Epidemiology and Procedural Protections for Workplace Health in the Aftermath of the *Benzene* Case*

Carl F. Cranor†

In the Benzene *case, the Supreme Court held that OSHA can regulate workplace carcinogens only if quantified data show a significant risk of harm at current exposure levels. In the author's view, the* Benzene *decision undermines the remedial purpose of the OSH Act, leaving workers at risk when scientific testing, due to inherent limits of cost and design, overlooks toxic effects. The author urges allowing OSHA to regulate workplace toxins on the basis of policy considerations when quantitative risk assessment is impossible.*

I

INTRODUCTION

In 1970, a mounting toll of workplace illnesses, injuries and deaths moved Congress to enact the Occupational Safety and Health Act. With its passage, Congress took long-needed action to diminish workplace exposure to carcinogens and other toxins. In keeping with the Act's remedial spirit, Congress authorized the Occupational Safety and Health Administration to promulgate rules setting ceilings on toxin exposures.

In the *Benzene* case,[1] the Supreme Court enunciated the evidentiary requirements for promulgating such occupational health standards. A plurality of the Court[2] held that the Occupational Safety and Health Administration,[3] before promulgating occupational health stan-

* Earlier drafts were read by Professor Alvin Klevorick, Yale Law School, and Helaine Pleet, M.D., M.P.H., formerly of the Centers for Disease Control and the California Department of Public Health. I profited immensely from both; however, I alone am responsible for any defects that remain.

† Associate Professor of Philosophy, University of California (Riverside). M.S.L. 1981, Yale Law School; Ph.D. 1971, Philosophy, University of California (Los Angeles); B.A. 1966, Mathematics, University of Colorado.

1. Industrial Union, AFL-CIO v. American Petroleum Inst., 448 U.S. 607 (1980) [hereinafter referred to as the Benzene case or American Petroleum Institute].

2. Justice Stevens wrote the plurality decision, joined by Chief Justice Burger and Justices Powell and Stewart. Justices Powell and Rehnquist filed separate concurrences.

3. Hereinafter referred to as OSHA or the agency.

dards, must show that workers already face a "significant health risk" at existing levels of exposure. Dicta in the case suggest that OSHA must demonstrate that significant health benefits likely will accrue at a reduced exposure level. Moreover, it must prove these conditions by a preponderance of the evidence.[4]

While there is good reason to believe that the Court wrongly interpreted the Occupational Safety and Health Act,[5] the primary focus of this article is elsewhere. I argue that the decision and its presuppositions have undesirable consequences for OSHA's attempts to promulgate environmental exposure standards under conditions of scientific uncertainty.

The Court and even some commentators seem to assume that OSHA *can* provide the relevant scientific information to demonstrate a "significant risk" of material health impairment when a risk in fact exists. This assumption is mistaken. Statistical limitations, and the conditions and costs of epidemiological research, make it virtually impossible to obtain scientific results in a wide range of cases. Thus, a researcher can fail to detect a "significant health risk" even though a threat to health exists. The *Benzene* decision misapprehends the nature of the scientific tools at hand. Scientists and technicians using animal and epidemiological studies—the foundations for setting exposure standards—will face nearly insuperable difficulties in satisfying the evidentiary requirements of the *Benzene* decision.

When the inherent limitations of scientific testing prevent the concrete establishment of a potentially serious risk to health, policy considerations such as the health of our workforce should play a more active role in guiding the promulgation of environmental exposure standards. The Court, however, has attached talismanic significance to raw scientific data and to the quantification of scientific results as the foundation for occupational health standards. Allowing policy considerations to fill the interstices of scientific knowledge would introduce flexibility into OSHA rule promulgation, and would be a justifiable expression of our concern for the well-being and safety of our workforce.

II

THE *BENZENE* CASE STANDARD

The Occupational Safety and Health Act authorizes the Secretary of Labor (the Secretary) to promulgate "occupational safety and health

4. 448 U.S. at 642-44, 653.

5. 29 U.S.C. §§ 651-678 (1976) [hereinafter cited as the OSH Act or the Act]. See the dissent's arguments, 448 U.S. at 688-724, and Note, *Cost-Benefit Analysis for Standards Regulating Toxic Substances Under the Occupational Safety and Health Act: American Petroleum Institute v. OSHA*, 60 B.U.L. REV. 115, 123-41 (1980).

standards."[6] Section 3(8) of the Act defines such standards as those which require "conditions, . . . practices, means, methods, operations or processes, *reasonably necessary or appropriate* to provide safe or healthful employment and places of employment."[7] Section 6(b)(5) directs the Secretary, when promulgating occupational health standards for "toxic materials or harmful physical agents," to

> set the standard which most adequately assures, to the extent feasible, on the basis of the best available evidence, that no employee will suffer material impairment of health or functional capacity even if such employee has regular exposure to the hazard dealt with by such standard for the period of his working life.[8]

The same section mandates:

> Development of standards . . . shall be based upon research, demonstrations, experiments, and such other information as may be appropriate. In addition to the attainment of the highest degree of health and safety protection for the employee, other considerations shall be the latest available scientific data in the field, the feasibility of the standards, and experience gained under this and other health and safety laws.[9]

Since the early twentieth century, benzene has been recognized as a toxic substance.[10] Exposure to ambient concentrations of 20,000 parts benzene per one million parts air (20,000 ppm) will result in death within minutes. Brief exposure to milder airborne concentrations (250-500 ppm) causes nausea, breathlessness, vertigo and other unpleasant reactions. Extended or chronic exposure to even lower concentrations (25-40 ppm) has been shown to cause non-malignant but potentially fatal blood abnormalities such as aplastic anemia, leukopenia, and dysfunctional blood marrow, as well as chromosomal aberrations.[11]

In 1969, the American National Standards Institute adopted a benzene limit of 10 ppm, which OSHA adopted without rulemaking as a national consensus standard in 1971.[12] Although a number of studies had suggested a link between benzene exposure and leukemia, it was not until the 1970's that "several additional studies reported a statistically significant increased risk of leukemia among workers occupationally exposed to high levels of benzene and concluded [that] benzene

---

6. 29 U.S.C. § 655 (1976). The Secretary of Labor has delegated this authority to the Assistant Secretary of Labor for Occupational Safety and Health, the chief executive officer of OSHA. The Secretary and OSHA thus are referred to interchangeably in this article.

7. 29 U.S.C. § 652(8) (1976) (emphasis added).

8. 29 U.S.C. § 655(b)(5) (1976).

9. *Id.*

10. American Petroleum Inst. v. Occupational Safety & Health Admin., 581 F.2d 493, 498 (5th Cir. 1978), *aff'd*, 448 U.S. 607 (1980).

11. *Id.*

12. *Id. See* 29 U.S.C. § 655(a) (1976), granting the Secretary this authority. The standard is codified at 29 C.F.R. § 1910.1000 Table Z-2 (1977).

was a leukemogen . . .,"[13] that is, a substance tending to induce leukemia.

In light of this new evidence, OSHA proposed to lower the permanent benzene standard, even though it could not determine a safe exposure level. The Secretary argued that section 6(b)(5) compels the lowest exposure limit possible when safe levels of a proven carcinogen are unknown.[14] This approach parallels the agency's generic cancer policy:

> [W]henever there is a quantum of proof—either from animal experiments, or, less frequently, from epidemiological studies—that a substance causes cancer at any exposure level, an emergency temporary standard [will] be promulgated immediately, requiring employers to provide monitoring and medical examinations and to reduce exposures to the lowest feasible level.[15]

Public hearings were held during the summer of 1977. The following February, OSHA promulgated a permanent benzene standard which reduced the maximum exposure limit from the national consensus standard of 10 ppm to 1 ppm.[16]

Various benzene producers sought pre-enforcement review of the new standard.[17] In *American Petroleum Institute v. Occupational Safety and Health Administration*,[18] the Fifth Circuit invalidated the standard on the ground that OSHA had exceeded its delegated rulemaking authority. The court held that OSHA failed to demonstrate that the 1 ppm exposure limit was "reasonably necessary or appropriate to provide safe and healthful employment," as required by section 3(8).[19] The Fifth Circuit concluded that section 6(b)(5) did not give OSHA "unbridled discretion" to adopt standards designed to create a risk-free workplace regardless of cost.[20]

On certiorari, the Supreme Court affirmed the judgment of the Fifth Circuit.[21] In a plurality opinion written by Justice Stevens,[22] the Court ruled that the 1 ppm exposure limit standard was not supported by appropriate findings. The new benzene standard did not rest on a *finding* that exposure to 10 ppm of benzene *would* cause leukemia while

---

13. 581 F.2d at 498 (footnote omitted). *See* studies cited *id.* at 498 n.13.

14. 448 U.S. at 613, 624-25, 637.

15. *Id.* at 645 n.51.

16. 581 F.2d at 498, 499. OSHA's statement in support of the reduction is published at 43 Fed. Reg. 5,918-70 (1978).

17. The OSH Act provides that any person "adversely affected" by a health and safety standard may challenge the validity of that standard in an appropriate United States court of appeals. 29 U.S.C. § 655(f) (1976).

18. 581 F.2d 493 (5th Cir. 1978), *aff'd*, 448 U.S. 607 (1980).

19. *Id.* at 510.

20. *Id.* at 502.

21. Industrial Union Dep't, AFL-CIO v. American Petroleum Inst., 448 U.S. 607 (1980).

22. Joining Justice Stevens were Justices Stewart, Powell, and Chief Justice Burger. The Chief Justice and Justice Stewart joined in the opinion in its entirety, while Justice Powell joined only in Parts I, II, III-A, III-B, III-C and III-E. *Id.* at 611, 664.

exposure to 1 ppm *would not*; OSHA acted upon an *assumption*, according to the plurality, that exposure to 10 ppm of benzene *might* cause leukemia and that the number of such cases *might* be reduced by lowering the permissible exposure level to 1 ppm.

The plurality interpreted section 6(b)(5), which empowers the Secretary to promulgate occupational health standards "reasonably necessary or appropriate to provide *safe or healthful* employment," as requiring the Secretary to make a threshold finding that a workplace is "unsafe" before issuing a toxic material exposure standard.[23] According to the plurality, "safe" does not mean risk-free, and a workplace is not "unsafe" unless it poses a "significant risk of harm" to the worker.[24]

Even a finding of significant harm may not suffice. The dicta intimate that the Secretary must further show improved safety at the new lower levels.[25] And the burden of proof, for an informal and non-adversarial rulemaking, is unusually heavy: both findings apparently must be supported by a preponderance of the evidence.[26] The OSH Act itself specifies a relatively probing scope of judicial review, the substantial evidence test:[27]

> As we read the statute, the burden was on the Agency to show, on the basis of substantial evidence, that it is *at least more likely than not* that long-term exposure to 10 ppm of benzene presents a significant risk of material health impairment . . . [t]he closest [OSHA] came to making a

---

23. *Id.* at 642, interpreting 29 U.S.C. § 655(b)(5) (1976) (emphasis added).

24. *Id.*

25. *Id.* at 642-44.

26. The Court cited no independent statutory authority for this newly imposed burden of proof.

The same considerations that govern the allocation of the burden of proof are relevant to its weight. The agency as the challenger should bear a heavier burden of proof in situations where congressional policy favors preserving the status quo. F. JAMES & G. HAZARD, CIVIL PROCEDURE 252-53 (1977); C. MCCORMICK, HANDBOOK OF THE LAW OF EVIDENCE 785-89 (1972). The opposite is true in the context of occupational health. The remedial bent of the OSH Act, which was enacted in response to "the grim history of our failure to heed the occupational health needs of our workers," S. REP. No. 1282, 91st Cong., 2d Sess. 2 (1970), supports a minimal quantum of proof. JAMES & HAZARD at 249-53. The fact that industry, not OSHA, controls access to exposure data, does likewise. *Id.*

27. Section 6(b)(5), 29 U.S.C. § 665(b)(5) (1976), requires the agency to support its findings with substantial evidence. The substantial evidence test is a departure from the more deferential "arbitrary and capricious" standard which typically governs informal rulemaking proceedings. *Compare* section 10(e)(2)(A) of the Administrative Procedure Act, 5 U.S.C. § 706(2)(A) (1976). Although the *Benzene* case did not discuss the meaning of the substantial evidence requirement, the plurality gave this comparatively stricter standard of review prominence throughout the opinion. Since the decision turned on OSHA's failure to find a significant risk of material health impairment, the Court did not need to decide whether OSHA had satisfied the substantial evidence test. 448 U.S. at 640. *See also infra* note 31.

Subsequent circuit court decisions have differed over the degree of judicial deference the substantial evidence test warrants. For respectively intrusive and deferential conceptions of the OSH Act "substantial evidence" standard after the *Benzene* case, compare Texas Independent Ginners Ass'n v. Marshall, 630 F.2d 398, 404 (5th Cir. 1980) with United Steelworkers v. Marshall, 647 F.2d 1189, 1207 (D.C. Cir. 1980).

> finding that benzene presented a significant risk of harm in the workplace was its statement that the benefits to be derived from lowering the permissible exposure level from 10 to 1 ppm were "likely" to be "appreciable."[28]

The plurality noted in dicta that its newly announced significant risk requirement is not a "mathematical straitjacket," and that OSHA is not required to conclude with scientific certainty that such a risk exists before regulating toxic substances in the workplace.[29] Nevertheless, the plurality's view of the substantial evidence test[30] apparently forces OSHA to rely on quantified data rather than partly or wholly on policy considerations when it promulgates exposure standards.[31]

Writing separately, Justice Powell concluded that the OSH Act requires the agency "to determine that the economic effects of its [health] standard bear a reasonable relationship to the expected benefits."[32] Justice Rehnquist concurred in the judgment, but concluded that sec-

---

28. 448 U.S. at 653. Justice Powell did not join in this section of the pluarlity's opinion.

This passage seems to authorize the reviewing court to independently assess whether a preponderance of the evidence exists in determining whether the agency meets the substantial evidence test. In effect, this would import a traditional trial court burden of proof standard into the judicial scope of review. This reading, if correct, would contradict the Court's own pronouncements which interpret the substantial evidence test as requiring "something *less* than the weight of the evidence. . . ." Consolo v. Federal Maritime Comm'n, 383 U.S. 607, 620 (1966) (emphasis added). In *Consolo*, the Supreme Court held that an agency finding cannot be reversed for lack of substantial evidence merely because substantial evidence supports a *contrary* outcome. *Id.* at 618. There the Court remarked:

> [T]he possibility of drawing two inconsistent conclusions from the evidence does not prevent an administrative agency's finding from being supported by substantial evidence.

Id. at 620.

29. 448 U.S. at 656. For lower court elaboration of the "significant health risk" standard and its methodological requirements, see Pratt & Whitney Aircraft v. Secretary of Labor, 649 F.2d 96 (2d Cir. 1981); Texas Independent Ginners Ass'n v. Marshall, 630 F.2d 398 (5th Cir. 1980); American Textile Mfrs. Inst. v. Donovan, 617 F.2d 636 (D.C. Cir. 1981).

30. *See supra* note 27.

31. Responding to the dissent, the plurality sympathized with the need for value judgments in rulemaking:

> [W]hile the Agency must support its finding that a certain level of risk exists by substantial evidence, we recognize that its determination that a particular level of risk is "significant" will be based largely on policy considerations.

448 U.S. at 655 n.62. The opinion conceded that section 6(b)(5), which allows the Secretary to regulate on the basis of the "best available evidence," relaxes the seemingly stringent substantial evidence test:

> As several Courts of Appeals have held, this provision requires a reviewing court to give OSHA some leeway where its findings must be made on the frontiers of scientific knowledge.

Id. at 656. The opinion went on, however, to define such leeway in terms of the *data* that would suffice, implying that *some* empirical data is necessary. *Id.* Nowhere did the plurality ever suggest that policy considerations alone could constitute substantial evidence, e.g., in the face of a "no effects" finding. *See infra* text accompanying notes 58-60, 103-104, 114-16.

32. 448 U.S. at 667 (Powell, J., concurring). The plurality did not reach the issue of cost-benefit analysis, since "the Secretary did not make the required threshold finding" of a significant health risk. *Id.* at 640.

tion 6(b)(5) of the Act constituted an invalid delegation of legislative power, insofar as it applies to:

> [A]ny toxic substance or harmful physical agent for which a safe level . . . is, according to the Secretary, unknown or otherwise "infeasible."[33]

Four justices dissented[34] from the judgment affirming the lower court's invalidation of the 1 ppm benezene standard, calling the plurality's analysis "both extraordinarily arrogant and exraordinarily unfair,"[35] and in flagrant disregard of restrictions on judicial authority.[36] After reviewing the administrative record and the evidence compiled by the Secretary in support of the new occupational health standard for benzene, the dissent concluded that the Secretary's determinations were supported by "substantial evidence"[37] and therefore that the 1 ppm standard should withstand judicial review.

The dissent found "particularly embarrassing" the plurality's construction of the OSH Act as requiring the Secretary to determine that a "significant risk of harm" exists before an occupational health standard may be promulgated.[38] The dissenting justices agreed that the "significant risk" standard did violence to congressional intent to the extent that it rendered superfluous the first sentence of section 6(b)(5), which requires the Secretary to promulgate the standard "which most adequately assures . . . that no employee will suffer material impairment of health."[39]

In response, the plurality downplayed the harshness of its standard.[40] Despite qualifying language, however, the plurality gave prominence to the "significant risk of harm" test. Although the test is

---

33. *Id.* at 688 (Rehnquist, J., concurring).

34. Justices Marshall, Brennan, White and Blackmun.

35. 448 U.S. at 695 (Marshall, J., dissenting).

36. *Id.* at 688 (Marshall, J., dissenting).

37. *See supra* note 27.

38. 448 U.S. at 709 (Marshall, J., dissenting).

39. *Id.*, *quoting* section 6(b)(5), 29 U.S.C. § 655(b)(5) (1976). *See* Baird, *INDUSTRIAL UNION DEPARTMENT, AFL-CIO v. AMERICAN PETROLEUM INSTITUTE: Limiting OSHA's Authority to Regulate Workplace Carcinogens Under the Occupational Safety and Health Act*, 9 B.C. ENVTL. AFF. L. REV. 623, 670-71 (1981).

The commentators roundly criticize the effort to ground the "significant risk" requirement in section 3(8), which limits occupational health and safety standards to those "reasonably necessary or appropriate" to provide a safe workplace. *See supra* text accompanying note 7. For a suggestion that section 3(8) necessarily gives broad discretion to OSHA, since there was no legislative debate to the contrary, see Comment, *Avoiding the Use of Cost-Benefit Analysis in the Context of Occupational Safety and Health, The Requirement of Significant Risk; INDUSTRIAL UNION DEPARTMENT AFL-CIO v. THE AMERICAN PETROLEUM INSTITUTE*, 22 B.C. L. REV. 1149, 1156 (1981) [hereinafter cited as Boston College Comment]. One observer remarked that the phrase "reasonably necessary or appropriate" "almost never [is] used to impose a requirement other than that agency actions bear a reasonable relation to statutory purposes." *The Supreme Court, 1979 Term*, 94 HARV. L. REV. 75, 246 (1980).

40. *See supra* notes 29-31 and accompanying text.

accordionlike,[41] a strict interpretation would certainly hamper OSHA's generic policy on carcinogens.[42] To the extent that OSHA cannot rely upon its generic cancer policy, it will have to evaluate suspect substances on a case-by-case basis,[43] costing time and expense.

Those who argue that OSHA will be forced to postpone indefinitely many of its regulations are surely correct. Even so they still may be too optimistic. The commentators,[44] like the Court, seem to assume that the agency *can* provide the relevant information to demonstrate a significant risk of material health impairment when one exists. The remainder of this article will examine why this assumption is mistaken in many particular circumstances and logically impossible in others.

III

FACT AND POLICY ISSUES IN EPIDEMIOLOGICAL STUDIES

The *Benzene* case plurality implicitly committed itself to scientific models that do not correspond to the practical, statistical or policy realities in which they must be applied. In order to establish a quantifiable risk of harm to workers from a disease like cancer, scientists must rely upon either animal studies or epidemiological studies of humans exposed to the potential carcinogen. As this section will discuss, both are dependable indicators that a substance is a carcinogen at some exposure level, but neither can reliably assess the risk at particular exposure levels.

A. Measuring Risks

Relative risk and attributable risk are the two basic measures of health risks in the workplace that quantitative risk assessments generate.[45] The first statistic, relative risk, is the ratio of the incidence rate of disease for those exposed to a disease-causing substance to the incidence rate among those not exposed:[46]

$$\text{Relative risk} = \frac{\text{incidence rate among exposed}}{\text{incidence rate among nonexposed}}$$

41. *See supra* note 31. *See supra* note 29 for subsequent lower court cases applying the standard.

42. Boston College Comment, *supra* note 39, at 1169. *See also* Baird, *supra* note 39, at 683; Comment, *The Supreme Court's New Occupational Health Standard for Benzene Exposure: Regulated Industry's Triumph Over Employee Health*, 1981 UTAH L. REV. 525, 531, 533. *See supra* text accompanying note 15.

43. Boston College Comment, *supra* note 39, at 1169.

44. *See, e.g., id.* at 1169-70.

45. I consider the only concept of attributable risk relevant to our purposes. *See* J. MAUSNER & A. BAHN, EPIDEMIOLOGY: AN INTRODUCTORY TEXT 322-23 (1974) for a discussion of several other concepts of risk.

46. *Id.* at 322.

For instance, if the incidence rate of lung cancer in the general "nonexposed" (i.e., non-smoking) population is 7/100,000, and the incidence rate among heavy smokers is 166/100,000, the relative risk is 23.7.[47]

Attributable risk is the "arithmetic . . . difference in incidence rates between an exposed group and a nonexposed group."[48] Thus, in the case of lung cancer, the risk attributable to heavy smoking is 166/100,000 minus 7/100,000 or 159/100,000. Since attributable risk measures how many cases can be avoided by making a workplace as safe as the general environment, it is crucial for estimating the costs and benefits of removing a health hazard.[49]

B. Difficulties with Animal Studies

At the 1978 OSHA hearings on the agency's new carcinogen policy, most if not all of the expert witnesses agreed that "animal evidence alone should serve as the basis" for regulating carcinogens.[50] A host of reasons recommend this advice. To begin with, all but one, or possibly two, of the chemicals that "cause cancer in humans . . . are known to cause cancer in animals." Of these chemicals, most that induce cancer in one mammalian species also induce cancer in others. The pathological development of tumors in various species of animals resembles that in man. Finally, human and animal molecular interactions show a close resemblance in laboratory testing.[51]

Animal studies have several advantages over their human counterparts. Few industrial chemicals have been adequately tested by epidemiological studies to discover whether they cause cancer in humans. For one thing, epidemiological studies are notoriously insensitive, and a survey that fails to show positive results merits little weight. Moreover, it is still too soon to observe the effects of newer synthetic chemicals, given the long latency period of cancer. Ethical considerations also come into play. There is no moral justification for waiting for "evidence of harm in exposed workers when risks can be established rela-

47. Figures are from *id.*, *quoting* Doll & Hill, *Lung Cancer and Other Causes of Death in Relation to Smoking*, 2 BR. MED. J. 1071, 1073 (1956).

48. MAUSNER & BAHN, *supra* note 45, at 322.

49. All of these risk measures are potentially misleading. No one statistic tells policymakers all they need to know. Relative risk may exaggerate the risk of a rare disease. One could have a relative risk of 24 whether the prevalence of the disease in the population is 1/100 or 1/10,000 as long as exposure to some health hazard makes one 24 times more likely than the general population to contract it. Attributable risk may overestimate risks unless one is explicit about the denominator. In many ways relative risk is best, used in conjunction with the risk rate to the general population (7/100,000 in the smoking example) for it provides a measure of the danger of an exposed group relative to the general population. In addition, as we will see, it is the easiest to provide through scientific studies.

50. 45 Fed. Reg. 5,061 (1980).

51. Testimony of Dr. Arthur Upton, Director, National Cancer Institute. *Id.*

tively quickly by animal experimentation."[52]

There are serious problems, however, in using animal studies to establish health risks to humans. Detecting an incidence of cancer as low as .01% (1/10,000) in experimental animals, for example, "would require hundreds of thousands of animals."[53] Since such experiments would be prohibitively expensive, researchers instead expose relatively small groups of experimental animals to high doses of a suspected carcinogen and then use "biologically reasonable models in extrapolating the results to estimate risk at low doses."[54] For example, OSHA identifies at least three relatively simple models for extrapolating from high dose to low dose levels in animals.[55] These models produce results that differ by as much as a factor of 1,000, or three orders of magnitude.[56] Extrapolating from animals to human beings compounds the uncertainty.[57] A striking illustration of these difficulties is provided by a recent OSHA survey of private estimates of the lifetime risk from exposure to 1 ppm vinyl chloride. Among the thirteen studies surveyed, the estimates of risk varied a millionfold.[58]

Further, suppose, as the plurality intimates, that the OSH Act requires quantifying the risk to workers by a preponderance of the evidence.[59] If this means OSHA must have at least fifty-one percent confidence in the results, the plurality's requirement defies established scientific reasoning. It is an accepted principle of nondeductive scientific reasoning that the greater the number of hypotheses available to explain given data or to predict a certain result, the lower the probability that any one of the hypotheses is correct.[60] For example, if two mutually exclusive hypotheses both explain the results of an experiment, the probability that either one is true cannot exceed fifty percent. With numerous hypotheses for extrapolating from high to low dose responses in animals and several others for extrapolating from animals to humans, a respectable researcher cannot claim that the chosen risk quantification is more probably true than false. Were OSHA required to quantify the risks to humans, it could not in good conscience satisfy the plurality's preponderance of the evidence requirement.

---

52. Testimony of Dr. David Rall, Director, National Institute of Environmental Health Sciences. *Id.*

53. NATIONAL ACADEMY OF SCIENCES, DRINKING WATER AND HEALTH 55 (1977), *quoted at* 45 Fed. Reg. 5,087 (1980).

54. *Id.*

55. 45 Fed. Reg. 5,184 (1980). *See also* Leape, *Quantitative Risk Assessment in Regulation of Environmental Carcinogens*, 4 HARV. ENVTL. L. REV. 86, 100-03 (1980).

56. *Id.* at 98-100; 45 Fed. Reg. 5,185 (1980).

57. *See generally id.* at 5187-97.

58. *Id.* at 5200.

59. *See supra* text accompanying notes 27-31.

60. *See* L. WRIGHT, BETTER REASONING: TECHNIQUES FOR HANDLING ARGUMENT, EVIDENCE AND ABSTRACTION 92-97 (1981).

The preceding discussion assumes that the suspect substance is a known carcinogen, at least in animals. Animal studies that show "no harmful effect" pose problems of an entirely different order. One might suppose that such optimistic results rule out the risk of harm to human beings. That is not the case, for a "no effects" or negative study in one animal species does *not* show a suspected substance is safe for animals or humans. Because the difficulties of "no effect" or "negative" studies are even greater than those of positive studies, OSHA will rely on them as indicators that a substance is not carcinogenic only under very special circumstances. First, there must be no positive tests to the contrary, for positive tests always supersede negative tests. Second, the results of negative tests must be conclusively established, and established for two different animal species.[61] The second requirement is so difficult to satisfy that one researcher has said that in order to establish with "95 percent confidence that a realistic dose of just one substance causes fewer than one cancer in a million subjects would require a test involving at least 6,000,000 animals."[62]

C. Difficulties with Epidemiological Studies

Epidemiological studies of human beings exposed to a substance in the past or present provide the best scientific evidence that a substance is carcinogenic at specific levels of exposure. Such studies, however, face a number of practical and theoretical difficulties.[63]

1. Discovering Risks

Human health risks at particular exposure levels can be detected either through cohort or case-control epidemiological studies. A cohort study compares the incidence of disease in a group exposed to relatively high levels of a health hazard with the incidence of disease in a group representative of the general population.[64] In a case-control study, "people diagnosed as having a disease (cases) are compared with persons who do not have the disease (controls)."[65] Clearly, fewer people are needed in a case-control than in a cohort study, for only those

61. 45 Fed. Reg. 5,087.

62. Mantel & Schneiderman, *Estimating "Safe" Levels, A Hazardous Undertaking*, 35 CANCER RESEARCH 1379, 1383 (1978). McGarity, *Substantive and Procedural Discretion in Administrative Resolution of Science Policy Questions: Regulating Carcinogens in EPA and OSHA*, 67 GEO. L.J. 729, 734 (1979), makes the same point, referring to Schneiderman, Mantel & Brown, *From Mouse to Man—Or How to Get from the Laboratory to Park Avenue and 59th Street*, 246 ANNALS OF THE N.Y. ACAD. OF SCI. 237, 241 (1975). *See infra* text preceding note 104. The example quoted is clearly one extreme, but it suggests the problems researchers face with "no effect" studies.

63. I am indebted to Dr. Helaine Pleet for advice about and information for the discussion in this section.

64. MAUSNER & BAHN, *supra* note 45, at 312-13, 322-25.

65. *Id.* at 312-13.

with the disease, not those exposed to a risk factor, are the objects of examination. In either case, a positive correlation between a risk factor and the disease means that those exposed will tend to develop the disease and those not exposed will tend not to develop it.

Case-control studies are essentially retrospective. The researcher takes a group that has contracted a disease, compares characteristics of that group and its environment with those of the general population, and tries to isolate factors that might have caused the disease. Cohort studies can be retrospective or prospective. In a prospective study, a sample population exposed to a potential disease-causing factor is followed forward in time. Its disease rate is then compared with the disease rate of a group selected randomly from the general population. In a retrospective study the same method is employed, only using historical data. The researcher studies the cold record of a group of people exposed to some suspected disease-causing factor over some time period to establish their disease rate. That rate is then compared with the disease rate for a similar group from the general population.

Each kind of study has its advantages and its problems. The case-control study can provide an estimate of relative risk,[66] and incurs little expense, due to the small sample size.[67] Likewise, the case-control approach is especially suited to the study of rare diseases.[68] Its disadvantages, however, are numerous. It requires thorough and sophisticated diagnosis to insure a properly representative control group.[69] The incidence rate cannot be derived, for there are no appropriate denominators for the populations at risk.[70]

The case-control approach, like the retrospective cohort study, requires historical information about its subjects. This creates problems of accuracy and documentation.[71] Information about past events regarding doses and exposure may not be available. In some cases, records have not been kept; in others, no one measured the risk factor. Sometimes it is difficult to separate and measure the effect of one risk factor compared with another.[72] For example, rubber workers are exposed to vinyl chloride, polychlorinated biphenyls, chloroprene, selenium compounds, benzidrine and its salts, aniline, carbon tetrachloride and benzene, all of which are either suspected or federally regulated

66. *Id.* at 316, 318. The estimation can be made only under certain assumptions, since the magnitude of the incidence rates cannot be derived. *Id.*; *see infra* text accompanying note 70.

67. *Id.* at 318.

68. *Id.* at 328.

69. *Id.* at 315.

70. Unlike a retrospective cohort study. *Id.* at 316.

71. *See infra* text accompanying notes 78-87.

72. MAUSNER & BAHN, *supra* note 45, at 93. These same problems plague retrospective cohort studies. *See infra* text accompanying note 82.

carcinogens.[73] Case-control studies also run the risk of bias, since both the informant and the interviewer know the subject has the disease.[74]

In contrast, a prospective cohort study is free from bias. And cohort studies yield incidence rates and attributable risk as well as relative risk.[75] But cohort studies, particularly prospective ones, have their drawbacks too. All cohort studies require large numbers of subjects and a long followup period, which increases with the latency period. Predictably, then, such studies are costly.[76] In a prospective study, subjects may drop out. (In a retrospective study, they may be difficult to trace.) Criteria and method may change as the years progress. Finally, since most carcinogens have a latency period of five to fifty years,[77] there are ethical problems in exposing people to suspected carcinogens for the period a prospective cohort study requires.

2. *Practical Problems*

The costs and bioethical aspects of a prospective cohort study prompt most epidemiologists to rely on case-control or retrospective cohort studies. For this reason, it is worthwhile examining the numerous practical difficulties inherent in relying on historical information. Frequently, industry data on workplace exposure to harmful substances is inadequate.[78] When this is a problem, epidemiologists must resort to a worker's duration of employment as a measure of total exposure. The proper interpretation of this data, like any indirect measurement, is understandably a point of controversy,[79] and in any event, often companies fail to keep this information.[80] Even if such data exists, it does not necessarily reveal *which* employees actually worked in the contaminated quarters.[81]

As was mentioned above, workers are often exposed to more than one chemical agent, which makes both case-control and retrospective cohort studies much more difficult, if not impossible, to conduct.[82] In addition, the dosage of exposure frequently varies over time.[83]

Job mobility and population heterogeneity also take their toll. Since there is considerable job mobility in American employment, the

73. Schottenfeld & Haas, *Carcinogens in the Workplace*, 29 CA-A CANCER J. FOR CLINICIANS 144, 156-59 (1979).

74. MAUSNER & BAHN, *supra* note 45, at 320.

75. *Id.* at 323-24.

76. *Id.* at 324-25.

77. *See* representative latency periods for various causes in Schottenfeld & Haas, *supra* note 73, at 156-59.

78. 45 Fed. Reg. 5,040 (1980).

79. *Id.*

80. *Id.*

81. *Id.*

82. *Id. See supra* text accompanying notes 72-73.

83. *Id.* at 5,043.

effect of a carcinogen easily can be overlooked. The briefer the exposure, the longer the latency period of the disease.[84] Even if an epidemiologist has data for one population and its set of characteristics for either a cohort or case-control study, it is difficult to extrapolate to other populations and their characteristics.[85] Populations can vary in socioeconomic status, age at which exposure occurred, smoking history, and other factors that affect susceptibility.[86]

These practical problems make it difficult, perhaps nearly impossible, to obtain scientifically respectable results to quantify health risks to workers and to provide even the most rudimentary dose-response curve for a substance. In fact, one researcher has suggested that the relevant data is missing for most chemical substances and industrial processes.[87] As we have seen with regard to animal studies, there are more complex ways of estimating such information, but they introduce additional uncertainties which are incompatible with the preponderance of evidence requirement.

3. Theoretical Problems

Even without these practical difficulties, serious mathematical and statistical problems would remain. These statistical problems combined with practical constraints are incompatible with the Court's requirement that OSHA demonstrate a "significant health risk" before promulgating a more restrictive occupational health standard.

To see these points we must delve into the theory of hypothesis acceptance and rejection. The need is not to learn statistics, but rather to introduce enough terminology to characterize the main risk and proof variables with which epidemiologists must work.

In studying a carcinogen such as benzene, a scientist considers two hypotheses. The first posits that exposure to benzene is *not* associated with greater incidence of a certain disease, say, leukemia or aplastic anemia, than that found in the general population. The second hypothesis assumes that exposure to benzene *is* associated with a greater incidence of such diseases. In statistical parlance, the first is known as the null hypothesis, while the second is called the alternative hypothesis.[88] When should one reject the null hypothesis and accept the alternative hypothesis, and what errors might there be in accepting one or the other?

Since an epidemiological survey studies small samples of both exposed and unexposed populations, there remains the possibility that the

84. *Id.* at 5,040.
85. *Id.* at 5,041.
86. *Id.* at 5,042.
87. *Id.* at 5,044.
88. A. FEINSTEIN, CLINICAL BIOSTATISTICS 320-21 (1977).

study will erroneously favor a certain hypothesis by chance alone. That is, a researcher risks inferential errors from studying a sample instead of the whole population in question. The researcher runs the risk of making one of two kinds of mistakes. The study might show that the null hypothesis should be rejected (and the alternative hypothesis accepted) when in fact the null hypothesis is true. This is called a type I error.[89] Or, the study might show that the null hypothesis should be accepted when in fact the null hypothesis is false and the alternative hypothesis is true. This is a type II error.[90]

TABLE I

	Null hypothesis is actually true, e.g., benzene is not positively associated with cancer	Null hypothesis is false, alternative hypothesis is true, e.g., benzene is associated with cancer.
Null hypothesis is accepted	No error	Type II error
Null hypothesis is rejected and alternative hypothesis is accepted	Type I error	No error

In the abstract, the fact that two kinds of errors can be committed in such studies is of little interest. Instead, one wants to know the probability of committing such errors by chance alone. The probability of committing a type I error is normally designated α and the probability of committing a type II error is designated β.[91] Conventionally, α is set at .05 so that there is only a one in twenty chance of rejecting the null hypothesis when it is true. In other words, if α is .05, and the study shows the null hypothesis should be rejected, the chance of doing so correctly is ninety-five percent.[92] Convention also sets β between .01 and .20, when α is .05. When β is .20, one takes one chance in five of rejecting the alternative hypothesis as false when it is

89. *Id.* at 321-22.

90. *Id.* at 324-25. Table I is adapted from *id.* at 325 (Table I).

91. *See generally id.* at 320-34.

92. Walter, *Determination of Significant Relevant Risks and Optimal Sampling Procedures in Prospective and Retrospective Comparative Studies of Various Sizes*, 105 AM. J. OF EPIDEMIOLOGY 387, 391 (Table 2) (1977). I do not discuss how the various statistical variables are derived for a particular study from the raw data, but only wish to show the conceptual relationships among them.

true, for example, of saying benzene is not associated with leukemia when in fact it is.[93]

Perhaps more important is 1 - β, or the "power" of a statistical test. Since the probability of accepting or rejecting the alternative hypothesis is 1, and since β is the probability of rejecting the alternative hypothesis (accepting the null hypothesis) when the alternative hypothesis is true, 1 - β is the probability of accepting the alternative hypothesis when the alternative hypothesis *is* true. Obviously this is a crucial statistic, for it tells the chance that benzene, for example, is indeed associated with cancer. When β equals .20, the power of one's statistical test is .80. This means the scientist has an eighty percent chance of accepting the alternative hypothesis as true when it *is* true.

One can think of α, β, and 1 - β as measures of the "risk of error" or "standard of proof." What chance of error is a researcher willing to take? Is a twenty percent (β=.20) chance of saying benzene does not cause cancer, when in fact it might, an acceptable risk? When workers may be contracting cancer even though a study shows they are not, is their health worth a twenty percent gamble?

Alternatively, we might think of α, β, and 1 - β as measures of proof. What standards of proof do we demand of researchers? Must benzene be condemned by mere preponderance of evidence, say fifty-one percent (1 - β = .51)? That is to say, must researchers be fifty-one percent sure that benzene is a carcinogen with high risk for workers before regulating it? Should OSHA be permitted to take a forty-nine percent chance (β=.49) that benzene is not a high risk carcinogen to workers, when in fact it might be? These questions only precede more complex matters, for the standards of proof demanded of statistical studies have implications for the costs of doing such studies and for the relative risks that can be detected.

In order to fully see these trade-offs we need two other variables: N, the total number of people studied in the exposed and unexposed groups, and δ, the relative risk one wants to detect.[94] At the outset of the study one usually sets δ at some value considered an unreasonable risk to health for public policy purposes, say, a relative risk of 2, 5, or 10.[95] The value chosen depends upon many factors, including the seriousness of the disease, its incidence in the general population, and how great a risk the exposed group justifiably should be expected to run.[96]

93. FEINSTEIN, *supra* note 88, at 324-25.

94. *See* FEINSTEIN, *supra* note 88, at 320-24.

95. That is, the incidence of disease in a group exposed to a risk factor would be two, five or ten times greater respectively than the incidence of disease in the general population. *See supra* text accompanying notes 46-47.

96. To be more precise about this, however, δ need not be set in advance, for it depends upon the number of people to be studied and the prevalence of the underlying disease, in addition to the values of α and β.

If one wishes to detect a very small relative risk between two groups, e.g., a relative risk of 2 for a rare disease, large numbers of exposed and unexposed individuals must be studied. A large relative risk, such as a risk of 6, requires fewer individuals to obtain statistically significant results.

The relation between the relative risk and the number of subjects raises a more general issue. α, β, δ and N are interrelated. If any three of them are known the fourth can be determined. Typically, one would not be interested in solving for α, and thus it is specified at the outset. A researcher who knows α, δ, and N can solve for β and thus discover what the power of the study (1 - β) would be. If the observed data support the alternative hypothesis, but the power of the test is fifty percent, the results are not conclusive. Alternatively, knowing α, N and β (or alternatively 1 - β) makes it possible to calculate the lowest risk for which one could test. For example, it might be impossible to test for a relative risk smaller than 6, if one had an insufficient number of people to study.

Because the variables are interdependent, crucial trade-offs are inevitable. Consider four alternatives.

(1) In the first example, suppose we want to discover whether a suspected carcinogen, C, is associated with a particular cancer L. Suppose the incidence of L in the general population is $^{8}/_{10,000}$, and we want to be ninety-five percent sure that if our study shows that there is not an association between C and L, none exists. Thus we set α at .05. Suppose we set β at a typical value of .20, so we have only one chance in five of committing a type II error, and an eighty percent chance of statistically accepting the alternative hypothesis when it is true. Suppose further that we regard a relative risk of 3 (δ= 3) as a "serious" risk worth investigating for public health purposes.[97] Given these values, we would have to study 7,695 people exposed to C, and (I assume for the sake of simplicity) an equal number who are not exposed to obtain statistically significant results at a relative risk of 3.[98] That would be prohibitively expensive.

(2) Suppose we can study only 2,150 workers who were exposed to C. We want to be ninety-five percent confident (α = .05) of results favoring the null hypothesis and eighty percent confident (1 - β = .80) of any results that favor the alternative hypothesis when the prevalence

97. This is a complicated matter, for relative risk does not tell us certain information. A three-times normal risk of a certain kind of cancer may merit further investigation. If the prevalence of disease in the general population is $^{1}/_{100,000,000}$, the risk is not worth study, unlike a disease with a prevalence of $^{1}/_{100}$ or $^{1}/_{1,000}$. *See supra* note 49.

98. All the figures used in these examples are courtesy of Dr. Helaine Pleet and the computers at the Centers for Disease Control in Atlanta, Georgia. That information is summarized in the Appendix. Similar numbers can be calculated from the equation in note 100 *infra*. The particular numbers in alternative (1) come from line 3(c) of the Appendix.

of the underlying disease is $\frac{9}{10,000}$. What relative risk can we hope to detect? At best we only could detect a relative risk of 6, or two times what we thought was "serious" enough to warrant social attention.[99] Put differently, given α at .05, β at .20, and N at 2,150, a relative risk of 6 is the smallest risk we could detect with eighty percent confidence.

(3) Epidemiological studies are flexible enough, however, to detect a lower relative risk, say $\delta = 3$, by making some trade-offs. If we kept N and α constant ($N = 2,150$, $\alpha = .05$), β would have to be correspondingly raised to .44, lowering $1 - \beta$ to .56.[100] Because $\beta = .44$, there is a forty-four percent probability of mistaking a toxic substance for a benign substance by chance alone. We could only hope to find a relative risk of 3 among the 2,150 exposed workers if we were willing to foresake confidence in the results *and* if we were willing to take a considerable chance of leaving workers exposed to a possibly harmful carcinogen.

(4) Epidemiological flexibility allows a final alternative. If we had only 2,150 workers to study and wanted to be able to detect a relative risk as low as 3, we could increase α instead of β. With β set at .20, the resultant α would equal .33.[101] This creates a thirty-three percent chance of concluding that benzene is associated with a threefold increase in cancer when it is not. That is, we could be only sixty-seven percent confident of negative results, instead of having the ninety-five percent confidence researchers typically demand. A standard epidemiological text notes that "it is generally accepted in the medical literature that it is [only] safe to reject a null hypothesis when there is a less than five percent chance of being wrong (type I error)."[102] Thus, even though we can reach statistically significant results for a relative risk of 3 by increasing α to .33, this is inconsistent with current scientific methodology and would seriously undermine the credibility of the study.

---

99. *Id.* The particular numbers for alternative (2) come from line 5(c) of the Appendix.

100. β=.44 is a value computed from the following equation:

$$n = K\frac{(p_1q_1 + p_2q_2)}{(p_1 - p_2)^2}$$

Walter, *supra* note 92, at 387-88 (equation (2)). $K = (z_\alpha + z_\beta)^2$, p_1 = disease rate among exposed individuals, p_2 = disease rate among non-exposed, n = the size of each group (assumed to be equal), $q_1 = 1 - p_1$, $q_2 = 1 - p_2$, and z_α and z_β are the points in the normal distribution defined by the type I and II error rates respectively. The equation is for a one-tailed test of significance.

101. Based on computation from equation (2) in Walter, *supra* note 92, at 387-88.

102. A. RIMM, BASIC BIOSTATISTICS IN MEDICINE AND EPIDEMIOLOGY 201-02 (1980).

TABLE II

Cohort Study

Assume that the prevalence of disease L in the general population is 8/10,000. The study seeks to detect a relative risk of 3 ($\delta = 3$).

Alternative (1): $\delta = 3$, $\alpha = .05$, $\beta = .20$

best study, but too costly and time consuming to conduct

Must study 7,695 exposed and 7,695 unexposed workers

positive result: acceptable

Alternative (2):
$\alpha = .05$, $\beta = .20$, $N/2 = 2{,}150$

At best can test for $\delta = 6$

negative result: can only infer that relative risk is not as high as 6. Some chance that workers remain exposed to dangerous substance

Alternatives (2)-(4)

Alternative (3):

positive result: acceptable, but with only 56% confidence

$\alpha = .05$, increase β to .44, $N/2 = 2{,}150$

Can test for $\delta = 3$.

negative result: 44% odds that workers remain exposed to harmful substance

positive result: acceptable with 80% confidence. Disadvantage: using large α (= .33) may undermine scientific credibility and confidence in study

Alternative (4):
$\alpha = .33$, $\beta = .20$, $N/2 = 2{,}150$

Can test for $\delta = 3$.

negative result: 20% odds workers will remain exposed to harmful substance

From these examples we can construct the decision tree displayed in Table II. It is not immediately evident which alternative is the most attractive. Alternative (1) is excluded for reasons of cost. Alternatives (2) and (3) put workers at considerable risk, and alternative (4) risks undermining the credibility of the research because it is inconsistent with scientific practice. Whatever the final choice, however, the Court's "significant health risk" and quantification requirements impose hard choices on researchers. We can easily imagine cases in which the Court's requirement would make it statistically impossible to detect a significant health risk. Assume that a relative risk of 3 is considered legally significant for a disease that occurs in eight people of every 10,000. If there were only 1,000 workers to study (with α at .05 and β at .20), a relative risk could not be detected below 8.73, even if it turned out the substance in question *did* cause a threefold increase in mortality among workers. Workers would be exposed to a trebled increase in mortality although the study could not reliably detect the risk.[103]

Alternative (2) suggests some interesting results for "negative" or "no effect" studies. Assume a study is run on 2,150 exposed workers with α at .05 and β at .20, when the prevalence of the underlying disease is $^{8}/_{10,000}$. With these values, we only could be confident of detecting a relative risk of 6. But suppose no relative risk were detected, that is, the study was "negative" or showed "no effect" between the chemical C and the disease L. What could we infer? At most we would be justified in concluding that the relative risk was less than 6. It might be 5.8 or 1, but given the constraints on the study, we could not conclude so statistically. Thus, a "negative" result (i.e., accepting the null hypothesis) in this situation is of little value. Although we could not know it from the numbers of people studied, the workers might well face high risk. This example illustrates the general problem with "no effect" studies. Short of a positive association between the suspected substance and the disease, the most that can be inferred is that the relative risk to people is not as high as the relative risk tested for in the study.[104]

As striking as these examples are, they only intimate the statistical problems a cohort study of benzene-caused leukemia would pose. Alternatives (1) through (4) assume that the prevalence of the hypothetical disease L in the general population is $^{8}/_{10,000}$. In fact, however, the

103. This figure is taken from Walter, *supra* note 92, at 391 (Table 2).

104. By comparison, the *Benzene* case plurality noted that only 1,440 employees exposed to benzene worked in benzene plants. 448 U.S. at 616 n.6. Of course, thousands of other workers were exposed to benzene in other industries. To study a group of employees working in uniform conditions in a single plant, it might be difficult to obtain a large enough sample for a statistically significant result in a case-control study. One might of course be able to find exposure conditions in several different plants that would be uniform enough for a defensible study, but that would be fortuitous.

prevalence of leukemia (the carcinogenic disease related to benzene exposure) in the general population is between 8/100,000 and 12/100,000.[105] That is, it is ten times rarer than the hypothetical example assumed. As a consequence the difficulties sketched above are even greater for an epidemiological study of benzene exposure.

If we construct analogues to alternatives (1) through (4), the results change simply because leukemia is rarer than the hypothetical disease L. In alternative (1), holding α, β and δ constant and changing the incidence of L to eight per 100,000 for leukemia, we would have to study 77,087 people exposed to benzene and an equal number not exposed to obtain statistically significant results.[106] In alternative (2), holding α, β and δ constant and using the prevalence rate of leukemia, we would have to study 21,580 people in order to detect a relative risk of 6.[107] Pared back to 2,150 subjects, we could detect a statistically significant relative risk no lower than thirty-nine.[108] If we were forced to study as few as 1,000 people, we could reliably detect a relative risk no lower than 50.[109] Studying as many as 10,000, we could detect a relative risk of no less than 9.[110] In alternative (3), holding α, δ, and N constant, we would lack even fifty percent confidence in our results.[111] Even increasing α to sixteen percent would not decrease the chance of a type II error below fifty percent.[112] With a disease as rare as leukemia, and no more than 2,150 subjects, it is impossible to obtain statistically useful results in a cohort study. Analogues of alternative (4) lead to

105. DEMOGRAPHIC ANALYSIS SECTION, DIV. OF CANCER CAUSE & PREVENTION, NATIONAL CANCER INST., MONOGRAPH NO. 57, SURVEILLANCE, EPIDEMIOLOGY AND END RESULTS: INCIDENCE AND MORTALITY DATA 1973-1977 662-63 & Table 51 (1981).

106. This figure is from line 3(d) of the Appendix.

107. This figure is from line 5(d) of the Appendix.

108. This figure is from Walter, *supra* note 92, at 388 (equation (4)).

109. This figure is from Walter, *supra* note 92, at 391 (Table 2).

110. *Id.*

111. I put this point in a general way because of statistical problems. A sample of 2,150 people is so small compared to the disease rate of leukemia that the assumptions underlying the usual epidemiological equations no longer apply. The most important of these assumptions is that there is a normal distribution of diseased individuals throughout the population. In a group of 2,150 people on the average only .17 people would contract leukemia. Since people come in multiples of 1, a probability procedure is needed to estimate how many discrete individuals are likely to contract leukemia. The problem is analogous to one of estimating how many times a coin will come up heads if it is flipped 25 times. The chance of k individuals contracting leukemia (with a prevalence of 8/100,000) is given by the formula for a Poisson distribution:

$$p(k) = \frac{e^{-M}(M)k}{k!}$$

where e is the mathematical constant and M is the mean of the sample population being studied, when $\alpha = 1.5\%$, $\delta = 3$, and $\beta = 90.5\%$. Thus, the chances of making a type II error are nearly one hundred percent. When $\alpha = 16\%$ and $\delta = 3$, $\beta = 59.7\%$, one could not be more than forty percent confident that a positive outcome was correct. Professor David J. Strauss of the Department of Statistics, University of California at Riverside, pointed this out to me.

112. *See supra* note 111.

similarly unsatisfactory results. The upshot is that the rarer a disease, the greater the problems faced by epidemiologists.

These problems are incident to a cohort study. A case-control study which looks only at diseased people and compares them with a control group requires fewer subjects, thus lowering the costs. The trade-offs and statistical difficulties imposed are exactly the same, however.[113] The trade-offs involved between relative risk and type II errors mean that either study may conclude that workers face no risk when in fact they do. The demand for risk quantification prior to regulation, while preserving a patina of precision about health risks, runs a substantial chance of leaving workers exposed to potentially dangerous substances. The problems just discussed are, of course, exacerbated for the "significant health benefit" model, for reasons already considered.

4. The Idea of a Legitimate "No (Legally) Significant Health Risk"

Given the problems outlined above, the issue then becomes whether we should err on the side of optimism or on the side of caution in assessing the risk of harm. The difficulties with animal and epidemiological human studies suggest that caution is the wiser tack.

Recent work by the International Agency for Research on Cancer (IARC) reinforces this conclusion. IARC studied 368 substances selected on the basis of two criteria: "a) evidence of human exposure and b) evidence of carcinogenicity in experimental animals and/or of a

---

113. The following table shows the number of subjects needed to calculate various values of relative risk for cohort and case-control studies of congenital heart disease, a relatively common disease whose prevalence is .008 (8/1,000) in the general population. The table assumes $\alpha = .05$ and $\beta = .10$.

Relative Risk	Cohort Study	Case-control study
2	3,837	188
3	1,289	73
4	712	45
5	478	34
6	280	24
7	168	18

Taken from Schlesselman, *Sample Size Requirements in Cohort and Case-Control Studies of Disease*, 99 AM. J. OF EPIDEMIOLOGY 381, 382-83 (Tables 2 and 3) (1974). Note that the requisite sample size would have to become larger as the disease becomes rarer. *See supra* text accompanying notes 106-112.

Despite the dramatic difference in the numbers of people needed for cohort and case-control studies, both studies have similar statistical problems. In each, the lower the relative risk one seeks to detect, the higher N must be. In addition, if N is fixed, the lower the relative risk, the lower the power of the study ($1-\beta$), and the less confidence we can have in the outcome. This increases the chance of a type II error, that is, of test results that show that workers are not at risk when in fact they are. In any case, both studies must consider and separately account for the carcinogenic effects of additional substances to which the worker subjects were exposed. *See supra* text accompanying notes 72-73; Walter, *supra* note 92, at 393. Negative results in both studies have the same inconclusive effect. *See supra* notes 61-62, 103-04, *infra* notes 114-16, and accompanying text.

human risk."[114] The study concluded that present epidemiological data are too limited to reliably detect a threat to man in the vast majority of substances that are known to induce cancer in animals:

> [t]wenty-six chemicals or industrial processes were identified as associated with an excess carcinogenic risk to man . . . and for a further 221, some evidence of carcinogenicity in one or more animal species was found. For the majority of these 221 chemicals evidence of human exposure exists but no evaluation of the carcinogenic risk to man was made, either because no epidemiological studies or case reports were found (205 compounds), or because the results were inconclusive (16 compounds). For the remaining 121 chemicals, the available studies were inadequate to make an evaluation of the presence or absence of a carcinogenic effect. . . . Thus, of the 368 compounds reviewed in the IARC monographs, *less than 8% had received adequate investigation to judge whether or not they had carcinogenic effects in humans. This is a grossly inadequate basis on which to assess the percentage of human cancers "caused" by occupational exposure.*[115]

Testifying before OSHA, Dr. Irving Kessler of Johns Hopkins University indicated that deficiencies exist in the quality as well as in the quantity of epidemiological studies of occupational groups:

> In fact there have been relatively few controlled epidemiological studies of occupational groups. Rather, the large majority of human studies have been conducted upon hospitalized patients with and without a given disease. Only in the last few years, with the emergence of interest in occupational health hazards, have occupational studies been undertaken by epidemiologists. With a few exceptions, most of these could not, until recently, have been classified as well designed and controlled epidemiological studies. This is an emerging and very important field where methodologic approaches are still being developed.[116]

The *Benzene* case thus invites the criticism that its proof standards are scientifically illegitimate in the sense that failure to quantify a significant health risk may result from a number of factors irrelevant to establishing the scientific truth that a health risk exists. Further, the IARC results show that 92 percent of all known or suspected carcinogens have not yet been positively identified as being either dangerous to or safe for humans in the workplace at *any* exposure level. The evidence of the presence or absence of carcinogenicity at *particular* exposure levels is even less conclusive.

---

114. Tomatis, Agthe, Bartsch, Huff, Montesano, Saracci, Walker & Wilbourn, *Evaluation of the Carcinogenicity of Chemicals: A Review of the Monograph Program of the International Agency for Research on Cancer (1971 to 1977)*, 38 CANCER RESEARCH 877, 878 (1978).

115. Testimony of Dr. David Rall, Director, National Institute of Environmental Health Sciences, 45 Fed. Reg. 5,044 (1980) (emphasis added).

116. *Id. See supra* note 31 and accompanying text.

5. Policy Assumptions Implicit in Epidemiological Studies

The plurality opinion suggests that occupational health regulations ultimately must rest upon empirical facts rather than upon policy considerations.[117] The outcome of epidemiological studies themselves, however, rest on and are shaped by assumptions that are the equivalent of social policy considerations.

The decision tree at Table II provides an apt illustration. Alternative (1) must be rejected for reasons of time and cost, although it promises the most accurate results. Epidemiologists must consider social policy trade-offs when weighing the desirability of obtaining the most accurate information against the cost and morality of doing so. Some studies are prohibitively expensive; others are unethical.

Alternatives (2) and (3) involve the latter problem. These approaches illustrate the trade-offs that must be made between relative risks traded for and the chance of making a type II error. Either (2) or (3) takes the risk of leaving workers exposed to potentially dangerous carcinogens which the study could not detect. Whether or not epidemiologists should conduct studies which run such risks goes to the very design of a study and infects its product. With either (2) or (3), researchers may obtain a "no effect" outcome and have a false sense of security while workers' lives are endangered.

Finally, studying the same number of workers, scientists could adopt alternative (4) and increase the risk of a type I error to 33 percent. At stake is the credibility of the study. Almost all scientific journals require that statistical results run no more than a five percent chance of a type I error.[118] The choice between (4) on the one hand and (2) and (3) on the other requires balancing the potential loss of credibility against the chance of leaving workers exposed to potentially harmful chemicals.[119]

These observations support several inferences. Any facts produced by an epidemiological study under less than ideal research conditions are heavily infected by policy considerations. Beyond the questions of scientific respectability and costs, scientists must render judgments of the very kind the Act requires: how great a risk to workers' health should be countenanced? These choices are especially critical when a study shows "no effect" between exposure to a substance and the presence of disease. In planning their studies, scientists must make many of the same policy judgments apparently forbidden OSHA

117. *See supra* notes 30-31 and accompanying text. I only briefly indicate the nature of this problem, since a thorough discussion will appear in a forthcoming paper by Dr. Helaine Pleet.

118. *See supra* note 92 and accompanying text.

119. In some respects a study that risks a type I error, i.e., showing there is a risk to workers when in fact there is none, may be preferable to the other alternatives. However, it is not clear how credible such studies would be in the eyes of the scientific community.

when it promulgates rules. If epidemiologists must make such policy judgments in designing their studies, why cannot OSHA make them in deciding whether to regulate a substance?

IV

A Proposal For Occupational Health Standard Promulgation

Justice Marshall, writing for the dissent in the *Benzene* case, rejected the "significant health risk" standard and the evidentiary requirements which attend it. In his view, when the "magnitude of risk cannot be quantified on the basis of current techniques,"[120] the Secretary should be permitted to act primarily on policy grounds.[121] Justice Marshall provided few details but his comments suggest the following guidelines. First, the Secretary should act in accord with the remedial purpose of the Act,[122] which is highly averse to risks to workers' health. When risks cannot be quantified, they could be established by expert opinion,[123] akin to the practice of congressional committees. Second, an appropriate margin of error may be used to establish both risks and new exposure standards, subject to the "feasibility" requirement.[124] Finally, policy guidance is especially appropriate for known carcinogens, since any "deficiency in knowledge relates to the extent of the benefits rather than their existence. . . ."[125]

Decisions "to take action in conditions of uncertainty bear little resemblance to the sort of empirically verifiable factual conclusions to which the substantial evidence test is normally applied,"[126] and thus require a higher degree of deference. The spectre of unbridled agency discretion gives cause for concern. There exist, however, additional criteria which can facilitate a more inquisitive judicial review without forsaking worker health. Recognizing the risk-averse nature of the Act, the Court should require OSHA to articulate fully the policy considerations behind the agency's assessment of risks and its determination that more restrictive standards are appropriate. These would include, but would not be restricted to: evidence of carcinogenicity in humans; evidence of carcinogenicity in more than one animal species; and the likelihood that practical problems unrelated to the existence of an actual health risk led to negative results in human studies.[127] Further, when

120. 448 U.S. at 690.
121. *Id.* at 716.
122. *Id.* at 706.
123. *Id.* at 707.
124. *Id.* at 693.
125. *Id.* at 707-08.
126. *Id.* at 706.
127. McGarity, *supra* note 62, at 734.

OSHA relies on biological extrapolation models, it should explain why it prefers one to another.[128] The agency should also set forth and explain the relative risk a particular epidemiological study assumes as well as the power of the test.

Suppose workers challenge an OSHA standard because they believe it offers them insufficient protection. The Court should examine OSHA's reasons for not setting a more stringent standard. It should pay particular attention to "no adverse effect" animal or human epidemiological studies at the current exposure level. As we have seen, numerous practical considerations can undermine the scientific legitimacy of such a "no adverse effect" finding. And, there are good statistical reasons for being more wary of negative results than of positive results. For example, a study may have ignored risk-averse statistical variables. The relative risk considered "safe" may have been set high to cut research costs with the result that a "no observed effect" finding would fail to preclude the existence of a considerable risk. Even if OSHA found a statistically significant risk, and even if it were quantified, OSHA might not have considered it "significant" or one important enough to regulate.[129]

Suppose on the other hand that industry challenges the environmental standard as overly restrictive. Here the Court should look closely at any quantified risks in light of the "substantial evidence" test. If the risks were not quantified, then it would have to look closely at expert testimony and any extrapolation models which OSHA used to judge the risk "significant."

In sum, OSHA should quantify risks when the scientific data justifies doing so, but it should rely on guided policy considerations when such data is not available. Consider two examples. In the ethyleneimine controversy, OSHA found ethyleneimine (EI) was carcinogenic in two animal species.[130] Using extrapolation models the secretary found EI carcinogenic in human beings.[131] The reviewing court determined that the experiments were based upon substantial evidence and concluded that the Secretary had in effect made a legislative choice "that if carcinogenicity in two animal species is established, as a matter

128. Schottenfeld & Haas, *supra* note 73, at 152. The scientific community favors risk averse extrapolation models, which also accord with the risk conservatism of the statute.

129. The Court should scrutinize OSHA's policy reasons for believing such a risk was not monetarily worth protecting against, and its reasons for believing that large numbers of workers would not be affected. In particular, the Court should be especially critical of the benefit information on which OSHA relied as well as the cost information on which the economic "feasibility" of the standards was based, for such benefit and cost information is systematically biased toward industry. The costs of providing devices in the workplace to protect health are readily ascertainable, but the benefits of being free from disease are not so easily discerned.

130. Synthetic Organic Chemical Mfrs. Ass'n v. Brennan, 503 F.2d 1155, 1158 (3d Cir. 1974).

131. *Id.* at 1159.

of law §§ 6(a) and 6(b)(5) require that they be treated as carcinogenic in man."[132] The court held that there was substantial evidence in the record as a whole to support the Secretary's findings.[133]

This decision furnishes a model for allowing OSHA to meet its burden of proof based upon policy considerations. Even the *Benzene* case, which quoted the *EI* decision with approval,[134] might authorize such a model. Yet a strict interpretation of the "significant health risk" model would preclude such a decision.

Consider a second example. Assume there is evidence that a substance is carcinogenic in animals, and carcinogenic in human beings at high exposure levels. But assume further that poor recordkeeping impedes a retrospective cohort epidemiological study, while a case-control study is precluded because no one has yet died from exposure to the substance at levels of the consensus standard. Finally, assume that conservative extrapolation models from both animal studies and from high exposure levels in humans indicate that there is considerable risk to humans at consensus standards. In light of the risk-averse nature of the statute, the evidence of carcinogenicity in animals, and the fact that the appropriate epidemiological studies cannot be done because of poor recordkeeping, the Court should allow OSHA to proceed with rulemaking.

Or assume the same facts with one new embellishment. Posit an epidemiological study showing no risk to humans at consensus levels of exposure. Assume its sample was so small that no significant relative risk could be detected below a value of 9, for a disease with a prevalence in the population of $^{9}/_{10,000}$. With so small a sample, one runs the risk that 72 of every 10,000 workers exposed at the consensus standard will contract the cancer associated with the substance despite the study's apparent optimism. In this situation, OSHA would be justified in promulgating a new standard based primarily on reasons of policy.

V

CONCLUSION

Epidemiology, the chief research tool for discovering risks to human health in the workplace, faces numerous practical and statistical obstacles in detecting such risks. This fact alone should not prevent OSHA from setting higher workplace health standards; although this fact makes it, when scientific facts are uncertain, OSHA could supplement its factual conclusions with policy considerations consistent with the aims and goals of the OSH Act. However, because the Court has

132. *Id.*
133. *Id.* at 1160.
134. 448 U.S. at 657 n.64.

imposed a procedural rule which well may require the quantification of scientific results, in many cases OSHA will be effectively prevented from setting lower exposure standards even when toxic substances pose health risks.

Nevertheless, the Court's options remain open. It can reinforce the strict rule which makes it quite difficult for OSHA to set lower standards and which flies in the face of scientific method and knowledge, or it can permit a more relaxed procedural requirement, along the lines proposed here. A less rigid rule would be more consistent with the OSH Act's legislative purpose and would evidence a proper understanding of scientific methodology and the difficulty of setting exposure standards under conditions of uncertainty. Which choice is followed will have important consequences for the well-being and safety of our workforce.

Appendix

Representative Statistical Values for Prospective Epidemiological Studies of Diseases at Various Incidence Rates

Incidence of disease in general population:		Numbers of Subjects to be Studied N/2 exposed	N/2 unexposed	N total
1) δ =1.5 (relative risk) α =.05 (risk of type I error) β =.20 (risk of type II error)				
a) 8/100	a)	691	691	1,382
b) 8/1,000	b)	7,631	7,631	15,262
c) 8/10,000	c)	77,022	77,022	154,044
d) 8/100,000	d)	770,939	770,939	1,541,878
2) δ =1.5 α =.05 β =.05				
a) 8/100	a)	1,212	1,212	2,424
b) 8/1,000	b)	13,382	13,382	26,764
c) 8/10,000	c)	135,078	135,078	270,156
d) 8/100,000	d)	1,332,037	1,332,037	2,664,074
3) δ =3 α =.05 β =.20				
a) 8/100	a)	62	62	124
b) 8/1,000	b)	756	756	1,512
c) 8/10,000	c)	7,695	7,695	15,390
d) 8/100,000	d)	77,087	77,087	154,174
4) δ =3 α =.05 β =.05				
a) 8/100	a)	109	109	218
b) 8/1,000	b)	1,326	1,326	2,652
c) 8/10,000	c)	13,495	13,495	26,990
d) 8/100,000	d)	135,191	135,191	270,382
5) δ =6 α =.05 β =.20				
a) 8/100	a)	13	13	26
b) 8/1,000	b)	207	207	414
c) 8/10,000	c)	2,150	2,150	4,300
d) 8/100,000	d)	21,580	21,580	43,160

6) $\delta = 6$
$\alpha = .05$
$\beta = .05$

a) 8/100	a)	22	22	44
b) 8/1,000	b)	363	363	726
c) 8/10,000	c)	3,771	3,771	7,542
d) 8/100,000	d)	37,845	37,845	75,690