallest, such that the relative risk at 10% prevalence of human carcinogens is approximately 60 for  $\alpha$  and  $\beta$  at minimum estimates, and approximately 4 for  $\alpha$  and  $\beta$  at maximum estimates.

(3) The curves showing relative risk against changing prevalence of human carcinogens are not linear. As the  $\alpha$  and  $\beta$  estimates become smaller, they tend to increasingly resemble exponential growth curves which approach infinity as the prevalence of human carcinogens approaches zero. Therefore, the relative risks at low prevalence ranges are disproportionately high compared to high prevalence ranges.

(4) The impact of changes in the false negative error rate is greater than changes in the false positive error rate upon changes in the relative risk. This is illustrated by the fact that the enrichment curve is "better" when  $\alpha$  is minimum and  $\beta$  maximum than when  $\beta$  is minimum and  $\alpha$  maximum. But the relative risk curve is "better" when  $\beta$  is minimum and  $\alpha$  maximum than vice versa. This characteristic remains true even after taking into consideration the difference in range between the maximum and minimum  $\alpha$  and the maximum and minimum  $\beta$ .

The criterion being considered may be divided into two sub-

issues: Can the Ames Assay identify a subset of environments where industrial hygiene interventions will be worthwhile? and, Are Ames positive environments truly high in cancer risk?

The enrichment curves in figure 5-2 represent the maximum proportion of successes that could be achieved by instituting effective industrial hygiene controls in Ames positive but not Ames negative environments. This corresponds to the overall accuracy of the Ames Assay at various prevalence points.

1. When  $\alpha$  and  $\beta$  are minimum, the overall accuracy of the Ames Assay varies between 92% and 93% as the prevalence of truly carcinogenic exposures varies between 20 and 70%.

2. When  $\alpha$  is minimum and  $\beta$  is maximum, accuracy varies between 67% and 86% as the prevalence of carcinogenic exposures varies from 20% to 70%.

3. When  $\alpha$  is maximum and  $\beta$  minimum, accuracy varies between 81% and 87% as the prevalence of carcinogenic exposures varies from 20% to 70%.

4. When both  $\alpha$  and  $\beta$  are maximum, accuracy varies between 63% and 74% as the prevalence of carcinogenic exposures varies from 20% to 70%.

So the industrial hygienist could achieve 63 - 93% successful interventions if the underlying prevalence of carcinogenic exposures were 20% to 70%. If the hygienist were to err on the side of caution he/she would ignore the costs of controlling truly non-carcinogenic environments and concentrate only on controlling as many carcinogenic enviro-

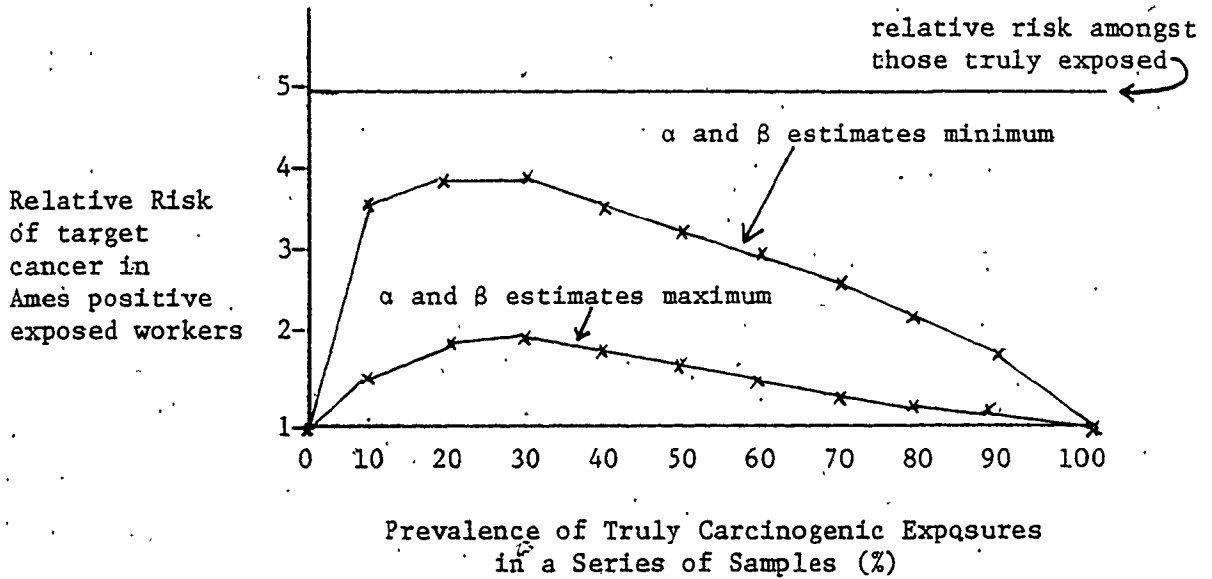
omments as possible. Considering the behaviour of the curves across the range of error rates, the hygienist would get best "protective effect" if he/she were able to use toxicological knowledge to identify a sampling situation with an a priori probability of carcinogenic exposures of 20 to 40%. If so, the hygienist could potentially achieve 40 - 80% successful interventions if there were 20% carcinogenic environments and 65 - 90% successful interventions if there were 40% carcinogenic environments.

How might the information contained in figure 5-3 assist in planning the exposure variable for a cohort study of occupational cancer mortality? The relative risks of exposure to human carcinogens are impressive at low prevalence rates of overall carcinogenic exposure. But the actual proportion of carcinogenic exposures amongst the Ames positive subset is nevertheless lower at low prevalence rates than higher ones. Will this property inhibit the epidemiologist seeking to use Ames Assay data to identify "at risk" groups?

Figure 5-4 explores this issue in relation to the following hypothetical example: The relative risk of mortality from a particular cancer is "5" amongst a group of workers who are consistently exposed to various industrial carcinogens, when compared with unexposed workers. The exposures are distributed such that some workers' environments are always Ames positive and others always Ames negative. If all the workers who were exposed to an Ames positive environment were lumped together in one exposure subgroup, what relative risk would be found for the cancer in question?

Figure 5-4 answers this question in relation to the high and low estimates of the Ames Assay error rates and at varying prevalences of carcinogenic exposures. The results clearly suggest that the usefulness of Ames Assay surveillance in identifying cancer risk depends very heavily on what the true error rates are. When  $\alpha$  and  $\beta$  are set at the minimum estimates, the relative risk amongst Ames Assay exposed workers remains above "3" throughout most of the prevalence range. When  $\alpha$  and  $\beta$  are set at the maximum estimates, the relative risk never reaches "2". These marked differences would have far-reaching impact on the power of a cohort study to detect true cancer risk and on the credibility of any association identified by such a study.

Figure 5-4: Use of Ames Assay Exposure Data to Identify "at risk" Groups



Sample Calculation:

- (a) agreement table when  $\alpha = .06$ ,  $\beta = .09$ , and prevalence of carcinogenic exposures = 10%

		Carcinogenic Exposure		
		(+)	(-)	
(+)	91	54	145	← Relative Risk = $\frac{(91 \times 5) + 54}{145} = 3.5$
(-)	9	846	855	← Relative Risk = $\frac{(9 \times 5) + 846}{855} = 1.09$
	100	900	1000	

Relative Risk = 5      Relative Risk = 1

∴ Computed Relative Risk =  $\frac{RR_{Ames+}}{RR_{Ames-}} = 3.36$

## CHAPTER VI

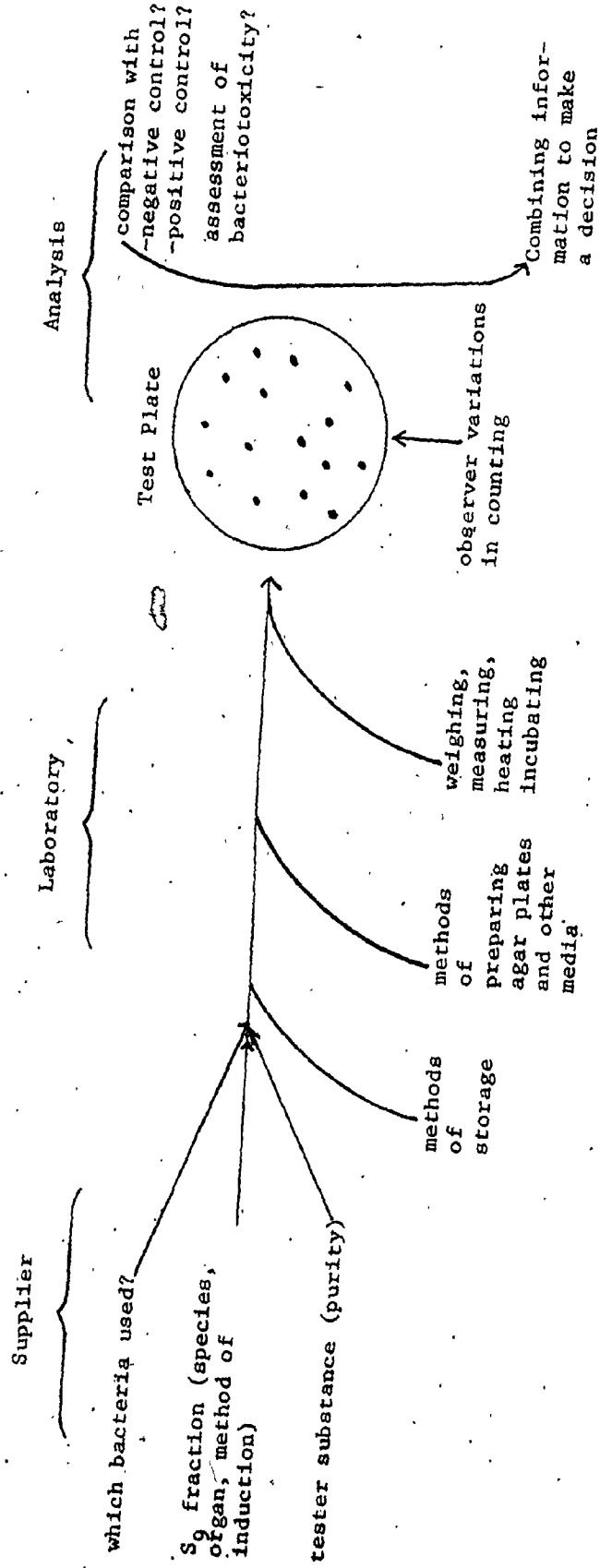
### WHAT DO COLONY COUNTS REALLY MEAN?

This chapter concerns two aspects of the Ames Assay which have been subject to only limited scientific scrutiny in the literature: reproducibility (precision) of Ames Assay counts and the use of Ames Assay dose-response relationships to predict the carcinogenic potency of test substances. The former aspect concerns the Ames Assay as both a qualitative (reproducibility) and quantitative (precision) predictor of carcinogenic risk while the latter primarily concerns its role as a quantitative predictor.

#### A. Reproducibility and Precision

Several components of Ames Assay variability which are known to affect the ultimate colony counts are illustrated in figure 6-1.<sup>49</sup> Despite much discussion in the literature concerning the impact of each variable and the need for a standardized protocol,<sup>55</sup> no adequate inter-laboratory or intra-laboratory studies of reproducibility or precision have yet been published.<sup>56</sup> (It should be noted that proposals for a standardized protocol have been made in the literature and that a multi-centre agreement study has been carried out, following this protocol. At the time of writing of this thesis, the data from that study was not yet available. However, it is clear from verbal presentation of the study plan, that the study did not deal adequately with several important methodologic issues. In particular, the method of selecting the tester substances was suspect.

Figure 6-1: Factors Affecting Colony Counts



Nonetheless, the data will soon be published in monograph form and may provide further insight into future research needs.) Instead of evaluating the bits of indirect evidence currently available in the literature, this section will discuss the sort of evidence needed to adequately assess reproducibility and precision. These issues may be separated into three questions:

1. Given standard procedures, what level of agreement can be reached within and between laboratories, concerning the assignment of pure substances as "mutagenic" or "non-mutagenic" and in determining the nature of the dose-response relationship for those found positive?

2. Do we know which protocol(s) should be proposed as standard, and assessed by large-scale agreement studies?

3. Are standard protocols adaptable to all test situations where the Ames Assay may be used? (If not, then what is the value of knowing about reproducibility or precision?)

In designing a protocol to address the first question, a number of important and difficult decisions regarding the selection of test substances, methods of laboratory analysis, and reporting of results would need to be made. Table 6-1 shows some of the objectives and methodologic problems in the design of a between-laboratory agreement study. Were a within-laboratory study contemplated, the problems concerning between-laboratory differences would be unimportant while the issue of the optimum interval between first and second test period would have to be addressed. Before such studies could be recommended, however, the challenges imposed by the latter two questions would have to be considered.



Table 6-1 lists seven laboratory variables which must be held constant in order to carry out a valid agreement study. The standard protocol requires consensus on the optimum choice for each variable. Previous evidence demonstrates that the different tester strains have different ranges of sensitivity to animal carcinogens<sup>5</sup>. Forster et.al.<sup>50</sup> demonstrated that varying the concentration of the S<sub>9</sub> fraction in relation to one test substance produced an "optimum point" where the greatest mutagenicity could be consistently found. Anderson et.al.<sup>51</sup> demonstrated that varying the concentration of the solvent (DMSO) at a single concentration of a tester substance could vary the number of revertant colonies per plate by a factor of six. Moreover, some substances showed increasing mutagenic activity with increasing solvent concentration while others showed decreasing mutagenic activity with increasing solvent concentration. Without more thorough investigation of these (and the other) variables,<sup>57</sup> a consensus regarding the optimum standardized protocol would be premature.

The third question concerns the possible trade-off between standardization and adaptability. Air samples, urine samples, and other impure materials must be extracted chemically before analysis. Each sample type poses different problems which require flexible approaches within the laboratory in order to achieve best results. The results of an agreement study based on pure substances would not be generalizable to laboratory testing linked to extraction procedures. It would be impractical to initiate large-scale agreement studies every time an investigator chose to modify a protocol to fit a new test situation. Moreover,

Table 6-1: Problems with Designing a Between-Laboratory Agreement Study

<u>Issue</u>	<u>Objectives of a Study</u>	<u>Methodologic Problems</u>	<u>Possible Solutions</u>
<p>1. Selection of Test Substances</p>	<p>To provide blinded laboratory investigators with a sample of tester substances which</p> <ul style="list-style-type: none"> <li>(1) contains substances relevant to the practical uses of the Ames Assay</li> <li>(2) minimizes the possibility of a priori diagnostic suspicion bias in the investigators because of tell-tale chemical and physical characteristics of the substances, or assumptions about the relative number of carcinogens and non-carcinogens in the sample.</li> </ul>	<p>(1) How to handle substances whose physical and chemical properties give them away?</p> <p>(2) What proportion of substances previously reported both Ames Assay positive and negative should be included?</p> <p>(3) How many substances must be tested to minimize Type II errors, without swamping the labs?</p>	<p>Include structural analogs for analysis, where possible</p> <p>Random selection of test substances would minimize bias if comprehensive list is reasonable</p> <p>Requires practical negotiating limited by budget restraints</p>
<p>2. Laboratory Analysis</p>	<p>To develop a standard protocol that can be met by all participating laboratories. Uniformity must be achieved in the following areas:</p> <ul style="list-style-type: none"> <li>(1) choice and handling of tester strains of bacteria</li> <li>(2) source, method of induction, and amount of S9 fraction used</li> <li>(3) source, handling, weighing, mixing of test substances</li> <li>(4) preparation of agar plates</li> <li>(5) choice, volume of solvent</li> <li>(6) conditions of incubation</li> <li>(7) number and type of control plates (positive and/or negative controls)</li> </ul>	<p>(1) How to handle differences in the level of experience of laboratory personnel</p> <p>(2) Under what conditions may samples be re-run? i.e. what to do about obvious blunders?</p> <p>(3) How to handle between-laboratory differences in quality of equipment and degree of control over the work-space</p>	<p>Carry out pre-testing of laboratory participants and bring together for joint indoctrination</p> <p>Rules must be spelled out beforehand. Photographs of contaminated plates, etc. must be provided</p> <p>Exclusion of poorer labs?</p> <p>Validity vs. generalizability:</p>
<p>3. Reporting of Results</p>	<p>To ensure blind colony counting, uniform decision-making regarding "positive" or "negative", and setting close-response relationships</p>	<p>(1) What about manual versus automatic colony-counting?</p> <p>(2) Will the rules of evidence influence the reported level of agreement between laboratories?</p>	<p>Requires pre-testing methods and decision as to one method based on repeatability and cost factors, etc.</p> <p>Reporting of raw data would allow various methods to be examined</p>

the quantities of test material extracted from environmental and biological sources may not be great enough to allow multi-bacteria, multi-dose testing, as per a standard protocol.

These constraints suggest three research areas which require further work:

1. The development of a standardized within-laboratory precision and reproducibility protocol that could be applied to each new procedure modification. Evidence of reproducibility and precision, as per the protocol, would have to be an important criterion in awarding grant money. The protocol would feature blind analysis and colony counting, and be adapted to the practicalities of a small laboratory situation.

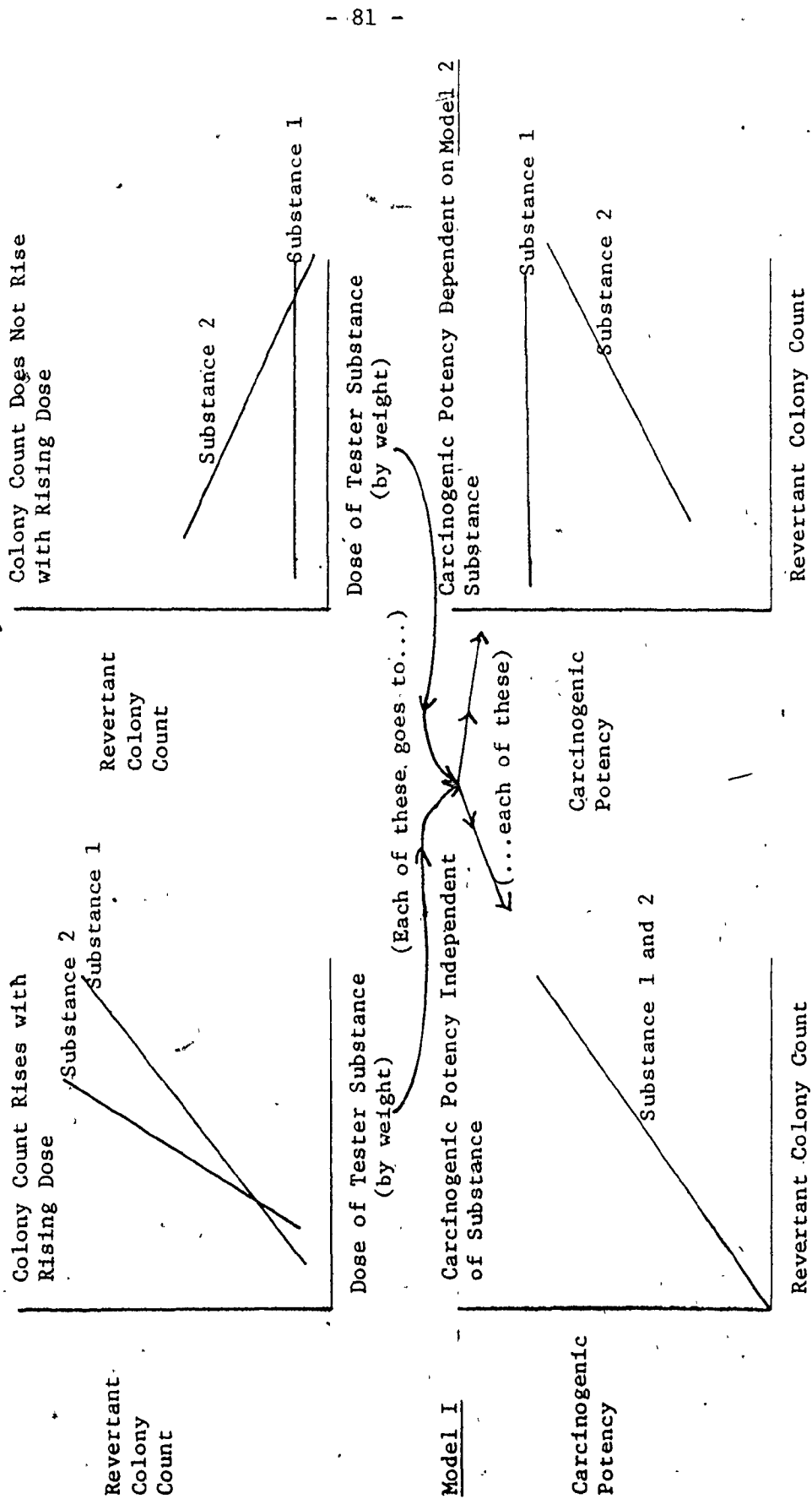
2. A large-scale, systematic initiative would need to be undertaken to isolate and describe the effect of each laboratory variable in a way that would allow a panel of experts to address the concept of an optimum standard protocol for analyzing pure substances.

3. Dependent upon research area 2, large-scale within- and between-laboratory agreement studies could begin.

B. Does Ames Assay Mutagenic Potency Predict Carcinogenic Potency?

Figure 6-2 illustrates an important issue in understanding the relationship between Ames Assay mutagenic potency and carcinogenic potency. In Model 1, the revertant colony counts serve as a common measure of carcinogenic potency, independent of the nature or dose of the test substances. Intrinsic differences in carcinogenic potency between substances affect the behaviour of the hypothetical curve relating the dose

Figure 6-2: The Relationship Between Colony Counts and Carcinogenic Potency

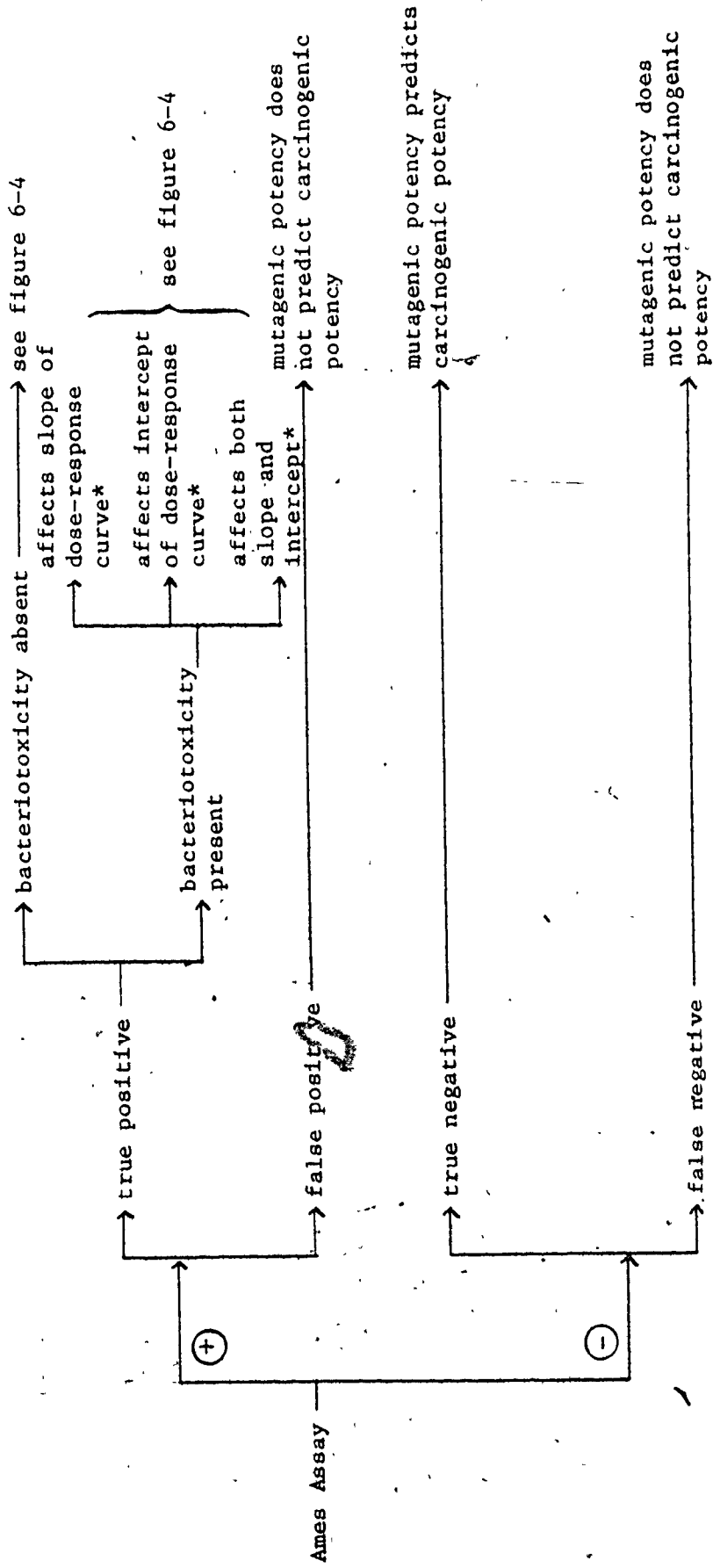


of each substance to the number of revertant colonies counted, but are unimportant when relating colony counts directly to carcinogenic potency. If this model were correct we could ignore the nature and quantity of the substances and use revertant colony counts as a valid measure of carcinogenic potency. Model 2 illustrates the opposite extreme, where carcinogenic potency is dependent on intrinsic properties of the test substance not reflected in Ames Assay revertant colony counts. If this model were correct, revertant colony counts would not be a valid measure of carcinogenic potency. Current research into the quantitative properties of the Ames Assay centres around the relationship between pure substances of known mutagenic potency and animal models of carcinogenic potency. How important is such research when the situation of environmental monitoring is considered?

Figures 6-3 and 6-4 show the results possible if Ames Assay revertant colony counts were used as a measure of carcinogenic potency in making comparisons among a series of different environments. (Problems associated with the sources of environmental material will be dealt with in the next chapter.) What factors limit the validity of this approach?

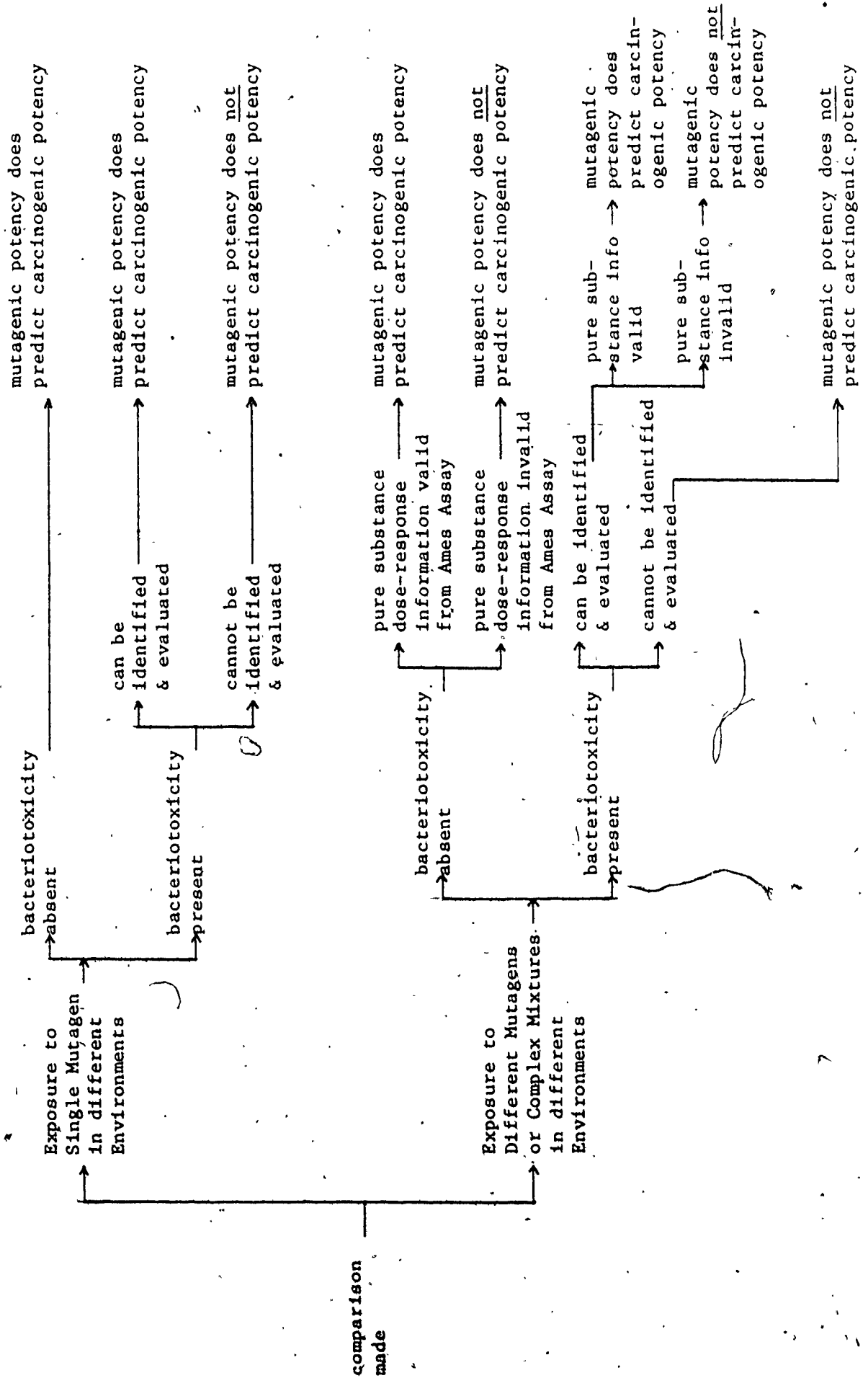
1. The sensitivity and specificity of the Ames Assay in relation to human carcinogens set a strict upper limit on the validity of the revertant colony counts as a measure of carcinogenic potency. By definition, all false positives and false negatives give incorrect dose response information while all true negatives give correct information. The uncertainty exists only among the true positives, where various fac-

Figure 6-3: Contingencies in Validating Ames Assay Mutagenic Potency on Environmental Samples



\* dose-response curve means the relationship between weight of mutagen and revertant colony count (see figure 6-2)

Figure 6-4: Contingencies in Validating Ames Assay Mutagenic Potency on Environmental Samples



tors affect the validity of the assumption that mutagenic potency predicts carcinogenic potency.

2. If bacteriotoxicity were present, it would have to be identifiable and evaluable if Ames Assay mutagenic potency were to be predictive of carcinogenic potency. If enough environmental material were available for analysis, bioassays of its bacteria-killing potency could be undertaken and their results used to adjust the revertant colony counts. If too little material were available (as may often be the case) toxicity bioassays would not be practicable. One statistical model has been developed recently which incorporates a parameter for bacteriotoxicity in the main analysis of mutagenic potency.<sup>52</sup> If valid, this would eliminate the need for separate toxicity bioassays. However, the technique awaits validation.

3. The nature of inter-environmental comparisons being made is very important. If the comparison concerns exposure to a single mutagen in different environments, the models illustrated in figure 6-2 would be irrelevant.

4. When comparisons are being made between different environments with differing mutagenic exposures, the models in figure 6-2 would be relevant. If Model 2 were true, then revertant colony counts from true positives would not be valid measures of carcinogenic potency. If Model 1 were true, then revertant colony counts from true positives would be valid measures of carcinogenic potency. However, without knowing which mutagens were false positives, this theoretical validity would not be translatable into practical validity.



These observations suggest that carrying out animal-based studies of the relationship between Ames Assay mutagenic potency and carcinogenic potency may not be very helpful in determining whether or not the Assay is a valid predictor of human carcinogenic risk.<sup>53,54</sup> Data such as that from the Dofasco study,<sup>43</sup> relating revertant colony counts at a given work place to cancer mortality at that work place, may be a more practical and direct approach to take..

## CHAPTER VII

### ISSUES IN ENVIRONMENTAL SAMPLING FOR AMES ASSAY MUTAGENS

The validity of the Ames Assay as a measure of human risk for environmentally-induced cancer depends upon the validity of the sampling procedure and the laboratory procedure. The lab procedure is a set manoeuvre and its validity depends upon its performance as defined in previous chapters. The sampling procedure is not set, so it must be optimized in relation to the intrinsic weaknesses and strengths of the possible choices.

The sampling procedure requires, in increasing order of specificity:

1. A choice of the sampling universe (ambient air or biological sample).
2. A choice of the specific material from which samples will be drawn (and therefore the sampling instrument).
3. A choice of sampling strategy.

Two practical objectives may be envisioned for the Ames Assay in the Occupational environment.

Objective 1 - Collecting data which could be used to identify job-related exposures to mutagens and monitor the effectiveness of industrial hygiene interventions in reducing or eliminating them.

Objective 2 - Collecting exposure data for the "dose" axis of a dose-response curve in a human morbidity or mortality study.

This chapter will explain my choice of air samples as superior in relation to both objectives and identify some uncertainties which may affect the validity of any sampling strategy.

A. Choice of Sampling Universe

Before applying decision-making criteria, what are the general positive and negative characteristics of air and biological samples?

Positive Characteristics of Airborne Samples

1. Sampling can be carried out without concern for problems of individual compliance among the exposed population. (If personal sampling is contemplated, compliance becomes relevant.)
2. Air samples can potentially be measures of collective exposure which can be used directly as the exposure limb of a dose-response curve.
3. Airborne samples can be collected which "represent" the substances presented initially to the body for metabolic activation. In so far as the Ames Assay mimics in vivo activation, the process will be biologically credible.
4. Airborne exposure data are comprehensive and exclusive in that they measure the mutagenic load of the environment of interest and not an aggregate of total body exposure to mutagens which may include extraneous exposures.

Negative Characteristics of Airborne Samples

1. Without personal sampling no direct accounting can be taken of personal exposure variability due to patterns of dust generation with material handling, the effects of local exhaust systems, and the dust-

trapping qualities of cotton clothing (ref.), to mention a few.

2. No allowance can be made for individual metabolic variability in the activation of pre-carcinogens and de-activation of carcinogens.

#### Positive Characteristics of Biologic Samples

1. It is possible that biologic samples are sensitive to individual exposure variables and differences in metabolic activation. This has not yet been directly examined.

2. Biologic samples may have high organ-specificity; urine samples might be highly sensitive to mutagenic exposures to the bladder while feces samples may be specially sensitive to gut exposure.

3. Previous studies, mentioned before, demonstrate the sensitivity of urine to certain environmental mutagens and suggest fecal sensitivity to others.

#### Negative Characteristics of Biologic Samples

1. Biologic samples are not specific to exposure in that they are affected by mutagens in drugs, diet, cigarettes, and other non-occupational sources.

2. The physiologic method of production of the biologic sample may make it a biased sample material (for instance, the kidneys do not excrete certain classes of mutagens).

3. Personal compliance with sampling regimes (especially for urine or feces) may be poor.

4. Collection of personal samples may be perceived as a screening test for individual cancer risk. This would lead to ethical problems con-

cerning disclosure of results and labelling certain people as more "at risk" than others, depending on Ames Assay results.

Four general criteria may be applied a priori to each sampling universe in relation to each objective.

1. Does the sampling universe comprehensively and exclusively tap the attributes which are the subject of study?

2. Does the sampling universe have any known or suspected biases in relation to the study objectives?

3. Is the universe accessible to sampling, given the practical limitations of the study protocol?

4. Does the selection of sampling universe carry with it any ethical problems?

How do these criteria apply to objective one: identifying work place mutagens and monitoring control procedures? (This discussion will deal with airborne contact, the commonest route of occupational exposure.)

The first criterion may be evaluated by working backwards from the theoretically possible results of air and biologic sampling to see which gives the best results. This is illustrated in figure 7-1 which shows 16 possible pairs of Ames Assay results before and after an industrial hygiene intervention in a work place. The combinations are presented as agreement or disagreement between biologic and air samples. For simplicity, the choices have been dichotomized. So, "A" (positive concordance of Ames Assay result between air and biologic samples before an intervention) can be paired with A, B, C, or D (the four possible concordant or discordant outcomes after intervention), and so on. Four

Figure 7-1: Possible Results of Comparing Air and Biological Samples Before and After Industrial Hygiene Intervention

		<u>Before Intervention</u>		<u>After Intervention</u>	
		Biological	Sample	Biological	Sample
Air Sample	(+)	A	A+B	(+)	A'+B'
	(-)	C	C+D	(-)	C'+D'
		A+C	B+D	A'+C'	B'+D'
		(+)	(-)	(+)	(-)
		Air Sample →			
				N	

N = number of workers sampled or number of work sites sampled, where there is a one-to-one correspondence between workers and work sites

Possible Combinations of Outcomes:

- A : AA', AB', AC', AB'
- B : BA', BB', BC', BD'
- C : CA', CB', CC', CD'
- D : DA', DB', DC', DD'

Total = 16 combinations possible

combinations, those beginning with "D", require exclusion, since they represent initial negative agreement and no intervention would be contemplated. What about the remaining 12? Combination AD<sup>1</sup> represents unequivocal success of the intervention, while AA', BA', and CA' represent unequivocal failure. These combinations are not helpful because they do not require us to choose between the results of airborne and biological monitoring. Combinations BC' and CB' involve mutagenicity switching from air to biologic sample (or vice versa) after an intervention. Both should be very unusual occurrences. That leaves 6 decision-making scenarios which have been grouped into 3 "disagreement" pairs for comparison.

Pair 1 BD' - Air mutagens are eliminated without evidence of exposure in biologic sample.

CD' - Mutagens are eliminated from biologic sample without evidence of airborne exposure.

If concern is directed to airborne exposure, the former result would be an expected success. The latter result would suggest that the mutagens had been eliminated by accident, or were unrelated to work.

Pair 2 BB' - Air mutagens present before and after intervention. No mutagenicity found in biologic sample.

CC' - Mutagens present in biologic sample before and after intervention; in the absence of air mutagenicity.

The former possibility is an unequivocal failure, while the latter is an expected failure.

Pair 3 AB' - Body fluid mutagens are eliminated following intervention while air mutagens are not.

AC' - Air mutagens are eliminated but not body fluid mutagens. In the latter case, the obvious response would be to look for non-occupational exposures. In the former case, the obvious response would be to consider the intervention a failure and try again to eliminate airborne exposure.

The disagreement pairs show that from a decision-making viewpoint, the exclusiveness of the air sample information makes it best for evaluating interventions designed to eliminate airborne exposure. Possible dietary, smoking, and other effects make the results of biologic sampling more confusing than useful for this task.

Without knowing the specific exposures (by chemical analysis of air and biologic samples), and knowing the pharmacology of the chemicals found, it is impossible to assess the potential for bias with biologic samples. So in most practical instances, there will be an unanswered suspicion of bias with biologic samples.

Meeting objective 1 will involve short-term sampling. This should help to reduce difficulty with compliance in donating biologic samples, since the energy of the investigators could go into getting good compliance on a small number of occasions.

Ethical considerations exist in relation to collecting biologic samples because they may be interpreted as having direct implications for individuals. This is not true for air samples, where the implications apply to groups, and don't allow the identification (misidentif-



ication) of "hypersusceptible" workers.

So according to the four a priori criteria, airborne samples are more suitable for identifying airborne exposures and evaluating industrial hygiene interventions than biologic samples.

Since objective two concerns measurement of human risk and may involve long-term sampling, the evaluation will be different. Under the following conditions biological samples best meet criteria one, two, and three:

1. Where concern is directed towards cancer at a particular site, which is closely associated with an easily obtainable body fluid.
2. Where the expected mutagens will be present in the fluid of choice if they have been absorbed.
3. Where general environmental data are believed, a priori, to misrepresent individual exposure.
4. Where short-term or infrequent sample requirements and good working relationships can overcome compliance problems.
5. Where true confounders (correlates of both exposure and outcome) also influence the mutagenic potency of the body fluid in a predictable way.

Under these conditions, air samples best meet the first three criteria:

1. Where inhalation of ambient air is the most important route of entry.
2. Where work habits and individual variables are believed, a-

priori, to be unimportant in influencing exposure.

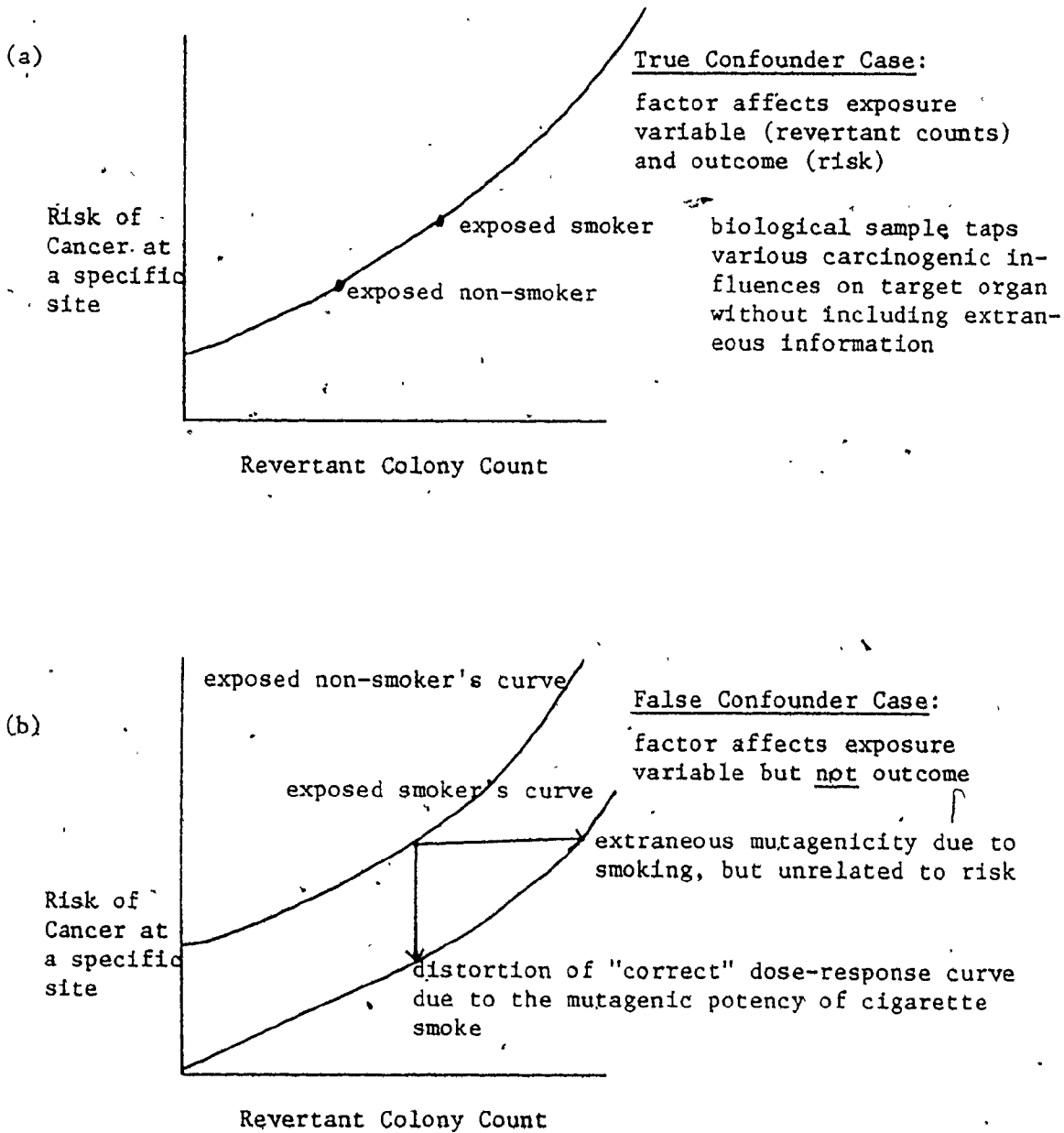
3. Where no hypothesis exists regarding the cancer site of interest or the site of interest is the respiratory tract.

4. Where frequent or long-term sampling is required.

5. Where other factors which influence the mutagenic potency of the relevant biologic sample are not also associated with the expected outcome.

The final point of each set of conditions addresses the same issue. The former point is based on the premise that Ames Assay revertant counts are a quantitative measure of carcinogenic potency in between-substance comparisons. The issue can be explained with an example. Compare an Ames Assay study of rubber workers with another group exposed to solvents. In the former case concern may be directed at bladder tumours so urine would be an appropriate fluid to analyze. Because smoking affects both urine mutagenicity and the risk of bladder cancer, counting revertant colonies in the urine samples will automatically account for the effects of smoking in a dose-response analysis of occupational exposure. In the solvent case, one might be concerned with lymphatic cancers. Perhaps analysis of blood samples could be contemplated. In this case, smoking would influence mutagenic potency, but is not thought to influence the risk of lymphatic cancer. So smoking influences would artificially inflate the colony counts and would interfere with dose-response analysis. These concepts are illustrated graphically in Figure 7-2.

Figure 7-2: Effects of True and False Confounders on the Use of Air and Biological Samples and the Ames Assay



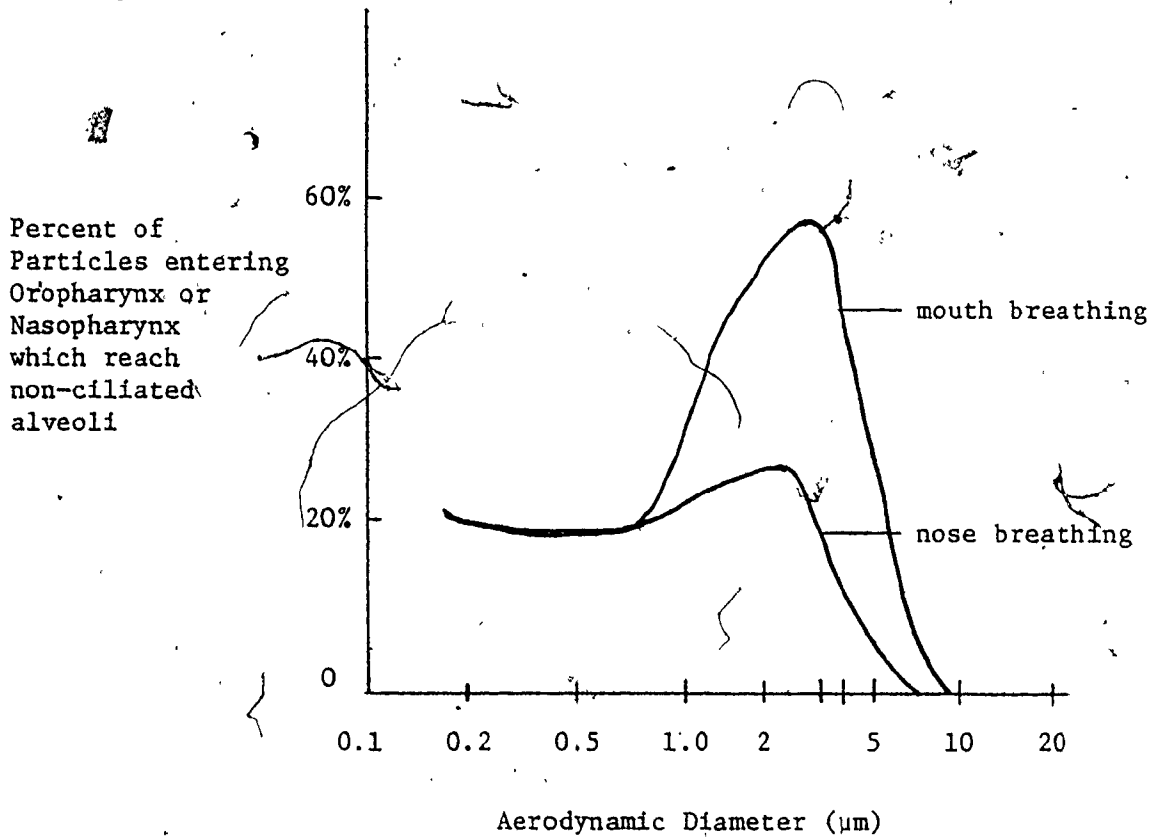
The ethical considerations are similar for Objective two as for Objective one. Because the conditions favouring airborne samples will occur more often than those favouring biologic samples, airborne sampling will tend to be more widely applicable as a source of long-term exposure data. The remainder of this chapter will deal with air samples and not biologic samples.

#### B. Choice of Specific Sampling Material

The first three criteria applied to choice of sampling universe apply as well to the choice of specific sampling material. Fulfillment of the criteria is based on the behaviour of mutagens in the air and their absorption into the human body.

It is known that many carcinogen-containing chemical families may persist for a long time in gaseous form under hot, industrial conditions. When they condense, they do so preferentially on to small particles with aerodynamic diameters of less than one, to approximately two, microns. Natusch<sup>58</sup> observed four- to fifty-fold differences in the concentration of metal carcinogens on fly ash as the effective diameter of the fly ash rose from 1 micron to 40 microns. He found a similar distribution for volatile chemical mutagens. On the exposure side, the size distribution of particles which deposit in the alveoli is known, as is the proportion of each size of particle which will deposit. This is expressed as the "alveolar dust curve" and is shown in figure 7-3. A rough cut-off size has been defined<sup>59</sup> for material which can get past the nasopharynx and oropharynx into the respiratory or gastroin-

Figure 7-3: The Alveolar Dust Curve



testinal tract. This is known as "inhalable dust". Little is known about particle size deposition in the bronchi or gut, or about the behaviour of carcinogenic gases not adsorbed to airborne particles.

Sampling technology is readily available for work with particulates, but not with gaseous samples. So the choice of studying particulates is based primarily on the nature of sampling instruments and not on any rigorous demonstration of validity. This suggests, a priori, that current sampling techniques cannot be comprehensive. Whether or not this will result in bias depends upon whether or not absorption of gaseous mutagens is a "major" source of total absorbed mutagens and whether or not the concentration and distribution of gaseous mutagens correlate positively with those absorbed to particulate matter. As the absorption of gaseous mutagens decreases and the positive correlation of gaseous and particulate mutagens increases, the opportunity for bias decreases, and vice versa.

Which particulates should be sampled? A dichotomous sampler has been developed under the auspices of the Environmental Protection Agency which simultaneously measures total inhalable dust (particles between 2.5 and 15 microns aerodynamic diameter) and "small" dust, which contains high concentrations of mutagens (that which is less than 2.5 microns aerodynamic diameter).<sup>60</sup> Two basic types of samplers exist which measure "respirable" dust, an engineer's approximation of alveolar dust.<sup>61</sup> In addition, measurement of total airborne dust is easily possible.

According to criterion one, respirable dust would be a poor choice because its size distribution characteristics are based on a different problem; measurement of alveolar deposit of fibrogenic dust. A comprehensive sample of airborne mutagens would include dust available to the bronchi, larynx, nasopharynx, and gut, rather than just the alveoli. Total airborne dust is relatively simple to collect but introduces a conservative bias since the large, non-inhalable dust will be low in mutagens. So its inclusion will decrease the relevant between-sample differences in mutagenic concentration.

The dichotomous sampler is preferable because the inhalable dust collector mode allows measurement of Ames Assay revertants per unit of biologically available dust and the "small" dust collector mode allows a direct measure of the mutagen-rich fraction. This latter is important since it should be the most sensitive indicator of between-sample differences in mutagenic concentration.

The dichotomous sampler is the most comprehensive and exclusive in relation to both Objective 1 and Objective 2. The primary drawbacks relate to uncollected gaseous mutagens and apply to all dust samplers. The ease of collection is reduced with the introduction of size selection. This is because the volume of air sampled must increase with increasing size selectivity for the total mass of collected particulate to be preserved. Increasing the sampler's flow rate may result in volatilization of mutagens at the collector site. So total dust samples may be preferable to dichotomous samples where the ambient dust level is

low. Otherwise, dichotomous sampling will maximize the sensitivity of airborne samples to between work site differences in mutagenic concentrations.

C. Choice of Sampling Strategy

The purpose of a sampling strategy is to generate data which can be combined into an estimate of exposure to mutagens. In the occupational environment, estimates are needed which allow for fair between-work site comparisons of exposure so that they can be related in the future to cancer rates among those who worked at the work sites under study. What are the most important characteristics that an investigator would wish to see represented in a summary estimator of exposure?

(1) It should account for each level of exposure; based on a knowledge of the distribution of levels of airborne mutagens with time.

(2) It should weigh different levels of exposure according to the biologically "correct" impact on cancer risk found at each level. This depends on a knowledge of the dose-response relationship between environmental mutagens and cancer risk, which may vary greatly between mutagens and is nonetheless unavailable at present.

A contingency chart (figure 7-4) has been developed in order to investigate the possibility of generating a valid summary estimate of exposure. The table illustrates the possible combinations of dose-response relationship and underlying distribution of airborne mutagens which may be found in an Occupational environment. The following sec-



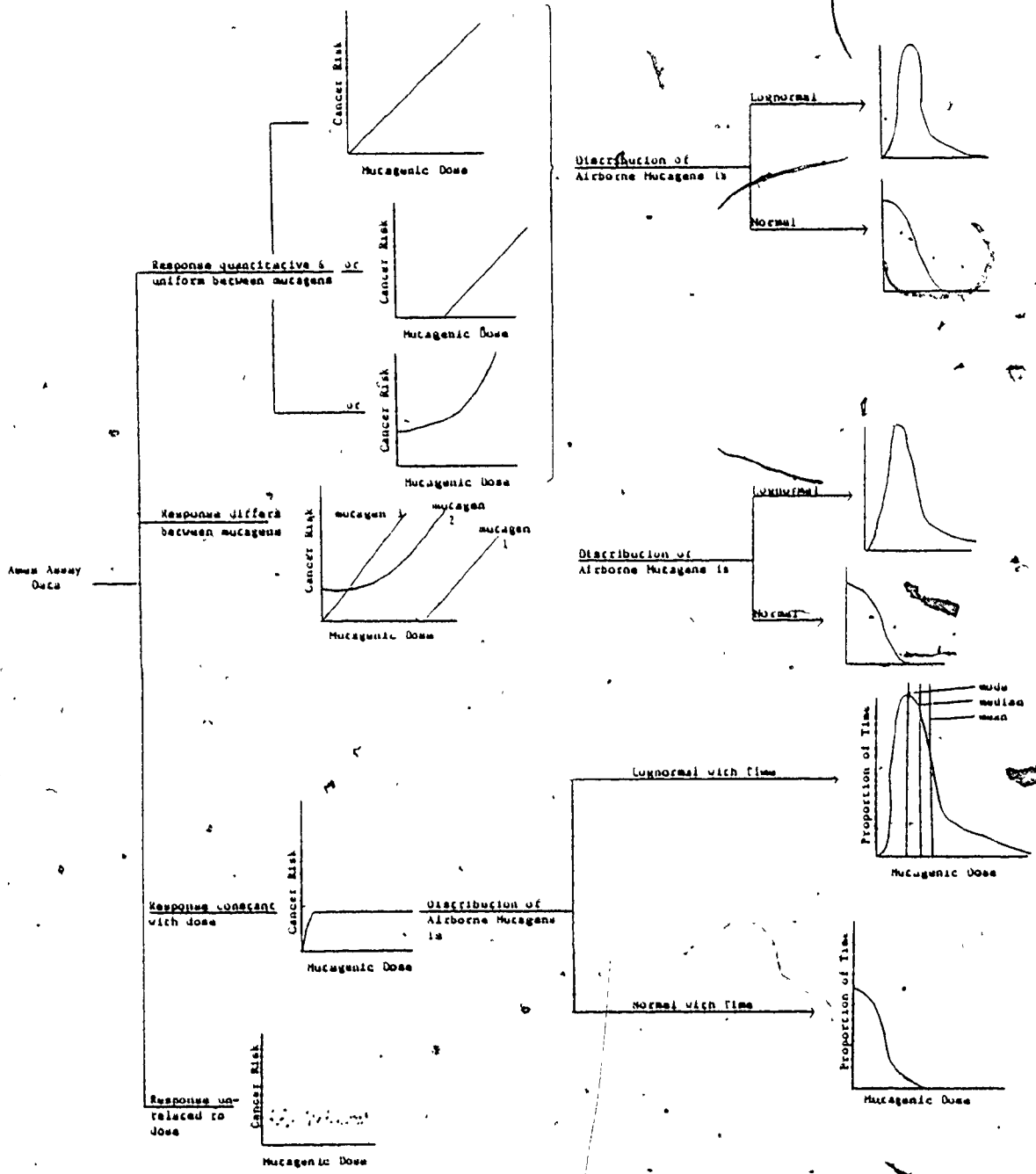
tion explains the distributional and dose-response alternatives and explores the implications for a summary estimate of exposure if each combination of distribution and dose-response relationship were true.

### I. Dose-response Models

The axis of the dose-response graphs are called "cancer risk" and "mutagenic dose". Mutagenic dose means the Ames Assay revertant colony count measured in a given airborne sample. Cancer risk means the magnitude of the contribution that a given mutagenic dose adds to the overall incidence of cancer in an exposed population. For any given sample, the contribution to risk will be vanishingly small and will actually exist as a probability of contributing to the carcinogenic process in any individual in an exposed population. So this contingency table does not concern cumulative dose. The way in which short-term exposures combine to affect observable cancer risk depends on the biologic variables outlined on the contingency table in the chapter on comprehensiveness.

Figure 7-4 has four dose-response limbs. The upper limb assumes that increasing exposure to Ames Assay mutagens means increased cancer risk, regardless of whether the comparison is made among different levels of a single mutagen or among different levels of different mutagens. The second limb assumes that different Ames Assay mutagens have different dose-response patterns. Comparisons of mutagenic potency between different substances would not result in accurate estimates of their relative contribution to cancer risk. The third limb assumes that exposure to Ames Assay mutagens means increased cancer risk; but that the

Figure 1-4. Contingency Chart of Possible Airborne Mutagen Distributions with Time and Mutagen Dose-Cancer Response Relationships



level of exposure is unrelated to the level of risk. The fourth (lowest) limb assumes that Ames Assay exposure is unrelated to risk.

## II. Models of the Distribution of Airborne Mutagens

The contingency chart gives two hypothetical distributions of air mutagens with time which encompass the plausible range: the normal and the lognormal. In general, airborne particulate levels in occupational environments meet four criteria conducive to the occurrence of the lognormal distribution:<sup>62</sup>

(1) The concentrations cover a wide range of values, often several orders of magnitude.

(2) The concentrations lie close to a physical limit (zero concentration).

(3) The variability of the measured concentration is of the order of the size of the measured concentration.

(4) There is a finite probability of very large values (or data "spikes") occurring.

Environmental variability occurs with changes in production routine, within-premises air movement, seasonal changes in air movement, and incidental occurrences, such as dust being freed from rafters by occasional heavy impacts. In an environment where there are few changes in production routine, little seasonal influence, and few incidental occurrences, the range of contaminant concentrations will narrow and the probability of data spikes will drop. Moreover, the presence of airborne mutagens is not constrained by the physical limit since it is possible that no mutagens will be present in the air at all. This is unlike the presence of general particulate matter, which is always present

throughout the atmosphere. Taking these factors into account, the range of plausible distributions of airborne mutagens with time is covered by the normal distribution overlapping zero and the lognormal distribution with a very large  $\sigma^2$ . There are two biologically relevant differences between these two distributions:

1. The presence of a long tail on the lognormal distribution; which represents the possibility of very high-dose, short-term exposures.
2. The compatibility of the normal distribution with zero exposure.

As given in figure 7-4, the distribution of airborne mutagens is based on a theoretical continuous measurement which records the instantaneous mutagen concentration in the air and combines the time intervals at each concentration to give a proportion of time at a given concentration. In practice, the shortest possible sampling interval will be determined by the minimum mass of particulate required for laboratory analysis, the maximum feasible sampling velocity (given the characteristics of the instrument), and the mean concentration of airborne particulate of the chosen size. So,

$$T = \frac{M \text{ (mg)}}{V \text{ m}^3/\text{min} \times \bar{P} \text{ mg}/\text{m}^3}$$

where T = minimum sampling time

M = minimum mass of sample required

V = maximum sampling velocity

$\bar{P}$  = mean airborne particulate concentration

As "T" increases (it may range from several minutes to several hours), the distribution of mutagens in randomly selected samples of time T will tend toward normal, regardless of whether the underlying distribution is normal or lognormal. The fact that instantaneous samples are highly autocorrelated over time means that there will be a countervailing tendency for the sampling distribution to maintain its underlying characteristics.

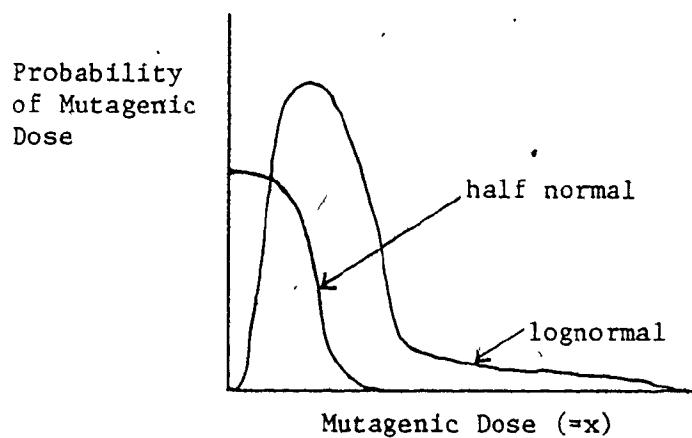
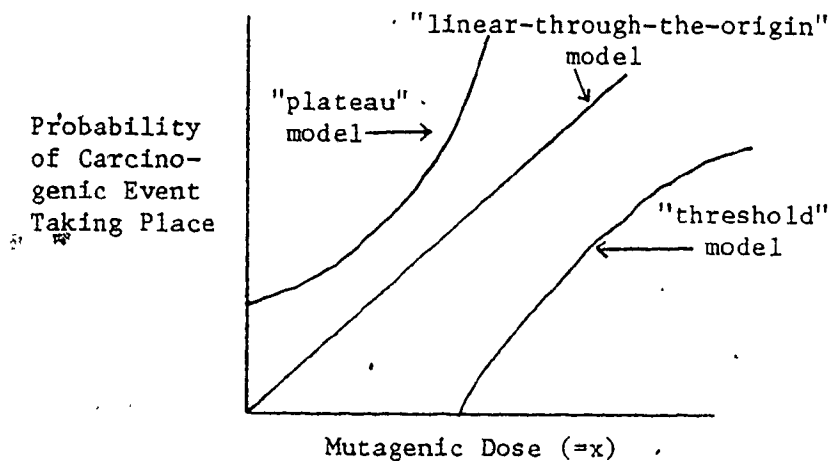
### III. Between-Work Site Comparisons of Exposure

Because of the limitations imposed by the minimum sampling period it is not possible to measure the underlying distribution of airborne mutagens. The limitations of our knowledge of carcinogenesis mean that we do not know what dose-response relationships are correct. By collecting a randomly selected series of samples of time T, we can construct sampling distributions of airborne mutagens and ascertain their means and variances. If we intended to make inferences on these means and variances; in particular, between work site comparisons of cancer risk, how could the underlying uncertainties mentioned above affect their validity?

The important uncertainties are illustrated in figure 7-5. The figure shows how the distribution of mutagens over time and the dose-response relationship between mutagenic dose and carcinogenic response may be combined to give a summary estimate of the probability of a carcinogenic event taking place (figure 7-5a). Below this is shown the index of carcinogenic risk calculable from the sampling distribution of airborne mutagens (figure 7-5b). The difference between the theoretical model of risk and the calculable index leaves room

Figure 7-5: Model of Factors Affecting Cumulative Risk

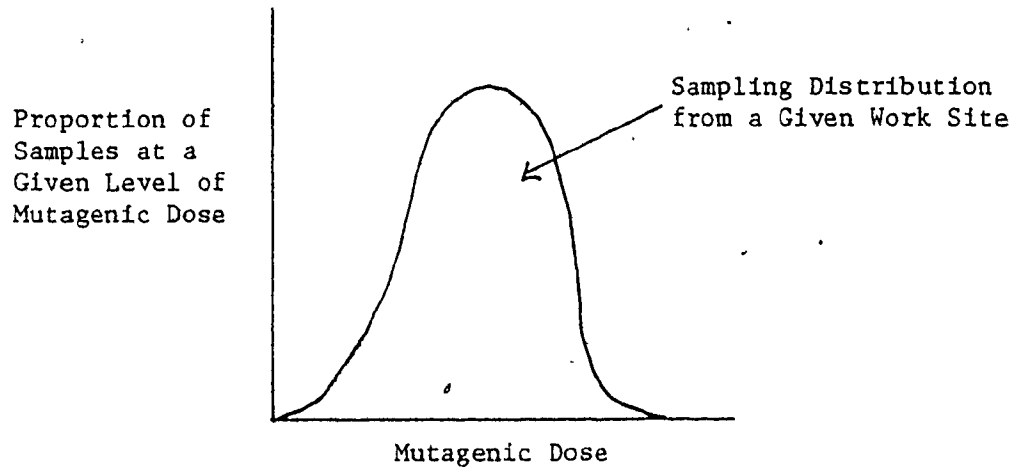
(a) Underlying Distribution of Mutagens and Dose-response Relationships



Model  $P[\text{Cancer}] = \int_x P[\text{Cancer}/x] \cdot P[x] dx$

Figure 7-5: Model of Factors Affecting Cumulative Risk (continued)

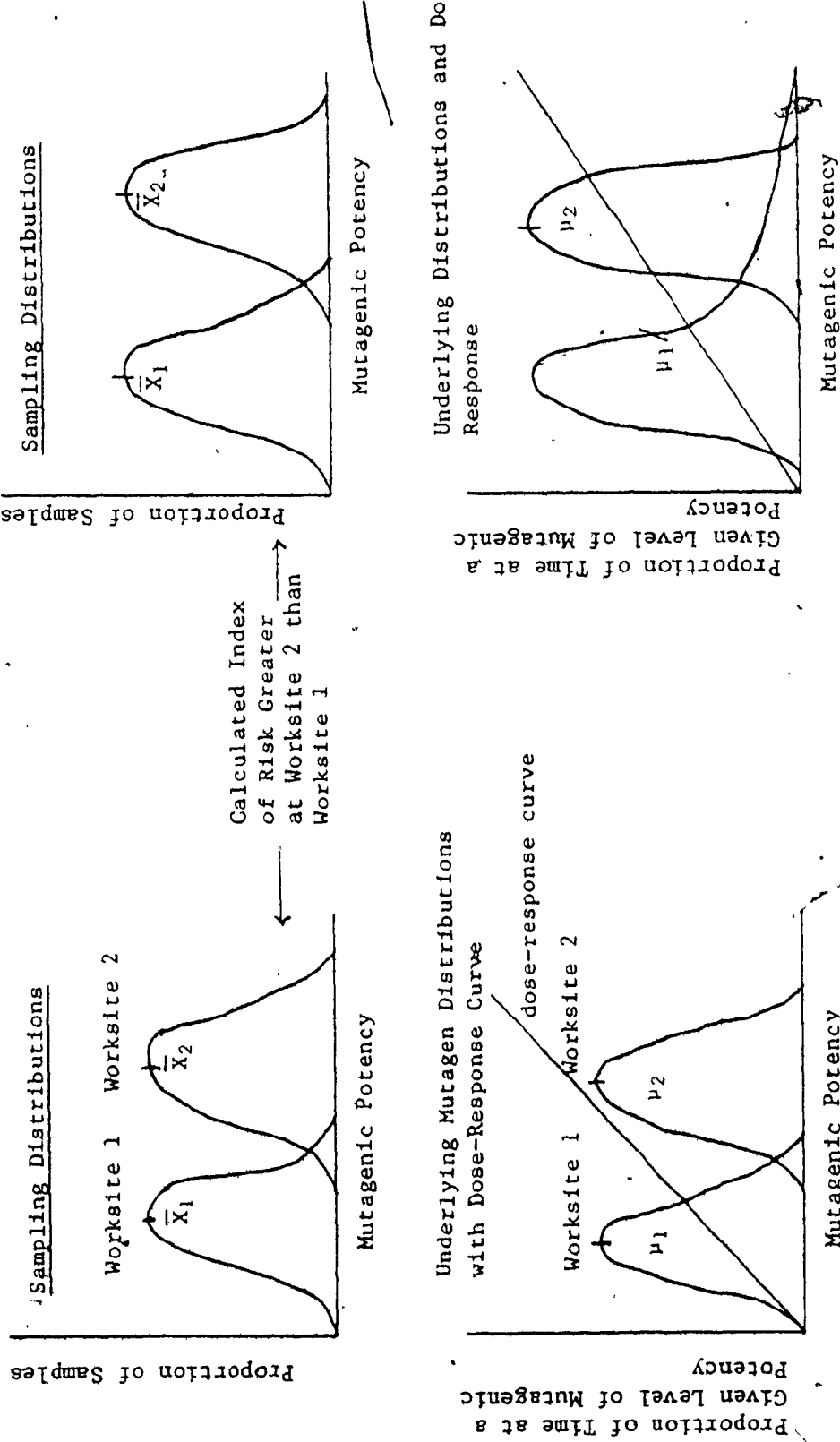
(b) Observations Made by Airborne Sampling



Calculated index of risk =  $\Sigma(\text{Mutagenic Dose}) \times (\text{Proportion of Samples at Dose})$

Figure 7-6: Risk Assessment when Dose-Response Curve is Linear and Passes Through Origin

- (a) Underlying Mutagen Distributions Normal (or both log normal)
- (b) Underlying Mutagen Distributions Different



"True" Risk at Work Site 2 > Work Site 1

80



for distortion of between-work site differences in risk. In the following examples, the nature of the underlying distributions of airborne mutagens and the dose-response relationships have been varied. The purpose of the exercise is to see whether between-work site differences in these factors could cause apparent reversal of the order of risk between work sites.

1. When the dose-response relationship is linear and passes through the origin, the between-work site comparison of risk will not distort the order of risks, regardless of the underlying distribution of airborne mutagens. This is illustrated in figure 7-6.

2. When the dose-response relationship at both work sites follows the threshold or the plateau model and the underlying mutagen distribution in both work sites is either normal or lognormal, differences in the variances of the sampling distributions will be sensitive to the "true" difference in effective exposure to airborne mutagens. This is because the biologic effect of a given dose is not directly proportional to the dose, but varies with dose. This is illustrated in figure 7-7.

3. When the dose-response relationship at both work sites follow the threshold or plateau model and the underlying mutagen distribution differs between work sites, the order of risk between work sites may be reversed.

Figure 7-8 illustrates the problem; showing that the long tail (corresponding to short-term data spikes) on the lognormal distribution will not necessarily influence the sampling distribution enough to guarantee

Figure 7-7: Risk Assessment with a Threshold Dose-Response Relationship when the Underlying Mutagen Distributions at Different Work Sites have Unequal Variance

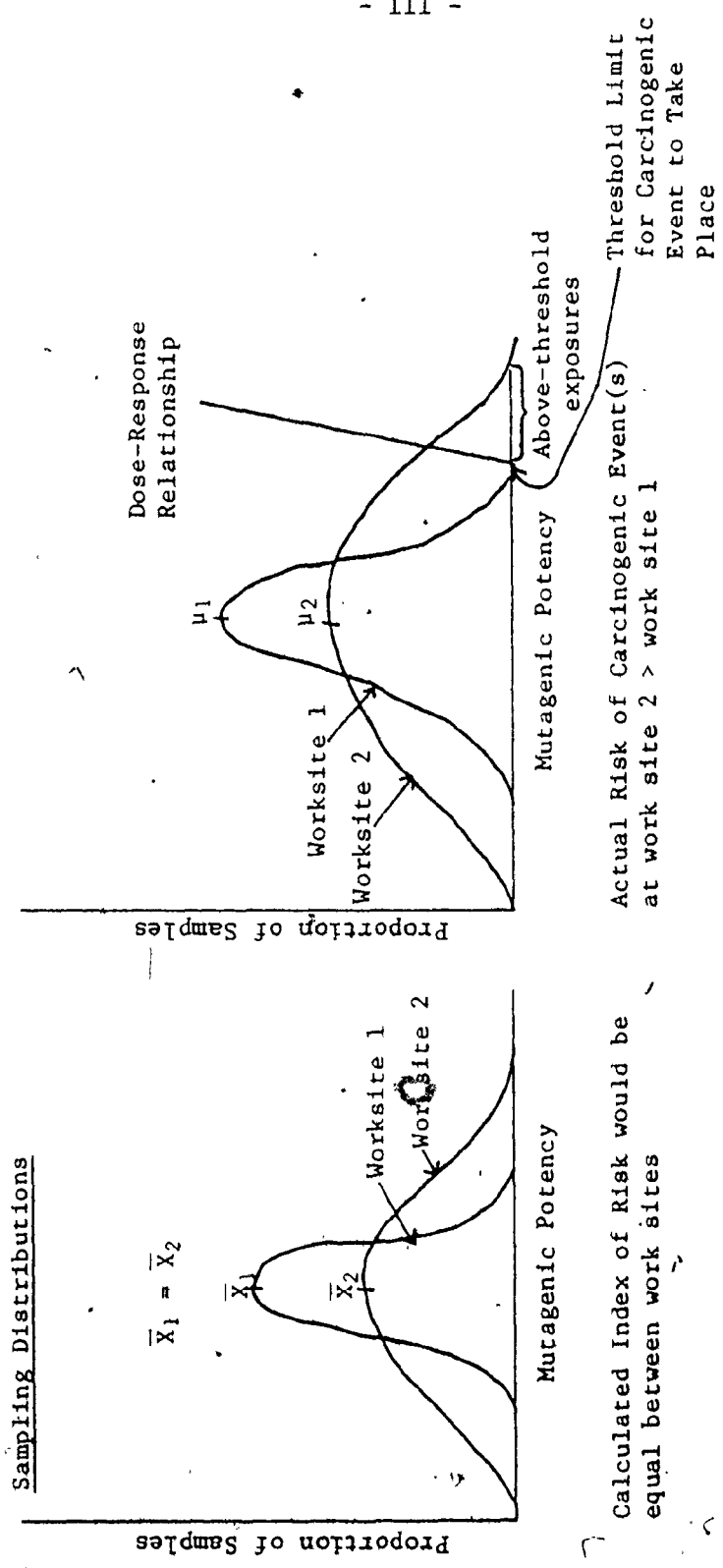


Figure 7-8: Risk Assessment and a Threshold (or Plateau) Dose-response Relationship

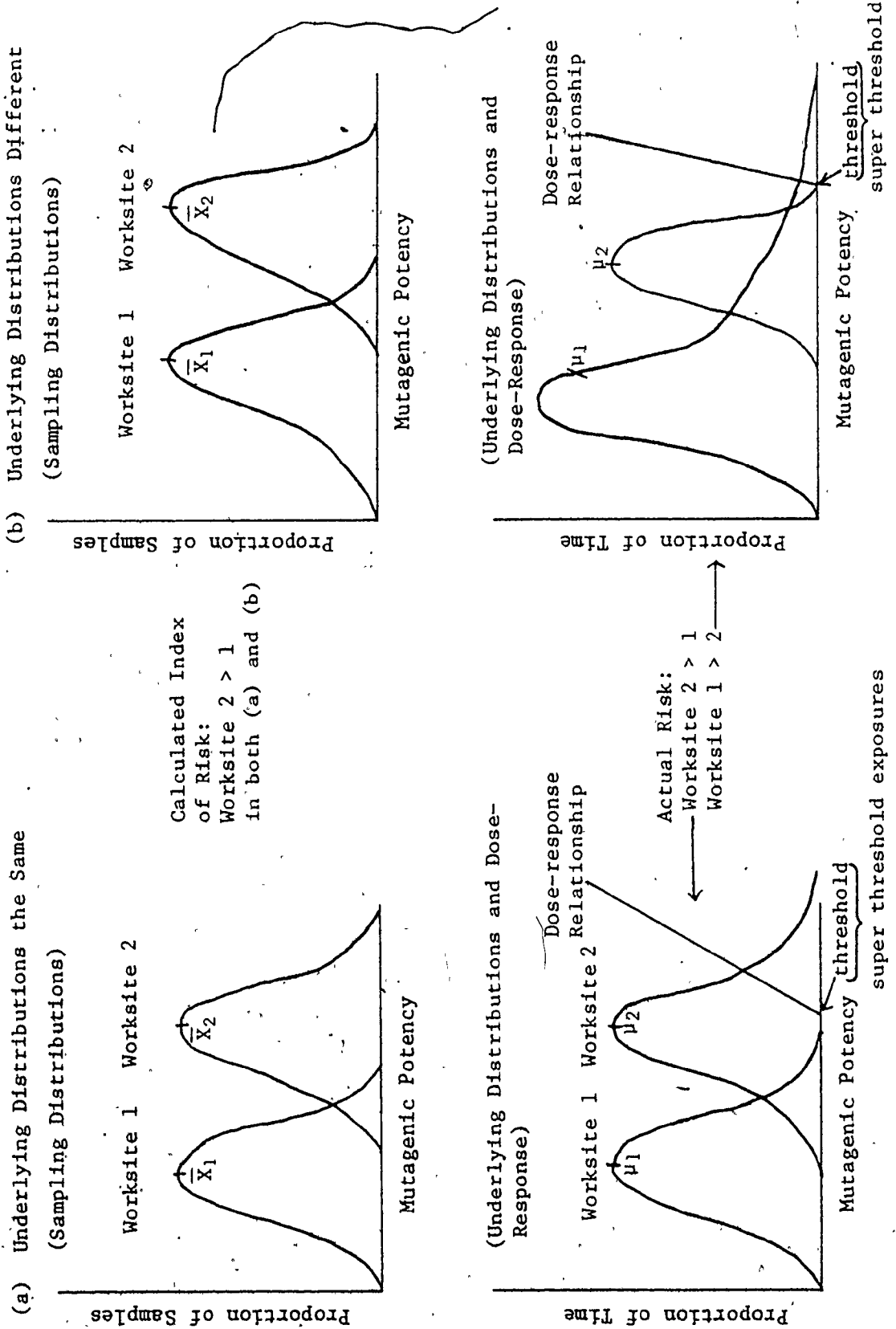
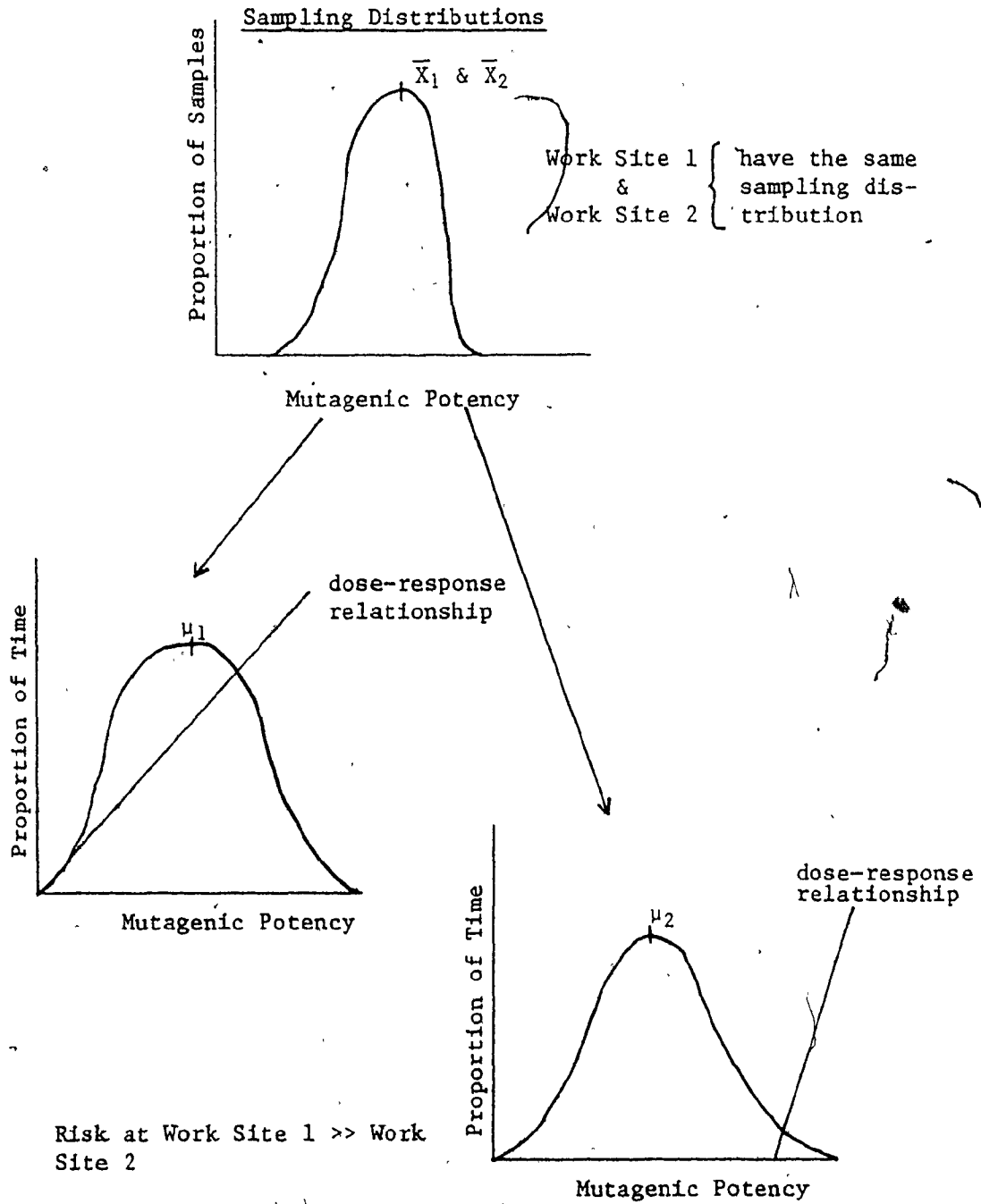


Figure 7-9: Risk Assessment with Between-Work Site Differences in the Mutagenic Dose - Carcinogenic Response Relationship



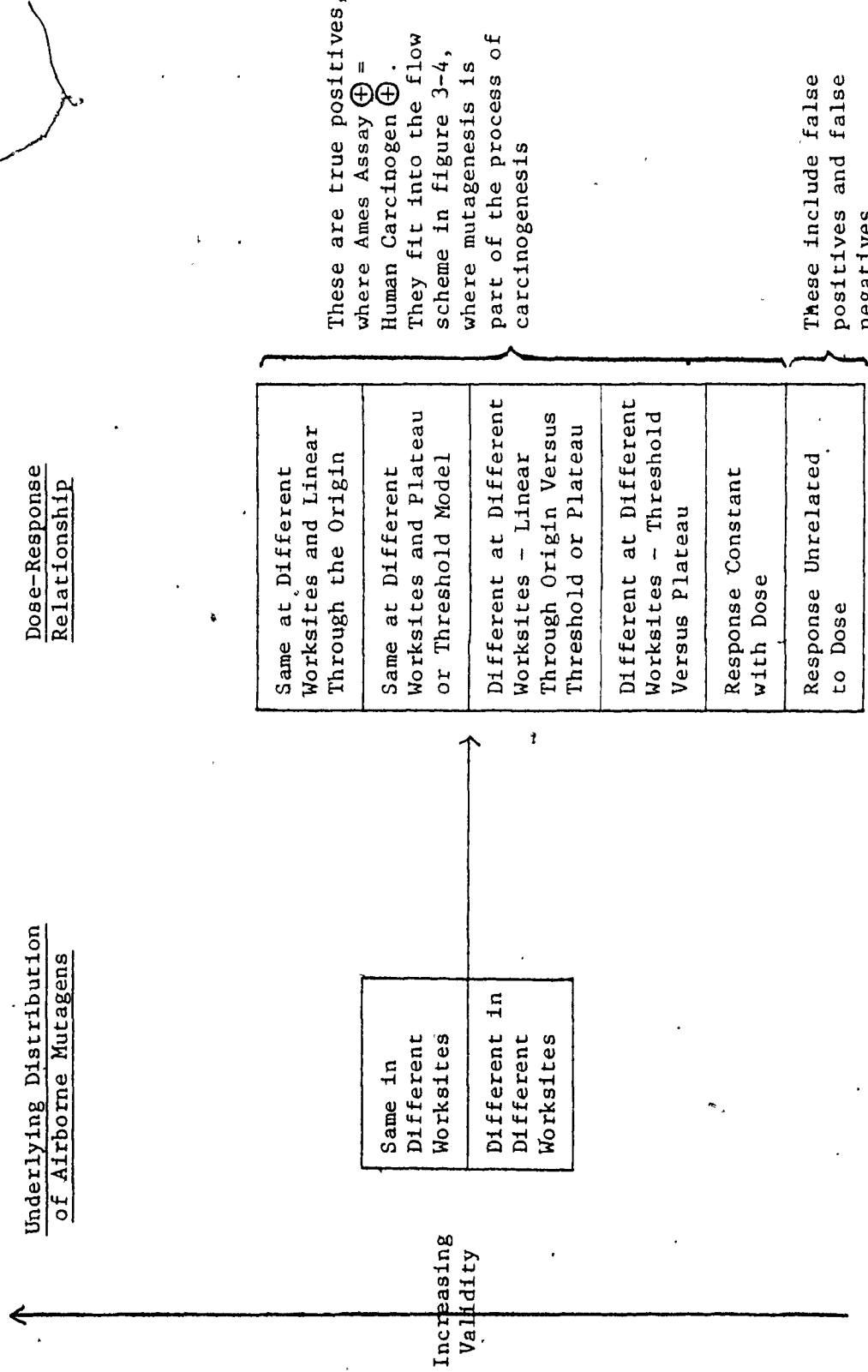
valid between-work site comparisons.

4. When the dose-response pattern differs between work sites, the biases inherent in the sampling distributions will be unequal and the comparison will tend to be invalid. This is true regardless of the underlying distribution of the airborne mutagens. Figure 7-9 illustrates this point.

5. When the dose response relationship has zero slope, the only relevant outcome is the presence or absence of airborne mutagens. If the mutagens present in different work sites have different intrinsic carcinogenic potencies unrelated to Ames Assay revertant colony counts, between-worksite comparisons of mutagenic potency would be meaningless.

These findings have been graphed in figure 7-10 to correspond with the flow scheme of comprehensive validity in figure 3-2. The overall accuracy of the Ames Assay as a predictor of human carcinogens is the link between the two flow schemes. Only the true positives will simultaneously give positive revertant colony counts and relate to one or more attributes of human carcinogenesis. Only true negatives will simultaneously give negative counts and not relate to the attributes of human carcinogenesis. Put together, these two diagrams show the most important factors which affect the possibility of the Ames Assay being able to detect observable differences in cancer risk due to airborne exposure.

Figure 7-10: Flow Scheme Relating the Limitations of Airborne Samples to the Comprehensive Validity of the In Vitro Procedure



## CHAPTER VIII

### RECOMMENDATIONS FOR USE OF THE AMES ASSAY AS AN EXPOSURE VARIABLE IN HISTORICAL PROSPECTIVE OCCUPATIONAL CANCER STUDIES

The most important priority for future research concerns the validation of the Ames Assay in relation to known human cancer risks. In terms of practical situations, this may involve true prospective data collection or collection of cross-sectional data which could be incorporated into historical prospective studies of occupational cohorts. The former approach is the most valid, but would be long, expensive, prone-to-failure, and limited in its applicability. The latter approach is attractive because of its potential for quickness and wide applicability. This chapter addresses the issues important in collecting cross-sectional Ames Assay data suitable for correlating with cancer risks identified in an historical prospective study of cancer mortality.

There are four principal design issues:

1. Defining discrete "work sites" or subenvironments within a work place which are reasonably homogeneous in terms of exposure.
2. Defining a sampling strategy which provides unbiased between-work site comparisons of exposure.
3. Standardizing a method of laboratory analysis which is practical and adaptable to small samples of dust, but which gives up a minimum of comprehensiveness.
4. Developing a method of data analysis which allows for sensible trans-

lation to employee exposure records.

1. Defining Discrete Work Sites

A panel of experts, including labour and management representatives and participating scientists would convene to define logical sub-environments. In determining them, the panel would have to consider the nature of the work being done, the area it is being done in, the way in which air currents move throughout the work place, any industrial hygiene data available, and any other consideration deemed suitable. The work sites would then be the basic units from which samples would be drawn. The reliability of work site definition would be checked by conducting Ames Assay pre-tests on a small series of grab samples taken simultaneously from various locations within each work site. A single dose Ames Assay with five replicate plates, using bacterial strain TA<sub>98</sub> with and without S<sub>9</sub>, would be the outcome measure. The revertant counts per plate would be compared using an unpaired t test or a one-way ANOVA (for more than two samples) for both the S<sub>9</sub><sup>+</sup> and S<sub>9</sub><sup>-</sup> sets of plates. Reliability would be deemed acceptable if no statistically significant differences were found between locations at the  $\alpha = 0.05$  level. If the reliability criterion were not met, the experts would have to redefine work sites and try again.

2. Defining a Sampling Strategy

The panel of experts would be required to assess the work place for factors which might influence exposure throughout the work place or at individual work sites. This would include the effects of daily work



patterns, the effect of seasons on ventilation in-plant, the periodic use of respirators, the presence or absence of repeating work cycles and any other factors deemed important. In the absence of any evidence of such patterns, an arbitrary sampling frame could be divided into consecutive time periods of sufficient length to allow for the dust requirements of the analytic method to be met. This would require a knowledge of the ambient dust levels and the amount of material necessary for analysis (discussed later). The consecutive time periods could then be randomly selected and air samples collected simultaneously at the various work sites throughout each designated time period. Simultaneous sampling is necessary to control for unidentified predictors of exposure during different time periods that may not have been considered by the panel of experts. Were such predictors present they would reduce the strength of non-simultaneously-collected samples in between-work site comparisons of exposure.

In order to achieve simultaneous sampling, total dust samples, rather than samples of respirable and/or inhalable dust would have to be collected. This is because the dichotomous samplers, which collect respirable and/or inhalable dust, are too expensive to be purchased in the large numbers required for simultaneous sampling. In order to assess the reliability of this method, the dichotomous sampler would be run simultaneously with the total dust sampler at each worksite on a rotation basis throughout the work place. Separate analysis of the material from the dichotomous sampler would be compared with the total dust samples by converting revertant counts from each to a measure of colonies per volume of air sampled. The values from the dichotomous sampler should al-

ways be slightly lower than the total dust sampler. Reliability depends upon whether or not the ratio:

$$\frac{\text{revertant colonies/volume (by dichotomous sampler)}}{\text{revertant colonies/volume (by total dust sampler)}}$$
 is approximately the

same at each worksite, but higher on a per milligram basis. If it is, no measurable bias is being introduced by using total dust samples. If it varies, a possible bias is being introduced. This must be reported as a problem in reliability, if present.

An alternate strategy may be to use personal samplers and pool the samples from all workers in a worksite. The difficulty of getting a large enough sample to analyze would be great unless many workers and many samplers were utilized. Moreover, the samples would be restricted to "respirable dust", which could introduce biases in relation to the total biologically-available dust.

If the effects of daily work patterns and seasonal changes were judged significant, stratification of time periods within the sampling frame would have to be considered. Proportional sampling by season and by "peak" or "slack" daily periods within each season would reduce the chance of significant random error if these factors had a marked effect on exposure. The usefulness of stratification may be evaluated after the fact by simply examining the data for evidence of anticipated variations in mutagen levels by the stratification factors.

If distinct work cycles were identifiable, these would define the sampling frame. The sampling frame would include at least one complete cycle, or a complete year if seasonal effects were important;

whichever were longer. When individual work site factors were considered important predictors of exposure, stratification would supercede the desirability of simultaneous sampling. So each individual work site would be blocked by within - worksite predictors and sampled proportionately within them.

### 3. Laboratory Analysis

The laboratory analysis must be simple and practical enough to allow for the collection of short term samples, since multiple-testing would require large amounts of dust and long sampling times. Pre-testing is necessary to assess reproducibility and accuracy, and to experiment with minimizing the number of plates needed per sample.

#### Protocol A - Reproducibility

A series of grab samples from the work place would be collected, separated into subsamples, coded, and sent to the laboratory for blind analysis. Analysis would be carried out using TA<sub>98</sub> and TA<sub>100</sub>, with and without S<sub>9</sub> activation, at a single dose level, using three replicates each. The data generated would be analyzed by a series of unpaired t-tests by an independent analyst with access to the codes. If statistically significant differences ( $\alpha \leq 0.05$ ) were found between the split samples in any analysis, the protocol would be repeated until no differences were found.

#### Protocol B - Accuracy

Pure mutagens would be divided into a series of measured doses and analyzed blind, using TA<sub>98</sub> and TA<sub>100</sub>, with and without S<sub>9</sub> activation.

The revertant colony counts, as counted by a second blinded technician, would have to show a linear increase with measured dose. Failure to do so would result in repeating the protocol.

Protocol C - Minimizing the Amount of Sample Needed per Analysis

Grab samples taken from each work site at peak work periods would be analyzed using TA<sub>98</sub> and TA<sub>100</sub>, with and without S<sub>9</sub> activation. If the results of analysis with one or the other strain yielded consistently higher results and no internal inconsistencies were found in the rank order of mutagenic potency by work site, then the less sensitive bacteria would be dropped from analysis. Also, if both strains were giving virtually identical results in both rank order and relative differences between work sites, TA<sub>100</sub> would be dropped because of historic difficulties found with its rate of spontaneous revertants. If either strain were dropped, periodic spot checks of agreement between strains would be inserted into the main protocol. Between-strain disagreement on the order of mutagenic potency between work sites would be noted as a problem in validity.

Protocol D - The Main Laboratory Protocol

Pending the results of protocol c, each sample would be collected to provide enough dust as follows:

a) TA <sub>98</sub> /+S <sub>9</sub>	dose 1	.5 mg x 3 replicates	=	1.5 mg
	dose 2	1.0 mg x 3 replicates	=	3.0 mg
	dose 3	1.5 mg x 3 replicates	=	4.5 mg
				<hr/>
				9 mg

LEAF 122 OMITTED IN PAGE NUMBERING

b) TA <sub>98</sub> /-S <sub>9</sub>	dose 1	.5 mg x 3 replicates	=	1.5 mg
	dose 2	1.0 mg x 3 replicates	=	3.0 mg
	dose 3	1.5 mg x 3 replicates	=	4.5 mg
				<u>9 mg</u>

and/or

c) TA <sub>100</sub> /+S <sub>9</sub>	dose 1	.5 mg x 3 replicates	=	1.5 mg
	dose 2	1.0 mg x 3 replicates	=	3.0 mg
	dose 3	1.5 mg x 3 replicates	=	4.5 mg
				<u>9 mg</u>

d) TA <sub>100</sub> /-S <sub>9</sub>	dose 1	.5 mg x 3 replicates	=	1.5 mg
	dose 2	1.0 mg x 3 replicates	=	3.0 mg
	dose 3	1.5 mg x 3 replicates	=	4.5 mg
				<u>9 mg</u>

Therefore, a range of 18 - 36 mg of dust would be needed for each analysis. For a margin of safety, 25 - 50 mg of dust would need to be collected. Sampling velocity would be held constant across all work sites. The minimum velocity possible would be determined by compromise of the need to minimize sampling time and the need to maximize the dust collected at the least dusty work site.

In practical terms, the samples could not be collected blind. But the samples, upon collection, would be coded and then analyzed blind in the laboratory. The basic laboratory methods, in terms of the source of Sq, quantity of solvent used, composition of top agar, conditions of incubation, etc. would follow Ames' 1975 standard protocol.<sup>63</sup> Any neces-

sary modifications required for the specific work place would have to be worked out in pretesting and fully documented. Modifications would be restricted to the methods of extraction of the material absorbed to the dust and would not affect the methods laid out in the Ames protocol.

The most important voluntary aspect of the protocol concerns the timing of laboratory analysis. The basic principle is that a series of samples from different work sites taken on a given day should, if possible, be put through the same analysis together. For instance, assume there were five work sites with samples collected during the same time frame and each analysis [ (a), (b), (c), (d) above ] requires 9 plates. If TA<sub>98</sub>/+S<sub>9</sub> were done one day, that would involve 45 agar plates plus controls. That is a practicable amount of work and allows for the same controls to be applied to all same-day samples within a single analysis. Negative controls (x5 replicates) are necessary for every set of plates which are analyzed and incubated together. The use of these plates will be discussed in the analysis section. Positive controls at one dose level are required for a known mutagen requiring S<sub>9</sub> activation (x5 replicates) and one not requiring S<sub>9</sub> activation (x5 replicates). These serve as quality control plates. If only S<sub>9</sub><sup>+</sup> samples are being run, only S<sub>9</sub><sup>+</sup> controls are required, and vice versa. If the positive controls come out negative or show a wide change from previously recorded revertant colony counts, the test plates analyzed at the same time are invalidated.

Missing data could come about if sampling instruments fail during sampling, too little dust is collected for analysis, laboratory mis-

takes are made, or positive controls fail. In order to avoid missing data, each randomly selected sampling interval would be backed up with a second "reserve" interval which would be sampled if errors have occurred. Constraints would have to be placed on the optional interval such that it was comparable to the original interval with regard to known predictors of exposure. Moreover, all work sites would have to be re-sampled during the optional interval in order to ensure comparability.

#### 4. Analysis and Application of Data to Worker Records

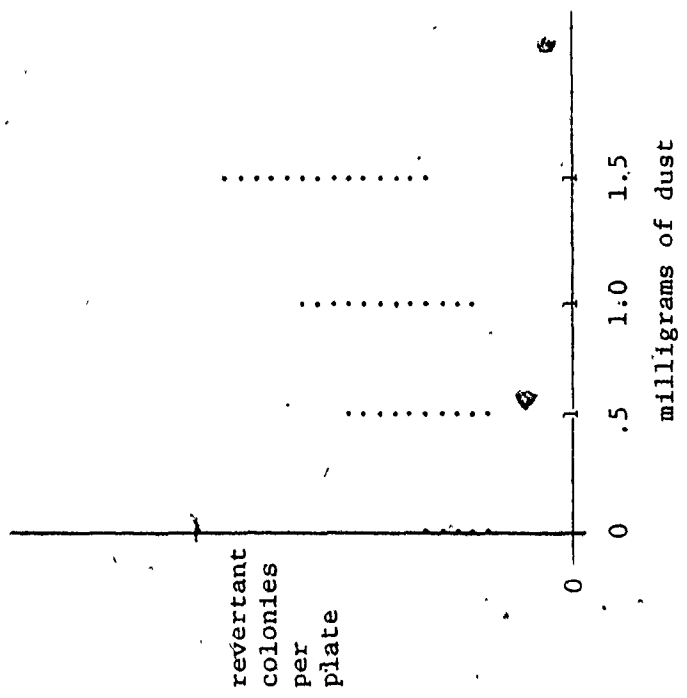
The data generated as a result of the sampling and laboratory analysis best lend themselves to multiple regression analysis as shown in figure 8-1. Each unique combination of bacterial tester strain and presence or absence of S<sub>9</sub> activation represents a unique analysis. Therefore, between 2 and 4 analyses would be generated by this protocol. The revertant colonies found on the control plates would be added into the regression analysis as the values at zero dose.

From the regression analysis would come:

- (a) a quantitative measure of mutagenic potency in revertants per milligram of dust sampled, with upper and lower 95% confidence limits around each regression line from each work site. By working backwards from the dust concentration at the individual work site while sampling was being carried out, these values can be used to generate a range of mutagenic potency in revertants per volume of air (see figure 8-2).
- (b) rank ordering of work sites by mutagenic potency per volume of air.

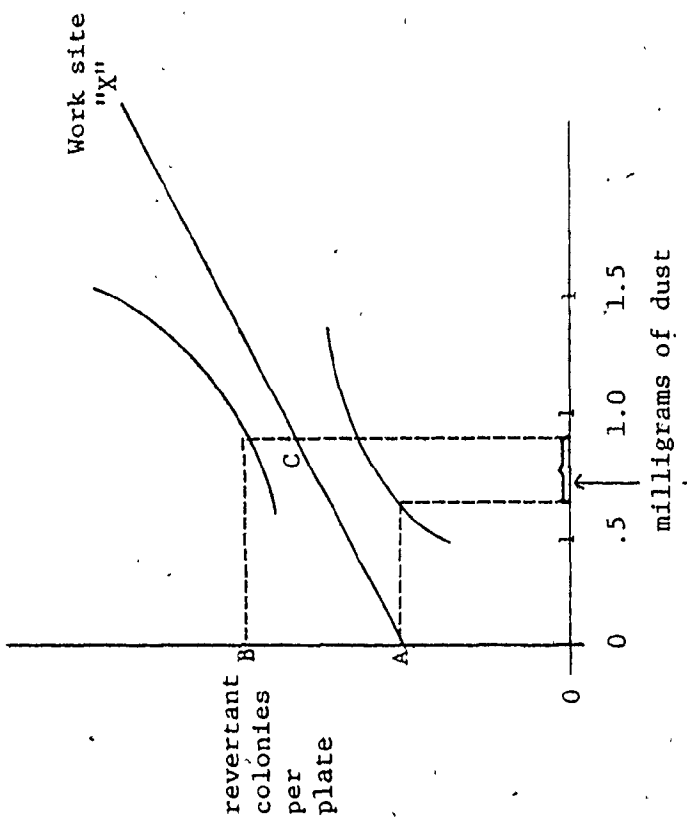


Figure 8-1: General Form of the Analysis



The Model:  $Y = \beta_0 + \beta_1 X + \beta_2 W + \beta_3 XW + E$   
 where  
 $Y$  = revertant colonies per plate  
 $X$  = volume of air sampled  
 $W$  = nominal variable denoting work site

Figure 8-2: Translating Milligrams to Standardized Volumes of Air Exposure



Amount of dust collected from standardized volume of air  $\pm$  CI  
 $\therefore$  Exposure at Work site "X"  $\approx$  C revertants per unit of air (C.I. = A  $\rightarrow$  B revertants per unit of air)

Direct comparison of negative control colony counts and work site samples incubated with them would give a dichotomous (yes/no) answer as to whether any mutagenicity was present. This would involve unpaired t-tests of revertant colony counts on the control plates and the work site plates using that dose group (either .5, 1.0, or 1.5 milligrams) with the highest mean mutagenicity. Recording of a yes/no outcome would be undertaken for each combination of bacteria and  $S_9$  for each sampling period.

One important issue in the multiple linear regression approach concerns violations of its basic assumptions, especially the assumption of homogeneity of variance. Other investigators have reported that the variance of Ames Assay revertant colony counts rises with rising mean value. Dofasco data confirms this observation. Various investigators have used square root transformations<sup>64</sup> or the Box-Cox approach<sup>65</sup> to stabilize variance. The most convenient approach to take would be to test a series of transformations between revertants<sup>-1</sup> and revertants<sup>1</sup> in an iterative manner, until near-homogeneity of variance is achieved. After the confidence limits have been calculated, they could be transformed back to revertants per unit of air sampled.

Bacteriotoxicity is a difficult potential drawback. Toxicity will be suspected when:

- (a) The revertant colony counts from a work site at any dose level fall below control levels.
- (b) The dose-response relationship from a work site on a given

day is flat or has negative slope.

Multi-dose pre-testing from all work sites would be necessary to assess the possibility of toxicity. If present, special potency assays would be necessary to estimate the level of bacteria-killing at the 3 test doses. These estimates would then be used as adjustment factors if similar patterns of toxicity showed up in the main analysis. If, as may be the more common case, toxicity is sporadic and variable, repeat sampling would be the best way around the problem. If it were persistent and variable in effect, exclusion of some data points may be required. One recent paper has developed a model with an in-built parameter for bacterio toxicity. An adaptation of this model may be suitable for persistent, variable toxicity.

Three sorts of exposure data were described earlier: quantitative work site levels of mutagenic exposure, rank ordering of work sites by exposure, and dichotomous summarization of work site exposure. In order to validate the Ames Assay, each level of data must be brought together with employee work histories to provide a summary estimate of exposure.

(a) Quantitative Data

The length of time each employee worked at each work site would be multiplied by the exposure level in revertants per volume of air to give a measure of revertant-months per volume of air exposure. The upper and lower confidence limits on each work site exposure estimate could be used to provide upper and lower estimates of exposure. Two to

four separate estimates would be available, depending upon whether both TA<sub>98</sub> and TA<sub>100</sub>, or just one bacteria were used. This would provide for several continuous exposure variables, based on all logical combinations of the estimates. Since the objective of research is to test the validity of the Ames Assay, each combination would be used as a hypothetical exposure variable, against which the observed risk of cancer mortality or morbidity at a particular site of interest could be compared.

(b) Rank Order Data

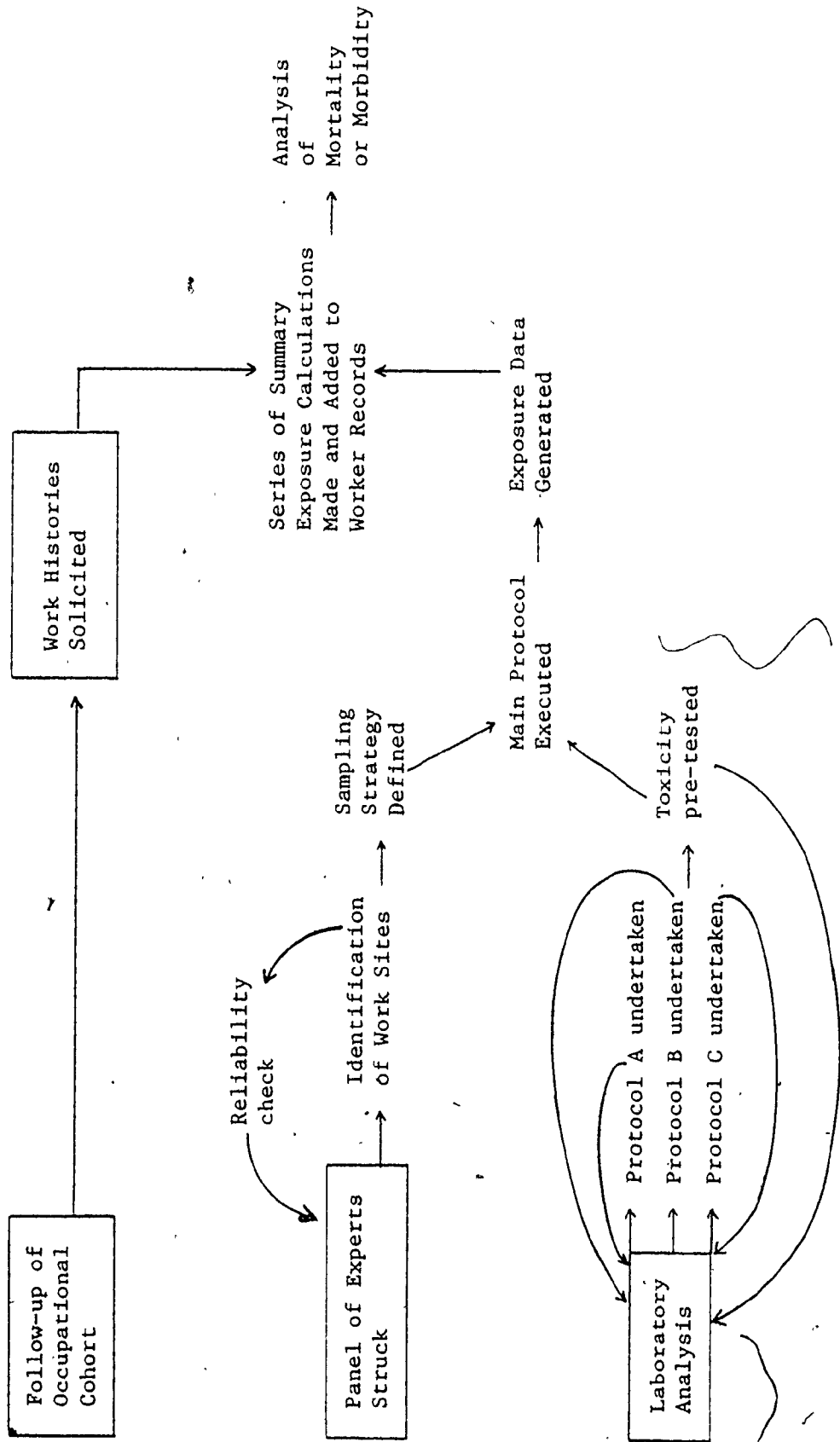
The rank order data would group work sites with similar levels of mutagen exposure. These groups could be divided into heavy, moderate, and light exposure and a worker's work history documented by these categories. These summaries could then be manipulated to see whether "length of heavy exposure" or "time since first heavy exposure" variables identified subgroups at excess risk or not.

(c) Dichotomous Data

Employee work histories are divided into time exposed to mutagens and time not exposed. The within work site data will give a proportion of samples which show any mutagenicity. The proportion would be applied to the time spent by a worker at that work site to give a "time exposed" estimate. These would be collected across the work sites where the employee worked. A gradient effect between proportion of work history exposed (given length of work history) and cancer risk would be sought.

Figure 8-3 shows how the various elements of sampling, laboratory

Figure 8-3: Flow of Events for Ames Assay Exposure Variable and Occupational Cohort Study



analysis, and generating exposure variables relate to each other and are linked to a historical prospective occupational cohort study. The limiting factors from the human data side are

- (1) the adequacy of occupational history records
- (2) the accuracy of morbidity and mortality information.

The former factor could invalidate linkage of exposure records if knowledge of where employees worked cannot be organized to correspond to the work sites as defined in the sampling protocol. Therefore, the definition of work sites must take into account existing knowledge of employee work histories. If knowledge is too mismatched with exposure and irretrievable in any other form, the validity of the Ames Assay will not be testable.

## CHAPTER IX

### CONCLUSIONS

In the preceding chapters, evidence has been presented concerning the Ames Assay's biologic credibility, comprehensiveness, and correlation with substitute outcomes for human cancer risk. The problems of treating Ames Assay revertant colony counts as quantitative measures of risk have been discussed. The effects of environmental variability on the validity of air sampling have been explored. Also, a presentation was made of the way in which plausible variations in sensitivity and specificity of the Ames Assay might affect its ability to detect carcinogenic and non-carcinogenic environments.

Table 9-1 gives the six criteria for a valid screening program for the early detection of disease<sup>66</sup>, and the analogous criteria for the Ames Assay as an environmental surveillance tool. It makes the point that the data presented in this thesis only concerns the first criterion. However, this point is pre-emptive in terms of the others, since an invalid procedure need not be contemplated as part of a valid screening program. So the priorities of this thesis and other research like it have not been misplaced.

As has been implied by the choice and content of the design component of this thesis, it is my opinion that the validation of a simple, standardized protocol against human cancer risk is the logical next step in the development of the test. This takes precedence over attempts to bring about improvements in the biological basis of the laboratory pro-

Table 9-1: Overall Progress in Validating the Ames Assay

Criteria for a Valid Screening Program	Analogous Criteria for Ames Assay in Environmental Surveillance	Issue Addressed by Thesis?
<p>1. Are screening manoeuvres able to detect disease which is likely to have an important impact upon health?</p> <p>2. Will the treatment of risk factors have a major impact upon the subsequent development of disease?</p> <p>3. What are the prospects that patients will comply with therapeutic regimens initiated as a result of screening programs?</p> <p>4. Do existing screening programs really alter the outcomes of the target disease?</p> <p>5. Are we misled by the traditional methods used in evaluating the clinical effectiveness or early detection programs?</p> <p>6. Have we considered the entire range of possible effects of screening, labelling of individuals as diseased, and long term therapy?</p>	<p>1. Can the Ames Assay identify a subset of environments where the cancer risk is truly greater than in other environments?</p> <p>2. Will Ames Assay surveillance manoeuvres lead to exposure-reducing industrial hygiene measures which actually reduce cancer risk?</p> <p>3. What are the prospects that companies will carry out product substitutions and/or effect engineering controls; and that workers will follow safety guidelines initiated as a result of Ames Assay surveillance?</p> <p>4. Has animal testing and the controls which have resulted, prevented any cancer?</p> <p>5. No. analogy: No traditional methods exist for evaluating the effectiveness of industrial hygiene interventions.</p> <p>6. Have we considered the entire range of possible effects of Ames Assay surveillance, labelling individuals as exposed, and industrial hygiene interventions?</p>	<p>Yes</p> <p>No</p> <p>No</p> <p>No</p> <p>No</p> <p>No</p> <p>No - only in relation to air versus biologic samples</p>

C



cedure or assess its precision or reproducibility. Moreover, information on the strengths and weaknesses of the Ames Assay in human studies is needed before batteries of short-term tests can be contemplated to improve the predictive accuracy of any one test.

The theoretical limitations on validity expressed in chapter seven cannot be overcome by practical sampling strategies, given current technology. Moreover, the difficulties found with documenting work histories, recording non-occupational risk factors, and ascertaining correct cause of death may impose such great limitations on the validity of occupational cancer studies to make sampling problems superfluous. Nevertheless, any strategy which would allow for reduced sampling times would allow for a more precise recording of the pattern of mutagenic exposures. I believe that the ability to identify short-term, high-dose exposures is an important capability in describing exposure to possible occupational carcinogens. So research into new technology for this purpose stands as an important priority.

Is the Ames Assay a valid predictive test of human environmental cancer risk? The Assay has the potential to identify subsets of environments more likely to be "carcinogenic" than the general environment. Whether or not this translates into genuine predictive validity awaits further documentation. This is why the implementation of a general exposure-data gathering protocol, such as that contemplated in chapter eight, is so important.

## REFERENCES

1. Devoret, Raymond: Bacterial Tests for Potential Carcinogens. *Scientific American*, 241:2 (1979), 40-49.
2. Sackett, D.L.: The Diagnosis of Causation. Mead Johnson Symposium on Perinatal and Developmental Medicine, 9 (1976), 3-9.
3. ----- Mutagenic Effects of Environmental Contaminants. New York: Academic Press, 1972.
4. Ames, B.N., Lee, F.D., Durston, W.E.: An Improved Bacterial Test System for the Detection and Classification of Mutagens and Carcinogens. *Proceedings of the National Academy of Science, U.S.A.*, 70:3 (1973), 782-786.
5. McCann, J., Spingarn, N.E., Kobori, J., Ames, B.N.: Detection of Carcinogens as Mutagens: Bacterial Tester Strains with R Factor Plasmids. *Proceedings of the National Academy of Science, U.S.A.*, 72:3 (1975), 979-983.
6. Ames, B.N., Durston, W.E., Yamaski, E., Lee, F.D.: Carcinogens are Mutagens: a Simple Test System Combining Liver Homogenates for Activation and Bacteria for Detection. *Proceedings of the National Academy of Science, U.S.A.*, 70:8 (1973), 2281-2285.
7. McCann, J., Ames, B.N.: "The Salmonella/Microsome Mutagenicity Test: Predictive Value for Animal Carcinogenicity". Origins of Human Cancer, Cold Spring Harbour: Cold Spring Harbour Laboratory, (1977), 1431-1450.
8. Collins, J.B., Pickard, L.: Self-Report Measures for Physical, Emotional and Social Health, Knowledge and Behaviour: Documenting Reliability and Validity. Final Report. Vancouver: Health Promotion Directorate, Health and Welfare Canada, 1979.
9. Ryser, H.J.P.: Chemical Carcinogenesis. *New England Journal of Medicine*, 285:13 (1971), 721-734.
10. Cairns, John: The Cancer Problem. *Scientific American*, 233:5 (1975), 64-72, 77-78.
11. Miller, E.C.: Some Current Perspectives on Chemical Carcinogenesis in Humans and Experimental Animals: Presidential Address. *Cancer Research*, 38 (1978), 1479-1496.
12. Purchase, I.F.H., Longstaff, E., Ashby, J., Styles, J.A., Anderson, D., Lefevre, P.A., Westwood, F.R.: An Evaluation of 6 Short-term Tests for Detecting Organic Chemical Carcinogens. *British Journal of Cancer*, 37 (1978), 873-883.
13. Rubin, Harry: Is Somatic Mutation the Major Mechanism of Malignant Transformation? *JNCI*, 64:5 (1980), 995-1000.

14. Armitage, P., Doll, R.: The Age Distribution of Cancer and a Multi-stage Theory of Carcinogenesis. *British Journal of Cancer*, 8:1 (1954), 1-12.
15. Armitage, P., Doll, R.: A Two-stage Theory of Carcinogenesis in Relation to the Age Distribution of Human Cancer. *British Journal of Cancer*, 9:2 (1957), 161-169.
16. Day, N.E., Brown, C.C.: Multistage Models and Primary Prevention of Cancer. *JNCI*, 64:4 (1980), 977-989.
17. Peto, R.: "Epidemiology, Multistage Models, and Short-term Mutagenicity Tests". Origins of Human Cancer. Cold Spring Harbour: Cold Spring Harbour Laboratory, (1977), 1403-1428.
18. Berenblum, I: The Cocarcinogenic Action of Croton Resin. *Cancer Research*, 1 (1941), 44-48.
19. Van Cuuren, B.L., Orris, L.: The Tumour-enhancing Principles of Croton Tiglium L. *Cancer Research*, 25 (1965), 1871-1875.
20. Bingham, E., Falk, H.L.: Environmental Carcinogens: The Modifying Effect of Cocarcinogens on the Threshold Response. *Archives of Environmental Health*, 19 (1969), 779-783.
21. Wattenberg, L.W., Lam, L.K.T., Speier, J.L., Loub, W.D., Borchert, P.: "Inhibitors of Chemical Carcinogenesis". Origins of Human Cancer. Cold Spring Harbour: Cold Spring Harbour Laboratory, (1977), 785-799.
22. Stutman, O.: "Immunological Surveillance". Origins of Human Cancer. Cold Spring Harbour: Cold Spring Harbour Laboratory, (1977), 729-750.
23. Beatson, G.T.: On Treatment of Inoperable Cases of Carcinoma of the Mamma: Suggestions for a New Method of Treatment, with Illustrative Cases. *Lancet*, 2 (1896), 104-162.
24. Nandi, S.: Hormonal Carcinogenesis: A Novel Hypothesis for the Role of Hormones. *Journal of Environmental Pathology and Toxicology*, 2 (1978), 13-20.
25. Kellermann, G., Shaw, C.R., Luyten-Kellermann, M.: Aryl Hydrocarbon Hydroxylase Inducibility and Bronchogenic Carcinoma. *New England Journal of Medicine*, 289 (1973), 934-937.
26. Cohen, G.M., Mehta, R., Meredith-Brown, M.: Large Interindividual Variations in Metabolism of Benzo (a) pyrene by Peripheral Lung Tissue from Lung Cancer Patients. *British Journal of Cancer*, 24 (1979), 129-133.

27. Painter, R.B., Howard, R.: A Comparison of the Hela CNA-Synthesis Inhibition Test and the Ames Test for Screening of Mutagenic Carcinogens. *Mutation Research*, 54 (1978), 113-115.
28. Rose, G., Barker, D.J.P.: Repeatability and Validity. *British Medical Journal*, 2 (1978), 1070-1071.
29. Julian, J., Muir, D., Roberts, R., Shannon, H.: Mortality Studies of Nickel Workers. *Chronic Diseases in Canada*, 1:3 (1980), 24.
30. Rinkus, S.J., Legator, M.S.: Chemical Characterization of 465 Known or Suspected Carcinogens and Their Correlation with Mutagenic Activity in the Salmonella Typhimurium System. *Cancer Research*, 39 (1979), 3289-3318.
31. Jenssen, D., Ramel, C.: The Micronucleus Test as Part of a Short-term Mutagenicity Test Program for the Prediction of Carcinogenicity Evaluated by 143 Agents Tested. *Mutation Research*, 75 (1980), 191-202.
32. Heddle, J.A., Bruce, W.R.: "Comparison of Tests for Mutagenicity or Carcinogenicity Using Assays for Sperm Abnormalities, Formation of Micronuclei, and Mutations in Salmonella". Origins of Human Cancer. Cold Spring Harbour: Cold Spring Harbour Laboratory, (1977), 1549-1557.
33. Fleiss, J.L.: Statistical Methods for Rates and Proportions. New York: J. Wiley and Sons, 1973.
34. Yamasaki, E., Ames, B.N.: Concentration of Mutagens from Urine by Adsorption with the Nonpolar Resin XAD-2: Cigarette Smokers have Mutagenic Urine. *Proceedings of the National Academy of Science*, 74:8 (1977), 3555-3559.
35. Falck, K., Grohn, P., Sorsa, M., Vainio, H., Heinonen, R., Holsti, L.R.: Mutagenicity in Urine of Nurses Handling Cytostatic Drugs. *Lancet*, 1 (1979), 1250 - 1251.
36. Harris, C.C.: The Carcinogenicity of Anticancer Drugs in Man: A Hazard in Man. *Cancer*, 37 (2 suppl.) 1976, 1014-1023.
37. Wall, R.L., Clausen, K.P.: Carcinoma of the Urinary Bladder in Patients Receiving Cyclophosphamide. *New England Journal of Medicine*, 293:6 (1975), 271-273.
38. Moller, M., Dybing, E.: Mutagenicity Studies with Urine Concentrates from Coke Plant Workers. *Scandinavian Journal of Work, Environment, and Health*, 6 (1980), 216-220.
39. Hakama, M., Kilpikari, I.: Cancer Risk Among Rubber Workers. *Journal of Toxicology and Environmental Health*, 6 (1980), 1211-1218.

40. Monson, R.R., Fine, L.J.: Cancer Mortality and Morbidity Among Rubber Workers. *Journal of the National Cancer Institute*, 61:4 (1978), 1047-1053.
41. Falck, K., Sorsa, M., Vainio, H.: Mutagenicity in Urine of Workers in Rubber Industry. *Mutation Research*, 79 (1980), 45-52.
42. Reddy, B.S., Sharma, C., Darby, L., Laakso, K., Wynder, E.L.: Metabolic Epidemiology of Large Bowel Cancer: Fecal Mutagens in High- and Low- Risk Population for Colon Cancer: A Preliminary Report. *Mutation Research*, 72 (1980), 511-522.
43. Gibson, E.S., Martin, R.H., Lockington, J.N.: Lung Cancer Mortality in a Steel Foundry. *Journal of Occupational Medicine*, 19:12 (1977), 807-812.
44. Tomatis, L.: "The Value of Long-term Testing for the Implementation of Primary Prevention". Origins of Human Cancer. Cold Spring Harbour: Cold Spring Harbour Laboratories, (1977), 1339-1357.
45. Poirier, L.A.: "Mutagenic Screening Tests for Carcinogenicity", Screening Tests in Chemical Carcinogenesis. IARC Scientific Publications No. 12. Lyon: International Agency for Research on Cancer (1976), 15-24.
46. McCann, J., Choi, E., Yamasaki, E., Ames, B.N.: Detection of Carcinogens as Mutagens in the Salmonella/Microsome Test: Assay of 300 Chemicals. *Proceedings of the National Academy of Science*, 72:12 (1975), 5135-5139.
47. Odashima, S.: The Cooperative Development in Japan of "Methods for Screening Chemicals for Carcinogenicity", Screening Tests in Chemical Carcinogenesis. IARC Scientific Publications No. 12. Lyon: International Agency for Research on Cancer (1976), 61-75.
48. Hui, S.L., Walter, S.D.: Estimating the Error Rates of Diagnostic Tests. *Biometrics*, 36 (1980), 167-171.
49. Ashby, J., Styles, J.A.: Factors Influencing Mutagenic Potency In Vitro. *Nature*, 274 (1978), 20-22.
50. Forster, R., Green, M.H.L., Priestly, A.: Optimal Levels of  $S_0$  Fraction in the Ames and Fluctuation Tests: Apparent Importance of Diffusion of Metabolites from Top Agar. *Carcinogenesis*, 1 (1980), 337-345.
51. Anderson, D., McGregor, D.B.: The Effect of Solvents Upon the Yield of Revertants in the Salmonella/Activation Mutagenicity Assay. *Carcinogenesis*, 1 (1980), 363-366.

52. Stead, A.G., Hasselblad, V., Creason, J.P., Claxton, L.: Modelling the Ames Test. *Mutation Research*, 85 (1981), 13-27.
53. Meselson, M., Russell, K.: "Comparisons of Carcinogenic and Mutagenic Potency". Origins of Human Cancer. Cold Spring Harbour: Cold Spring Harbour Laboratory, (1977), 1473-1481.
54. Ames, B.N.: Identifying Environmental Chemicals Causing Mutations and Cancer. *Science*, 204 (1979), 587-593.
55. deSerres, F.J., Shelby, M.D.: Recommendations on Data Production and Analysis Using the Salmonella/Microsome Mutagenicity Assay. *Mutation Research*, 64 (1979), 159-165.
56. Cheli, C., DeFrancesco, D., Petrullo, L.A., McCoy, E.A., Rosenkranz, S.: The Salmonella Mutagenicity Assay: Reproducibility. *Mutation Research*, 74 (1980), 145-150.
57. Stoltz, D.R., Stavric, B., Iverson, F., Bendall, R., Klassen, R.: Suppression of Naphthylamine Mutagenicity by Amaranth. *Mutation Research*, 60 (1979), 391-393.
58. Natusch, D.F.S.: Potentially Carcinogenic Species Emitted to the Atmosphere by Fossil-Fueled Power Plants. *Environmental Health Perspectives*, 22 (1978), 79-90.
59. National Research Council. "Effects of Inhaled Particles on Humans and Animals: Deposition, Retention, and Clearance". Airborne Particles. Baltimore: University Park Press, (1979), 107-145.
60. Anderson Samplers Incorporated: "Model 3000 Dichotomous Sampler (Virtual Impacter)". Bulletin 3000, 7-79. Atlanta: Anderson Samplers Incorporated, 1970.
61. Hamilton, R.J., Walton, W.H.: "Selective Sampling of Respirable Dust". Inhaled Particles and Vapours. London: Pergamon Press, 1961.
62. Leidel, N.A., Basch, K.A.: Statistical Methods of the Determination of Non-compliance with Occupational Health Standards. Cincinnati: U.S. Department of Health, Education, and Welfare (NIOSH 75-159), 1975.
63. Ames, B.N., McCann, J., Yamasaki, E.: Methods for Detecting Carcinogens and Mutagens with the Salmonella/Mammalian Microsome Mutagenicity Test. *Mutation Research*, 31 (1975), 347-364.

64. Weinstein, D., Lewinson, T.M.: A Statistical Treatment of the Ames Mutagenicity Assay. *Mutation Research*, 51 (1978), 433-434.
65. Irr, J.D., Snee, R.D.: "Statistical Evaluation of Mutagenicity in the CHO/HGPRT System". Mammalian Cell Mutagenesis: The Maturation of Test Systems. Cold Spring Harbour: Cold Spring Harbour Laboratory, (1979), 263-273.
66. Sackett, D.L.: Screening for Early Detection of Disease: To What Purpose. *Bulletin of the New York Academy of Medicine*, Second Series, 51:1 (1975), 39-52.