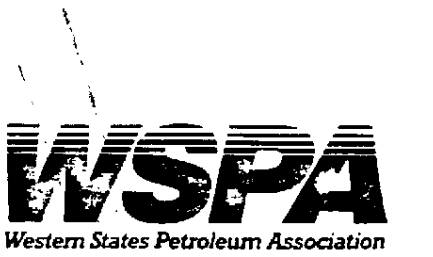

007043

December 22, 1989

RECEIVED
DEC 29

M E M O R A N D U M

To: **Proposition 65/Benzene Work Group**

From: **Tiffany Heitzenrader** TAH

Subject: **Review of New Proposal and Draft Journal Article by Tony Cox**

Please review the attached and submitted any comments to Bob Wilkenfeld (415) 231-6018 by December 31st. Both items will be discussed at the January 10th Prop 65/Benzene Meeting. (A meeting notice and agenda will be sent out shortly.)

Happy Holidays! See you in 1990...

505 No. Brand Blvd., Suite 1400 • Glendale, California 91203 • (818) 545-4105

To: Dr. Robert Wilkenfeld, Chevron Environmental Health Center
From: Dr. Louis Anthony Cox, Jr., Cox Associates
Re: Proposal to Develop Methods for Dealing with Scientific Uncertainties in Biologically-Based Cancer Risk Modeling
Date: 11/20/1989

In response to your recent request, I am pleased to submit this proposal on behalf of Cox Associates, to undertake an applied research project to develop new methods for dealing with scientific uncertainties in biologically-based risk assessment. As we discussed, the goal of this project will be to help make the conceptual advantages of biologically-based modeling more available in practiice by providing practical methods to use it in the presence of realistic data limitations and knowledge gaps. The following sections discuss the motivations for this effort in greater detail and propose a set of technical tasks, together with estimated resource requirements, for applying advanced methods from artificial intelligence and related fields to meet the needs of risk assessment practitioners attempting to use biologically-based modeling methods today.

1. Background and Motivation

Biologically-based risk analysis of chemical carcinogens can potentially change the way risk assessments are done. The premise of biologically-based risk models is that more realistic and more defensible estimates of the low-concentration human health effects of exposures to chemical carcinogens can be achieved by incorporating more biology into the calculations. This contrasts with the prevailing statistical risk assessment paradigm, which seeks to estimate mathematical relations between measures of exposure and frequencies of health responses in animals, and then extrapolate the estimated dose-response patterns to man. Where the biologically-based approach requires modeling and measuring what goes on inside an animal exposed to a chemical, the statistical approach is concerned only with the probability relation between directly observable input (exposure) and output (carcinogenic response), without requiring measurement or modeling of the detailed causal mechanisms connecting them.

Limitations of the statistical risk assessment paradigm have been comparatively well researched and discussed. The problems of extrapolating from animal responses in high-dose experimental regimens to human responses in low-dose occupational settings have

been discussed both by both industry and regulatory analysts. By providing more points of measurement along the causal chain linking exposure and response, biologically-based risk modeling offers a constructive approach to help solve these extrapolation problems and invites collection and utilization of new sorts of biological data to reduce (or at least more accurately quantify) uncertainty about risk.

The limitations of biologically-based risk modeling have been less well explored because the major techniques -- physiologically-based pharmacokinetic (PB-PK) modeling, cell kninetics and cytotoxicity modeling, and stochastic models of cancer induction and progression -- have only recently been developed to the point where practical applications to risk analysis problems seem potentially useful. The most important limitations arise from the requirements for accurate estimates or measurements of internal parameters (e.g., chemical-specific blood:air and tissue:blood partition coefficients; metabolic reaction rates; cytotoxic effects on specific cell subpopulations; cell birth, death, and differentiation rates as a function of metabolite concentration levels; and so forth.) While these data elements are logically necessary to model carcinogenic processes, they can be difficult and expensive to obtain. In some cases, such as when the metabolite(s) responsible for carcinogenesis are unknown, it may be years before the relevant science base advances far enough to provide the answers needed for biological modeling. In other cases, it may turn out that a chemical is inducing cancer by a completely unanticipated mechanism (as was the case with butadiene-induced activation of a murine leukemia retrovirus a few years ago.) If the mechanism is unknown, detailed biological modeling may be impossible.

For biologically-based risk modeling to succeed, it must include practical, sophisticated approaches to dealing with knowledge gaps and incomplete information. Otherwise, it will either be inapplicable to the many cases where little is known about the causal mechanisms of carcinogenic action, or else will be forced to rely on *ad hoc* assumptions (such as those pointed to by critics of the recent Thorslund *et al* report on benzene) that may weaken both the credibility of the approach and its predictive power. Rigorous ways of dealing with scientific uncertainties while making effective use of partial knowledge and data must be built into the foundations of a viable risk modeling methodology.

We believe that biologically-based risk modeling is the way of the future. The current statistical risk assessment paradigm (e.g., the linearized multistage model) contains intrinsic flaws that tend to produce inaccurate, generally overestimated, risk estimates.

Some of these stem from purely statistical considerations (e.g., inadequate representation of measurement errors for exposure estimates, response parameter heterogeneities, biases from censored survival times, omitted covariates, and structural changes in model form as a function of dose, in most risk statistical risk analyses.) Others result from omitted biological realities (e.g., the nonlinear dynamic responses of internal dose levels to transient external occupational exposures during an eight-hour period; or the random evolution of individual hazard functions in response to target cell population dynamics. The latter implies that no static dose-response function can adequately describe the probable time to response from a given exposure.) For us, therefore, the question is not whether a biologically more realistic alternative to the statistical risk assessment paradigm needs to be developed, but rather how soon it can be accomplished and made practical. Figuring out how to deal with scientific uncertainties is a crucial piece of this process.

2. Proposal

We propose a one-year, 78k, research effort to address the joint problems of (1) How to draw sound (and correctly qualified) interim conclusions about health risks from incomplete biological information using the techniques of biologically-based cancer risk assessment; and (2) How to use the dependencies of uncertainties about conclusions on uncertainties about facts to design empirical research programs to reduce uncertainties. These correspond roughly to the areas of inference and sequential design in statistics. However, a major goal of this effort will be to provide methods that are of practical value when statistical approaches are not. Specifically, we will emphasize the use of *partial* causal models and non-quantitative knowledge (typical of the real state of knowledge for chemicals such as benzene and butadiene) to constrain biologically-based quantitative risk estimates. Two types of deliverables are proposed, as follows:

(i) A set of *technical reports* on methods for dealing with uncertainty in biologically-based risk analysis, with applications to model-based guidance of empirical research strategies. [By "coping with" or "dealing with" uncertainties in risk analysis, we mean drawing sound and useful qualified risk conclusions from existing information, despite its imperfections, while generating insights into how to strengthen these interim conclusions (in order to support more confident or more accurate risk estimates) by performing empirical research.]

(ii) Sample *computer programs and worked-out examples* that implement and illustrate the proposed approaches for two chemicals, benzene and butadiene, for which there is both

considerable scientific information but also large remaining knowledge gaps about biological mechanisms. These examples will demonstrate how the ideas developed in the technical reports can be applied to these two chemicals of practical interest.

The intended contents of these deliverables are described more fully in our technical approach section. The proposed choice of benzene and butadiene as case studies is made so that this project can benefit from complementary projects on biologically-based risk modeling for these chemicals being proposed to the American Petroleum Institute and the Chemical Manufacturer's Association, respectively. We have also proposed to Carol Hauth of Arco, for possible funding by the AIHC, related theoretical work on improving the underlying logical and mathematical theories of uncertainty analysis. The effort proposed here stresses development of methods for practical application to risk analysis in the near term (one to four years) building on existing technologies in various disciplines. The AIHC effort, if funded, will focus on improving the long-term foundations on which future applied methods will be built.

3. Technical Approach and Estimated Resource Requirements

Our technical approach consists of five tasks, as follows. Estimated resource requirements of professional time, dollars, and calendar time are shown in parentheses for each task. They total 78k for 780 hours spread throughout a 12-month period.

Task 1: Prepare a taxonomy of data elements, detailed steps, and uncertainties encountered in biologically-based risk modeling. (80 hours, 8k, two months.)

Task 2: Review and explain the most relevant advanced techniques for dealing with the uncertainties identified in Task 1 (210 hours, 21k, four months.)

Task 3: Develop and illustrate methods for deriving sound and useful conclusions about human cancer risks from chemical carcinogens, using biologically-based risk modeling, in the presence of scientific uncertainties. (280 hours, 28k, six months.)

Task 4: Develop and illustrate methods for deriving efficient research strategies to reduce uncertainties about risk by collecting additional empirical data. (130 hours, 13k, three months.)

Task 5: Prepare technical reports and worked-out examples (including any necessary computer programs) illustrating the uncertainty-management techniques in Tasks 1-4. (80 hours, 8k, spread throughout the other tasks to cover report preparation, revision, and production expenses.)

We plan to deliver a series of four technical reports, one for each technical task, together with disks containing all computer programs developed to support the examples. Reports will delivered approximately ten weeks apart for the first ten months of this twelve-month project. During the final two months, the reports will be revised and finalized for submission as papers to peer-reviewed journals.

The orientation of this proposed research is to identify, develop, demonstrate, and promote the application of new ideas and methods to practical uncertainty-management questions in biologically-based cancer risk modeling. Therefore, we are primarily concerned with developing each technique far enough so that it can reasonably be evaluated. Development of software fully implementing the uncertainty-management

methods developed, perhaps as part of an integrated biologically-based risk modeling package, would be premature and will not be undertaken as part of this project. However, we will emphasize methods, examples, and demonstration programs that suggest what high-quality software can do for uncertainty management in cancer risk modeling. This will contribute to a longer-term goal: to advance biologically-based modeling and uncertainty-management techniques far enough by the end of 1990 to justify undertaking in 1991 the construction of an integrated software package to support biologically-based risk analysis in practical (e.g., regulatory) applications.

The motivation for and intended technical content of each of these five tasks is described in greater detail in the following paragraphs.

Task 1: Prepare a taxonomy of data elements and uncertainties encountered in biologically-based risk modeling.

Different kinds of uncertainty must be handled by different methods. Some of these are illustrated in Table 1. The purposes of Task 1 are as follows:

(i) Provide a step-by-step "how-to-do-it" review and exposition of biologically-based cancer risk modeling methodology.

This will cover the major stages of (a) PB-PK calculation of internal doses from external exposures; (b) Cell kinetics modeling of cytotoxic effects and the proliferation of stem cells; (c) Stochastic modeling of cancer induction via dose-dependent genotoxic transformations and/or changes in cell proliferation rates; (d) Stochastic modeling of the latency period and of tumor progression.

(ii) Identify the data elements (e.g., biological measurements) and knowledge ideally required for each step, along with useful proxy variables and default assumptions for each; and

(iii) Catalogue the uncertainties that arise at each point when relevant knowledge or data elements are missing, partial, unreliable, and/or inconsistent.

These uncertainties will be illustrated by concrete examples using benzene and butadiene. The rest of the report will then develop conceptual and empirical strategies and technical

Table 1: Examples of Some Types of Uncertainties Encountered in Risk Analysis and Techniques for Dealing with Them

o *Uncertainty due to noise or measurement error* in the experimental observations used to estimate metabolic rate parameters, equilibrium partition coefficients, and diffusion rate coefficients can be quantified *statistically* if measurement error characteristics are known.

o *Uncertainty due to missing data* for these numerical parameters can often be treated by a combination of two ideas from *numerical analysis:* sensitivity analysis (which addresses the question of what joint values of the unknown parameters are consistent with observed dynamic behavior, e.g., with the time courses of blood concentration levels observed in an inhalation experiment); and iterative "relaxation" methods (which use an initial guess at the unknown parameter values to make predictions of observable behaviors, and then use the difference between the predicted and observed behaviors to improve the initial guess. Under a wide range of conditions, iteratative application of this basic technique will greatly reduce uncertainty about the possible joint values of the unknown parameters. Detailed algorithms have been developed in the computer science literature on constraint programming languages and in the biostatistics literature on nonparametric maximum-likelihood estimation, NPMLE.)

o *Uncertainty due to inconclusive tests or assays* can be quantified statistically in those rare cases where their reliabilities and precisions (or more general error characteristics) are stable and known. For most assays used in risk analysis (e.g., SCE, micronucleated erythrocytes, changes in mitotic index, formation of hemoglobin adducts, and so on, in the case of known or suspected leukemogens) well-defined error probabilities are unknown. However, practical conclusions may still be reached using *"evidence-propagation"* methods to combine (possibly conflicting) evidence from different sources and tests. For example, finding that a chemical is not a mutagen in the Ames *Salmonella* test does not prove that it is not a genotoxic carcinogen, but it certainly provides probabilistic evidence that can help to support this conclusion. Evidence-propagation methods (e.g., Pearl, 1988) seek to use such inconclusive but relevant evidence to calculate numerical degrees of support for conclusions.

o *Uncertainty due to ignorance of causal structure or mechanisms* is often best approached using *assumption-based reasoning* techniques from artificial intelligence (Etherington, 1988; Smets *et al,* 1988; Pearl, 1988.) For example, speculations about whether benzene metabolites induce leukemia solely by cytotoxic stimulation of previously damaged stem cells, by exerting genotoxic effects themselves, or by as-yet unanticipated mechanisms (e.g., phenol inhibition of gap junction formation in bone marrow stroma or direct chemical reaction with cell surface rexeptors) can be treated as altenative hypotheses to be used in generating a set of alternative risk estimates. Automated identification of maximal internally consistent (but pairwise inconsistent) subsets of hypotheses can be a valuable aid both in theory formation for biologically-based risk analysis and in designing empirical experiments to discriminate among alternative assumption sets.

o Uncertainty due to ignorance of functional forms for causal relations between numerical variables (e.g., between concentration time series of metabolites and consequent death rates of susceptible cells) can often be bounded or rendered innocuous by *qualitative reasoning* techniques for analysis of dynamic systems (Karp, 1989, cf p. 63.)

tactics (from applied mathematics, computer science, and statistics) for coping with these uncertainties.

Task 2: Review and explain the most relevant advanced techniques for dealing with the uncertainties identified in Task 1.

Reasoning usefully with uncertain evidence, facts, and theories is, for better or worse, an intrinsically difficult problem. Many techniques have been developed to address various aspects of this challenge, ranging from relatively well-known applied probability and statistical methods (including statistical decision theory) to less understood (and often less defensible) methods from expert systems, artificial intelligence, computer science, applied mathematics, and numerical analysis. The purpose of Task 2 is to present a clear summary and synthesis of the main technical ideas from these fields as they apply to the specific problems and uncertainties arising in biologically-based risk analysis. This will provide non-specialists with the technical background to understand and evaluate the new techniques and strategies developed in this project. The presentation will be tightly focused on techniques that we judge to be potentially useful for the types of uncertainties identified in Task 1. It will be written for a relatively non-technical audience that nonetheless needs to understand how to deal with scientific uncertainties in quantitative risk analysis.

Task 3: Develop and illustrate methods for deriving sound and useful conclusions about the human cancer risks from chemical carcinogens, using biologically-based risk modeling, in the presence of scientific uncertainties.

This task tackles the first major question motivating this proposed research: how to draw the best possible conclusions about human cancer risks from realistic (relevant but typically very incomplete) observations, statistical evidence, and causal mechanisms using the conceptual framework of biologically-based risk modeling in conjunction with appropriate uncertainty-management techniques from Task 2. Simple computer programs will be created to illustrate the new techniques developed in this task, and their practical utility will be tested and evaluated by application to the two case-study chemicals, benzene and butadiene.

Task 4: Develop and illustrate methods for deriving efficient research strategies to reduce uncertainties about risk by collecting additional empirical data.

This task will create methods for using partial knowledge of a carcinogen's mechanisms of action to guide the acquisition of new empirical data, taking into account the costs, durations, and reliabilities of different tests and experiments and their relative expected contributions to reducing uncertainty about risk. We will use operations research techniques to study the problem of minimizing the resources (time and cost) required to achieve a given level of certainty about risk by achieving both greater understanding of causal mechanisms and greater confidence and precision in statistical estimates. The purely statistical issues (e.g., sample size and test selection) are well addressed in the literature on "expected value of information" in statistical decision theory. The contribution of this task will be to create analogous methods for quantifying the expected value of better knowledge of specific causal mechanisms in reducing uncertainty about health risks. The intent is to provide an analytic resource-management framework for addressing questions of research strategy such as "Which of the following applied research options would be most valuable for reducing uncertainty about the human health risks from occupational exposure to butadiene: (i) identification and isolation of the specific target cell populations affected; (ii) determination of whether the diepoxide metabolite, as well as the monoepoxide, is carcinogenic in B6C3F1 mice *in vivo*; (iii) more accurate measurement of detoxification rates in monkeys as a function of exposure concentration and duration; or (iv) more accurate modeling of the dynamic response of the human hematopoietic system to mitogens?" We do *not* propose to address the problem of evaluating competing basic research investigations that potentially provide benefits cutting across current and future applied research projects. Instead, we will focus on research strategies for efficiently reducing uncertainties about the human cancer risks from a specific chemical carcinogen. Finally, we will address the problem of selecting an applied research portfolio when different research options carry different costs and durations as well as uncertain outcomes. Throughout, current research options for benzene and butadiene will be used to provide realistic examples of risk research decision problems.

Task 4 is motivated by a belief that one of the greatest potential benefits from biologically-based risk modeling is the information that it provides about how the resolution of current scientific uncertainties can affect quantitative risk estimates. Enthusiasts of biologically-based risk assessment often mention uncertainty reduction for

risk estimates as a potential benefit. The purpose of this task is to help make that potential real by providing explicit methods for using biological knowledge (represented, e.g., by partial models) to guide applied risk-research.

Task 5: Prepare technical reports and worked-out examples (including any necessary computer programs) illustrating the uncertainty-management techniques discussed and developed in Tasks 1-4.

The deliverables from this task will document the research and will provide examples (and, where necessary, simple computer programs) to help illustrate and transfer the technology being developed here. As described above, each technical task will result in a technical report; these will be revised, finalized, and submitted for publication at the end of the project.

4. Project Administration

All work outlined here will be performed on a best-effort basis by Cox Associates in a timely and cost-effective fashion. Dr. Cox will lead all technical work on behalf of Cox Associates and will be responsible for the technical content and quality of all deliverables. Dr. Paolo F. Ricci of the UCLA School of Public Health will act as designated alternate project manager should the need arise.

Selected References

The following sources provide excellent technical background on some unconventional and non-statistical techniques of uncertainty analysis that are potentially useful for applied cancer risk analysis.

Etherington, D.W., *Reasoning with Incomplete Information.* Morgan Kaufmann, San Mateo, CA, 1988. (Requires some familiarity with contemporary mathematical logic.)

Karp, P.D., *Hypothesis Formation and Qualitative Reasoning in Molecular Biology.* Ph.D. thesis, Stanford University Department of Computer Science. Report No. STAN-CS-89-1263, Stanford University, June, 1989.

Martins, J.P. and S.C. Shapiro, "A model for belief revision," *Artificial Intelligence*, **35**, 25-79, 1988.

Pearl, J., *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann, San Mateo, California, 1988.

Smets, P. *et al* (eds), *Non-Standard Logics for Automated Reasoning,* Academic Press, New York, 1988. (Requires some background in mathematical logic and probability.)