Screening for Estrogen-mimicking Chemicals: An Assessment of the E-screen and Its Implications

by

Edmond Toy

B.A.S.
Civil Engineering and
Values, Technology, Science, and Society
Stanford University, 1993

Submitted to the Department of Civil and Environmental Engineering in Partial Fulfillment of the Requirements for the Degrees of

Master of Science in Technology and Policy
and
Master of Science in Civil and Environmental Engineering

at the

Massachusetts Institute of Technology
September 1995

Signature of Author: _____
Department of Civil and Environmental Engineering
August 11, 1995

Certified by: _____
David H. Marks, Ph.D.
James Mason Crafts Professor of Civil and Environmental Engineering
Director, Program in Environmental Engineering Education and Research
Thesis Supervisor

Certified by: _____
Richard de Neufville, Ph.D.
Professor of Civil and Environmental Engineering
Chairman, Technology and Policy Program

Accepted by: _____
Joseph M. Sussman, Ph.D.
Chairman, Departmental Committee on Graduate Studies

Screening for Estrogen-mimicking Chemicals: An Assessment of the E-screen and Its Implications

by

Edmond Toy

Submitted to the Department of Civil and Environmental Engineering
on August 11, 1995 in Partial Fulfillment of the Requirements for the Degrees of
Master of Science in Technology and Policy and
Master of Science in Civil and Environmental Engineering

There is a growing concern that chemicals in the environment may be affecting the endocrine systems of wildlife and humans. Some scientists believe that these endocrine disrupting chemicals are responsible for a wide variety of cancer and noncancer effects. Declines in the reproductive success of wildlife have been observed. In humans, declines in sperm quality and increased incidence of breast cancer and cancers of the male and female reproductive tract may be caused by exposure to endocrine disruptors from the time of fetal development through to adulthood. The effects are difficult to detect and not routinely examined by current chemical testing technologies and strategies.

This study examines the development of short term tests to detect endocrine disruptors. Bioassays that are quick, inexpensive, and reliable may play an important role in screening large numbers of chemicals for their endocrine disrupting characteristics. This research examines one particular test that has often been cited as a potentially valuable testing tool: the E-screen. This test identifies only those endocrine disrupting chemicals that mimic the hormone estrogen. While other hormones and other mechanisms of action are also of concern, a focused examination of the E-screen will provide insight into the general development of screening tests for endocrine disrupting chemicals.

This research investigates the factors that are likely to affect how the E-screen will be used and the ways that it can influence government and industry decision making. The scientific basis and the regulatory context of the E-screen are discussed. Insight is gained from interviews with a number of scientists in academia, government, and industry. In addition, a case study is presented of the most widely used short term mutagenicity test for cancer risk assessment, the Ames Test. Scientists and regulators have over two decades of experience with the Ames Test. Although the Ames Test and the E-screen are used to examine different scientific phenomena, the case study provides a framework for an assessment of the E-screen.

This assessment finds that the use and implementation of the E-screen will be affected by the level of public and political concern for endocrine disrupting chemicals, the predictive power of the test, the existence of a testing infrastructure (complementary tests), and the current regulatory framework. The test can fill several but certainly not all niches. At this point, it is premature to adopt the E-screen as part of a formal testing program. Most interviewed scientists believe that considerable research on both a testing infrastructure and the E-screen itself must still be pursued before the E-screen can be used as a basis for government and industry decision making. To facilitate this research, it is recommended that the test should be more widely disseminated; guidelines for its use should be developed; validation studies should be performed by scientists in academia, government, and industry; and validation studies should attempt to find the classes of chemicals for which the E-screen is most predictive.

Thesis Supervisor: David H. Marks, Ph.D.
Title: James Mason Crafts Professor of Civil and Environmental Engineering
Director, Program in Environmental Engineering Education and Research

Acknowledgments

The completion of a thesis typically marks the academic capstone of graduate school. It is difficult for me, however, to think of a suitable capstone for my two years at MIT when I consider the new things I've learned, the people I've met, and the truly incredible experiences I've had. These will surely not end with this thesis. So while I have many people to thank for helping me in this research, my thanks are certainly for much more than what you hold in your hands.

For encouraging me to sail through many uncharted territories, I must thank David Marks. He was always willing to let me do some exploring on my own and also gave me a pull when I needed it. The personal attention that Professor Ana Soto gave to my questions are much appreciated.

I must also thank my parents, who deserve more than can be expressed here. They have supported a son who has done things a little differently. I also thank my sister Lily for her encouragement.

As for my friends at MIT, they have made graduate school a surprisingly pleasant way to spend two years. My experience here would have been less rewarding without Kid Kelly, Jonathan Kleinman, Mort Webster, Narasimha Rao, Alice Yates, Nick Gertler, Alice Caponiti, Mark Stoughton, Tyson Browning, Gail Hickey, Rene Smith, Gooseberry's and TPP hockey. Dale Fried deserves a hearty thanks for his encouragement. Special thanks to Joe Bambenek for the little things that matter and for his lessons in spontaneity. Much credit must be given to the other members of the August Club, Andrew Green and Shanthi Muthiah; I cannot imagine how things would have turned out without them.

And to my constant companions this past year—Andrew Green and Erin O'Neill—I cannot express what wonderful friends you have been. Many roads must still be traveled.

Table of Contents

List of Tables and Figures

Tables

Figures

Abbreviations

ASTM	American Society of Testing and Materials
CCC	Chlorine Chemistry Council
CIIT	Chemical Industry Institute of Toxicology
DDE	1,1-dichloro-2,2-bis(p-chlorophenyl)ethylene
DDT	2,2-bis(p-chlorophenyl)-1,1,1-trichloroethane
DES	Diethylstilbestrol
EPA	Environmental Protection Agency
FIFRA	Federal Insecticide, Fungicide, Rodenticide Act
GAO	General Accounting Office
ICZ	Indole[3,2-b]carbazole
M	Molar
NRC	National Research Council
OSHA	Occupational Safety and Health Administration
PCB	Polychlorinated biphenyls
PMN	Premanufacturing notices
RPE	Relative proliferative effect
RPP	Relative proliferative potency
SAR	Structure-activity relationships
TSCA	Toxic Substance Control Act

Chapter One

Introduction

Tracing the history of environmental issues in the United States, certain icons have come to symbolize the struggle to preserve human health and the environment. Rachel Carlson painted a grim picture of a silent spring, the result of chemical pollution that was destroying bird populations. Also familiar symbols are the smokestack spewing clouds of acrid smoke into the air, the pipe that brings forth to rivers sewage and industrial wastes, and the leaking drums of toxic waste of Love Canal.

These problems led to the establishment of a number of important environmental policies to regulate pesticides, emissions into the air or water, and toxic chemicals. As the 21th century approaches, an emerging issue is poised to become a focus of environmental concern that could catalyze another round of statutory and regulatory actions. The concern is the possibility that chemicals, both synthetic and natural, may have the ability to alter the endocrine systems of wildlife and humans. Some scientists and policy makers believe that these chemicals—endocrine disruptors—may be responsible for a number of adverse effects that are not easily detectable.

Warning signs and concerns

Although endocrine disruption has been studied for many years, it has taken on a new urgency recently. In 1993, Congress convened a hearing to gather evidence from several key scientists in the field.[1] Much of the focus was on chemicals that disrupt the processes involving the hormone estrogen. Despite being known as the female hormone, estrogen is produced by both

---

[1] US Congress (1993). "Health Effects of Estrogenic Pesticides," hearing before the Subcommittee on Health and Environment o the Committee on Energy and Commerce, House of Representatives, October 21.

males and females and is critical in the development of the fetus and sexual maturation of both genders. Congressman Waxman emphasized that agricultural and industrial chemicals believed to mimic the effects of endogenous estrogen have been linked to "very serious health and environmental impacts."[2]

Much of the hearing was devoted to discussion of recent research on the possible links between hormonally active chemicals and the growing number of breast cancer cases. While it is generally believed that estrogen is a factor in causing breast cancer, it is uncertain whether estrogen-mimicking chemicals are increasing that risk. Jane Reese-Coulbourne of the National Breast Cancer Coalition stated that breast cancer is the most common form of cancer among women. In 1960, one out of fourteen women could expect to be diagnosed with breast cancer during her lifetime. Now, the odds are one in eight. Environmental factors, she argued, have not been adequately studied.[3] Provocative evidence of the effects of estrogen-mimicking chemicals was presented at this hearing. Mary Wolff described her research that found that breast cancer incidence was correlated with exposure to DDE, the main metabolite of the pesticide DDT.[4]

The effects of endocrine disruptors have also been observed in males. Louis Guillette of the University of Florida testified to the subcommittee about the effects that he has observed in wildlife. He has been investigating the waning reproductive success of alligators in Florida's Lake Apopka, which was the site of a major pesticide spill. Guillette found that hatching success has plummeted and the male alligators that survive have abnormal testes and abnormally small penises.[5]

Another alarming piece of evidence is the belief that human sperm quality has been declining for several decades. Some scientists hypothesize that male fetuses are exposed, via the mother, to endocrine disrupting chemicals that alter differentiation of the developing testes. Guillette pointed out that reports of declining sperm quality show that "every man...today is half

---

2 US Congress (1993). p. 1.
3 US Congress (1993). pp. 175-177.
4 US Congress (1993). pp. 16-18.
5 US Congress (1993). pp. 39-41.

the man his grandfather was, and the question is, are our children going to be half the men we are...."[6]

Developmental effects are particularly worrisome, argued Theo Colborn of the World Wildlife Fund. A parent exposed to endocrine disruptors may not show any signs of adverse effects, but its offspring may suffer the consequences. Often, the effects on the offspring are not detectable at birth and are observed only in the later functioning of the offspring. In addition, a small dose at a critical stage of pregnancy may be enough to alter development. Colborn stated that a mother's exposure to endocrine disruptors "could have more control over how their babies develop than the genes their children inherit from them and the training they give their children."[7]

Congressman Waxman stated that "perhaps the most important" finding of the hearing was that the risk posed by pesticides acting as endocrine disruptors is unknown: "EPA does not routinely evaluate pesticides for dangerous hormonal effects."[8] Ana Soto of Tufts University presented to the subcommittee her work to develop a test—the E-screen—that can identify estrogenic chemicals. She noted that in the past, a number of pesticides and industrial chemicals have been used and released into the environment before it was known whether they might be harmful. She told the subcommittee that this problem should be avoided; new chemicals should be tested before they are used. She stated that "simple, inexpensive, and reliable tests such as the E-screen assay should be used to check for hormonal effects before chemicals are released into the environment."[9]

Lynn Goldman, EPA's Assistant Administrator for the Office of Prevention, Pesticides, and Toxic Substances, told the subcommittee that the agency is attempting to "develop testing requirements to ensure that the public is protected from hormonal effects of pesticides." She continued:

> This [task] is not easy. Prior to issuing new test rules we need to be able to assure the public that the tests are predictive and scientifically reproducible and valid. We have had experiences in the past that are sobering. For example, in vitro testing for mutagenicity, that is genetic changes and bacteria were not adequate to predict cancer. Much evaluation of any in vitro tests would be needed along with animal

6 US Congress (1993). p. 57.
7 US Congress (1993). pp. 38-39.
8 US Congress (1993). p. 1.
9 US Congress (1993). pp. 19-20.

> testing to adopt any one test as a screening test of choice, but although this is not easy, the EPA will...move the science from the state of the art to routine testing. In the long run, we will probably have a strategy involving both short term assays and whole animal tests.[10]

The role of screening tests

The testimony given to Congress and the growing scientific literature on endocrine disruption suggest that chemicals should be routinely tested for their ability to alter hormonal functioning. Currently there are a number of tests being used for this purpose, and others are being developed. Some rely on whole animals, while others use cultured cells from animals or humans. The tests measure different endpoints. The variety of and uncertainty in the mechanisms by which endocrine disruptors operate make difficult their identification.

Short term tests could be of considerable advantage in chemical testing. Soto told Congress that "available animal tests to detect estrogenicity are time consuming and expensive to perform for mass screening." So far, Soto's short term assay, the E-screen, has confirmed the estrogenic abilities of chemicals previously shown to be estrogenic in animal tests.[11]

The research presented here examines the role of short term tests for the identification of endocrine disruptors. We investigate whether and how short term tests should be incorporated into strategies for the routine testing of chemicals. We will attempt to anticipate the important factors that will affect the development and use of short term tests. In addition, we will examine the ways in which short term tests will influence regulatory decisions on chemical testing, development, and control. Taking a broad context here, regulatory decisions include those decisions made by companies complying with or anticipating regulation.

In particular, the short term test that this thesis examines is the E-screen. There are several key questions that arise. At this point of development, will the E-screen be incorporated into regulatory decision making? Should it be incorporated? Is there enough concern for estrogen-mimicking chemicals to warrant the use of the E-screen? How will the test be validated? Against what standards? Under what circumstances will the E-

---

10 US Congress (1993). p. 67.
11 US Congress (1993). p. 19.

screen be influential? Importantly, we approach these questions from a variety of perspectives. We will examine the scientific, policy, and societal issues that form the context of environmental decision making.

Some of these issues were mentioned in Lynn Goldman's testimony to Congress. Of primary concern are the reliability and validity of short term tests that are candidates for use in testing strategies. She pointed to the experience of short term mutagenicity tests that were developed for cancer risk assessment. The connection is clear. Goldman implies that decisions regarding short term tests for endocrine disruptors should not repeat the mistakes of history. This thesis presents a case study of the Ames Test, the most widely used short term mutagenicity test for cancer risk assessment. The case study will serve as a framework for assessing the pitfalls and promises that are likely to be encountered in the use of screens for endocrine disruptors.

The E-screen is singled out in this research because it has been cited in the press and in the regulatory arena as one of the first short term tests for estrogenic chemicals. It is important to note that only a portion of the larger issue of endocrine disruption is being discussed here. Estrogen-mimicking chemicals are only one type of chemical thought to disrupt the endocrine system, and they may be responsible for only some of the observed effects. Although the E-screen tests only for estrogenic chemicals, the test will likely be a prototype for other short term tests in terms of their development, use, and influence. It must be emphasized here that the E-screen is used as a vehicle for discussing more general concepts for screening endocrine disruptors. Close scrutiny of the E-screen will provide lessons about screening tests in general.

Research methods

This thesis relies on evidence of several kinds. The first is the literature. Scientific literature and government documents were examined for both the case study of the Ames Test and the discussion of endocrine disruption. Examples include technical articles from scientific journals such as *Environmental Health Perspectives*, *Nature*, and the *Journal of Steroid Biochemistry*. EPA documents, the *Federal Register*, and other regulatory

documents were used. In addition, the Ames Test case study relies on political science literature to trace the sociopolitical history of the test.

The goals of this thesis, however, require the use of other types of evidence. Telephone interviews, primarily with scientists, were vital to the research.[12] In total, 25 interviews were conducted with people from EPA, industry, and academia. For the Ames Test case study, official EPA documents provide general information, but to gain insight into the details of decision making, interviews were required. The use of interviews was even more critical to understanding the decision making process within industrial laboratories.

Because the use of short term tests to screen for endocrine disruptors is a relatively new topic, the literature on their regulatory impacts is sparse. Meetings and conferences on endocrine disruption, in addition to a number of personal interviews, form a large part of the research presented here. Since 1993, the Massachusetts Institute of Technology has been engaged in a major interdisciplinary research project on the Management of the Future Uses of Chlorine. This project sponsored in November 1994, a symposium that brought together experts from academia, government, industry, and environmental advocacy groups. Endocrine disruption was a major topic of discussion. In addition, many insightful points and interviews came from a major meeting sponsored by the EPA's Health Effects Research Laboratory. At this Endocrine Disruptors Research Needs Workshop, hundreds of scientists gathered to discuss the cutting edge in research and how it could affect regulatory decisions.

In the next chapter, background is provided on endocrine disruption. We discuss some of the key concepts that make this issue problematic, and we briefly review some of the research into the effects of endocrine disruptors. In Chapter 3, we begin our examination of the E-screen. The development, basis, and early results of the test are presented. In Chapter 4, we turn to the case study on the Ames Test. Using the case study as a framework for analysis, we return to the E-screen in Chapter 5. We attempt to apply lessons learned

[12] In order to obtain candid information about decision making, interviews were often conducted with the understanding that persons interviewed would not be identified. For this reason, the names of the persons interviewed, and the firms or organizations at which they worked, are sometimes not revealed in the text.

from the experience of the Ames Test to the E-screen and other short term tests. Finally, we conclude in Chapter 6, where the main findings of this research are summarized and recommendations are made regarding the development, use, and influence of the E-screen and short term screening tests in general.

Chapter Two

Endocrine Disruptors in the Environment

The endocrine system is a complex orchestration of chemical signaling and gene expression that is responsible for critical functions from the time of conception, through development, and through maturation. The possibility that environmental chemicals may disrupt the normal processes of the endocrine system is a provocative issue. Endocrine disruption encompasses a wide variety of chemical causes, mechanisms of action, and effects. The science of endocrine disruption is fertile ground for new research. In this chapter, we briefly review endocrine disruption. First, the terms and mechanisms of endocrine disruption are discussed. Second, evidence of the potential impacts of endocrine disruptors is presented.

Endocrine Disruption

Hormones in the body are chemicals secreted by glands of the endocrine system. Hormones are transported throughout the body in the bloodstream. They interact, however, only with certain organs. The cells in these target organs contain molecules known as receptors, and hormones exert their effects by binding to receptors in target organ. The hormone-receptor complex generally alters the expression of genes related to cell function, growth, and differentiation.[1]

Of particular importance here are the male and female sex hormones, androgens and estrogens, respectively. Both types of hormones are produced by both sexes, although in differing proportions. Androgens include hormones

---

[1] Clark, James, William Schrader, and Bert O'Malley (1985). "Mechanisms of Steroid Hormone Action," in *Textbook of Endocrinology*, Jean Wilson and Daniel Foster, eds., Philadelphia: W.B. Saunders Company, p. 33.

such as testosterone. Target organs of androgens include the prostate and seminal vesicles, influencing the production of spermatozoa.

Estrogens are hormones that exert a wide variety of effects, most notably those dealing with the estrus cycle and female sex characteristics. Endogenous estrogens that are normally produced by the body include estradiol, estrone, and estriol, while synthetic estrogens include diethylstilbestrol and ethinyl estradiol. Endogenous estrogens are produced primarily by the ovaries but also by the placenta, the adrenal glands, and the testis. In addition, estrogens, known as phytoestrogens, are made naturally in some plants.[2]

Estrogen has several target organs, including the uterus. Estrogen leads to an increase in both the size and number (mitotic activity) of uterine cells. This uterotropic response can be observed in as little as four to six hours after the application of estrogen to an animal.[3] Another target organ of estrogen is the vagina. Estrogen causes proliferation of vaginal epithelium cells and the addition of several epithelial cell layers. Eventually, the uppermost layer will undergo cornification, which is a compression and hardening of cells.[4]

It is well known that exogenous chemicals can be applied to organisms to affect the functioning of the endocrine system. In some instances, this knowledge has allowed scientists to design therapeutic hormonal agents. Recently, however, concern has grown for the possibility that exogenous chemicals are inadvertently disrupting the endocrine system. Much of the focus has been on estrogenic chemicals. In 1979, the National Institute of Environmental Health Sciences sponsored its first symposium on Estrogens in the Environment.[5] The influences of estrogen on development was the focus of the second Estrogens in the Environment symposium in 1985,[6] while the third symposium in 1994 focused on global health implications. The issue of estrogens and other hormones in the environment became an urgent topic in

---

2 Paschkis, Karl, Abraham Rakoff, Abraham Cantarow, and Joseph Rupp (1967). *Clinical Endocrinology*, New York: Hoeber Medical Division of Harper & Row, pp. 534-558.

3 Clark et al. (1985). p. 55.

4 Paschkis et al. (1967). pp. 547-549.

5 McLachlan, John, ed. (1980). *Estrogens in the Environment*, New York: Elsevier Scientific Publishing.

6 McLachlan, John, ed. (1985). *Estrogens in the Environment II: Influences on Development*, New York: Elsevier Scientific Publishing.

the early 1990s. One important factor for this was the 1991 Wingspread Conference organized by Theo Colborn.[7] The conference proceedings were so important that Bern called it the "bible" of the endocrine disruption issue.

The Wingspread Conference brought together experts in anthropology, ecology, comparative endocrinology, histopathology, immunology, mammalogy, medicine, law, psychiatry, psychoneuroendocrinology, reproductive physiology, toxicology, wildlife management, tumor biology, and zoology. The scientists wrote a consensus document containing a provocative statement: "We are certain [that a] large number of man-made chemicals that have been released into the environment, as well as a few natural ones, have the potential to disrupt the endocrine system of animals, including humans."[8]

It is believed that endocrine disruption may be caused by chemicals acting in various ways.[9] First, endocrine disruptors may mimic the effects of endogenous hormones such as estrogen and androgen. Second, they may inhibit the effects of endogenous hormones. Third, they may change the normal synthesis and metabolism of hormones. Fourth, they may alter the quantity of hormone receptors.

The first two routes by which endocrine disruptors may work are the focus of much research. The theoretical basis for endocrine disruption is "relatively straightforward."[10] A lock and key analogy is often used. A receptor can be thought of as a lock. The intended hormone acts as a key, and the binding of the hormone to the receptor activates gene expression. The receptor will not, however, have perfect specificity; that is, foreign chemicals can bind to the receptor in place of the intended hormone and exert different effects. These chemicals can either mimic the effects of the endogenous hormone, or they can block the normal effects resulting from receptor binding. Hormone mimics, or *agonists*, increase hormonal activity, while hormone blocks, or *antagonists* (e.g., anti-estrogens and anti-androgens), reduce hormonal activity.

---

[7] Colborn, Theo and Coralie Clement, eds. (1992). *Chemically-Induced Alterations in Sexual and Functional Development: The Wildlife/Human Connection*, Princeton, New Jersey: Princeton Scientific Publishing.

[8] Colborn and Clement (1992). "Consensus Statement," p. 1.

[9] Soto, Ana, Carlos Sonnenschein, Kerrie Chung, Mariana Fernandez, Nicolas Olea, and Fatima Olea Serrano (in press). "The E-screen Assay as a Tool to Identify Estrogens: An Update on Estrogenic Environmental Pollutants," *Environmental Health Perspectives.*

[10] McLachlan, John (1993). "Functional Toxicology: A New Approach to Detect Biologically Active Xenobiotics," *Environmental Health Perspectives*, Vol. 101, No. 5, p. 386.

This description of endocrine disruption oversimplifies what are very complex and uncertain processes. There are several reasons why the identification of endocrine disruptors is so difficult. For example, chemicals can be metabolized into forms with more or less hormonally activity. Identification of endocrine disrupting chemicals is made difficult because laboratory tests may not be able to incorporate metabolic considerations. For example, the chemical methoxychlor is believed not to be estrogenic itself, but its metabolites are estrogenic.[11]

Recent research of Kelce et al. on the pesticide DDT and its metabolites illustrates the difficulty in identifying and characterizing endocrine disruptors.[12] There is considerable uncertainty in what scientists should be looking for because there are a variety of different mechanisms of action. Abnormalities observed in male reproductive health of wildlife and humans have coincided with exposure to estrogenic chemicals such as DDT. It has been postulated for several decades that DDT binds to the estrogen receptor and induces estrogenic effects in female rats. Although the effects of DDT in males are believed to be mediated through the estrogen receptor, the effects are also consistent with mechanisms involving the androgen receptor. Kelce et al. recently found that *p,p'*-DDE, the major metabolite of DDT, "has little ability to bind the estrogen receptor, but inhibits androgen binding to the androgen receptor, androgen-induced transcriptional activity, and androgen action...."

Richard Sharpe noted the significance of this finding.[13] Chemicals that are estrogenic ("feminizing") or anti-androgenic ("demasculinizing") can manifest themselves in the same way, but they work through the estrogen and androgen receptors, respectively. The work of Kelce et al. "means that 'DDT' has the capacity to act with both receptors." Sharpe points out that the most "remarkable aspect" of this research is that it has taken 50 years for scientists to discover the anti-androgenic abilities of DDT's metabolites. This fact is a reflection of the inadequacy of current methods of testing.

---

11 Soto et al. (in press).

12 Kelce, William, Christy Stone, Susan Laws, Earl Gray, Jon Kemppalnen, and Elizabeth Wilson (1995). "Persistent DDT Metabolite p,p'-DDE is a Potent Androgen Receptor Antagonist," *Nature*, Vol. 375, June 15, pp. 581-585.

13 Sharpe, Richard (1995). "Another DDT Connection," *Nature*, Vol. 375, June 15, pp. 538-539.

Another aspect of the problem is that endocrine disruptors can cause their effects at critical periods in the development of the fetus. The timing of the dose is of considerable importance. Endocrine disruptors may have different effects on the embryo and fetus than on the adult.[14]

In addition, there is the important issue that the definition of endocrine disruption is not entirely clear. What effect or endpoint should scientists be looking for when investigating endocrine disruptors? Take for example the hormone estrogen. According to Hertz, it can cause morphological, physiological, biochemical, and behavioral effects.[15] Estrogen has as its primary effect "the stimulation of mitotic activity in the tissues of the female genital tract." Other effects include "increased vascularity of the affected tissue, alteration of motility of affected muscular elements, alteration of fluid balance and of lipid and calcium metabolism....effects on the mammary glands...and differentiation of the central nervous system." It is not surprising that assays have been developed to measure different effects of estrogenic chemicals. It is not clear which effect is most appropriate as an endpoint to measure. Nevertheless, Hertz argues that "the sine qua non of estrogenic activity remains the mitotic stimulation of the tissues of the female genital tract. A substance which can directly elicit this response is an estrogen; one that cannot do this is not an estrogen." It is not clear, however, that the estrogenic activity—defined in this fashion—is ultimately responsible for the effects seen in wildlife and humans.

Evidence of effects

Complementing this evidence at the molecular level and from laboratory animals, there is a substantial body of literature on the effects of endocrine disruptors found in the environment. Table 2.1 lists chemicals with widespread distribution in the environment that have been identified as endocrine disruptors.

Correlations have been found between exposure to endocrine disruptors in the environment and effects in wildlife. It has been known for a number of

---

[14] Colborn and Clement (1992). p. 2.

[15] Hertz, Roy (1985). "The Estrogen Problem: Retrospect and Prospect," in *Estrogens in the Environment II: Influences on Development*, John McLachlan ,ed., New York: Elsevier Science Publishing, p. 1.

Table 2.1 Endocrine disrupting chemicals

Herbicides	Insecticides	Industrial chemicals
2,4,-D	Beta-HCH	Cadmium
2,4,5,-T	Carbaryl	Dioxin (2,3,7,8-TCDD)
Alachlor	Chlordane	Lead
Amitrole	Dicofol	Mercury
Atrazine	Dieldrin	PBBs
Metribuzin	DDT and metabolites	PCBs
Nitrofen	Endosulfan	Pentachlorophenol
Trufluralin	Heptachlor and epoxide	Penta- to nonylphenols
	Lindane	Phthalates
Fungicides	Methomyl	Styrenes
Benomyl	Methoxychlor	
Hexachlorobenzene	Mirex	
Mancozeb	Oxychlordane	Nematocides
Maneb	Parathion	Aldicarb
Metiram-complex	Pyrethroids	DBCP
Tributyl tin	Toxaphene	
Zineb	Transnonachlor	
Ziram		

Source: NRC Commission on Life Sciences (1995).

years that environmental concentrations of DDT can cause eggshell thinning and cracking to the extent of threatening the stability of bird populations.[16] In Florida's Lake Apopka, the site of a pesticide spill, the alligator population has been found to have declined.[17] Only 18 percent of laid eggs will hatch, and only half of those survive longer than 10 days. In other Florida lakes, the hatching success is around 90 percent. Guillette hypothesizes that DDE and perhaps other chemicals are acting as estrogens and causing these effects.

16 Colborn, Theo, Frederick vom Saal, and Ana Soto (1993). "Developmental Effects of Endocrine-Disrupting Chemicals in Wildlife and Humans," *Environmental Health Perspectives*, Vol. 101, No. 5, October, p. 379.

17 US Congress (1993). "Health Effects of Estrogenic Pesticides," hearing before the House of Representatives, Committee on Energy and Commerce, Subcommittee on Health and Environment, October 21, pp. 39-41.

A wide variety of other effects in wildlife have been documented.[18] These include abnormal thyroid function in birds and fish; decreased fertility in fish, shellfish, and mammals; decreased hatching success in fish, birds, and turtles; gender abnormalities (the masculinization and defeminization of females and/or the feminization and demasculinization in males) in fish, gastropods, birds, and mammals; and immunological effects in birds and mammals.

Evidence of effects in humans has also been found. One of the most dramatic examples of the potential human effects of estrogen-mimicking chemicals is the experience of diethylstilbestrol (DES), a synthetic estrogen.[19] For more than 20 years, DES was given to pregnant women to prevent premature birth and spontaneous abortion until the drug was banned for this use. Two to three million pregnant mothers are estimated to have taken this drug.[20] Daughters of women who took DES suffer from a wide variety of effects, including reproductive organ dysfunction, abnormal pregnancies, reduced fertility, deficient immune systems, and depression. DES-daughters also show increased rates of vaginal clear-cell adenocarcinomas.[21]

The DES experience suggests that estrogenic chemicals in the environment may have similar effects. Colborn et al. believe that estrogenic chemicals may have a role in the etiologies of a number of adverse health trends.[22] For example: there as been a 400 percent increase in ectopic pregnancies in the US between 1970 and 1987; incidence of cryptorchidism (undescended testes) in the United Kingdom has doubled during the same period; and an increase in prostate cancers was reported in the US between 1969 and 1986. Two of the most controversial and well-publicized health trends have been the increase in breast cancer and the decrease in sperm count.

Several studies have examined the link between breast cancer and endocrine disruptors. In 1993, Wolff et al. published a major study that focused on the effects of exposure to PCBs and DDE, the main metabolite of

---

18 Colborn et al. (1993). pp. 378-379.

19 Colborn et al. (1993). pp. 379-380.

20 Bern, Howard (1992). "The Fragile Fetus," in *Chemically-Induced Alterations in Sexual and Functional Development: The Wildlife/Human Connection*, Colborn and Clement, eds., Princeton, New Jersey: Princeton Scientific Publishing, pp. 9-10.

21 Colborn et al. (1993). pp. 379-380.

22 Colborn et al. (1993). pp. 380.

DDT.[23] From a population of 14,290 women in the New York University Women's Health Study from 1985 to 1991, they examined the sera from stored blood samples of 58 women with breast cancer and 171 matched control subjects . After adjusting for family history of breast cancer, lifetime lactation, and age at first full-term pregnancy, regression analysis showed a fourfold increase in breast cancer risk for women with DDE concentrations in the 90th percentile compared to women with DDE concentrations in the 10th percentile. For PCBs, the relative risk for women in the 90th percentile for serum PCB concentration was less than a twofold increase compared to women in the 10th percentile; this was a nonsignificant association. They concluded that breast cancer was strongly associated with DDE in serum but not with PCBs.

Shortly after this study, Davis et al. argued that established risk factors for breast cancer accounted for at most 30 percent of cases, and that a variety of estrogenic chemicals in the environment may account for recent trends in breast cancer.[24] In addition, they state that the risk depends on how estrogens are metabolized in the body. For example, some phytoestrogens can reduce cancer risk.

A more recent study on breast cancer calls into question these findings. Krieger et al. reported in 1994 that their epidemiological study found no relationship between breast cancer and blood serum levels of PCBs or DDE. Their sample population was 57,040 women; they examined the sera of 150 women with breast cancer and 150 control subjects. In addition to being a larger study than that of Wolff et al., this study considered the effects of race (also an indicator of socioeconomic status). Furthermore, this study focused on women with exposures prior to the federal restrictions on DDT. Consequently, these women had higher levels of DDE than in the Wolff et al. study, yet no increased breast cancer risk was observed.[25]

Evidence of the link between endocrine disruptors and decreases in sperm count is even less certain, but no less controversial. There is considerable debate about whether there has even been any observed changes

---

23 Wolff, Mary, Paolo Toniolo, Eric Lee, Marilyn Rivera, and Neil Dubin (1993). "Blood Levels of Organochlorine Residues and Risk of Breast Cancer," *Journal of the National Cancer Institute*, Vol. 85, No. 8, April 21, pp. 648-652.

24 Davis, Devra Lee, H. Leon Bradlow, Mary Wolff, Tracey Woodruff, David Hoel, and Hoda Anton-Culver (1993). "Medical Hypothesis: Xenoestrogens as Preventable Causes of Breast Cancer," *Environmental Health Perspectives*, Vol. 101, No. 5, October, pp. 372-377.

25 MacMahon, Brian (1994). "Pesticide Residues and Breast Cancer?" *Journal of the National Cancer Institute*, Vol. 86, No. 8, April 20, p. 572.

in sperm quality. Carlsen et al. reported in 1992 that their review of semen quality studies conducted between 1938 and 1991 showed a 42 percent drop in mean sperm counts.[26] The 61 studies included data on 14,947 men. Auger et al. reported corroborating evidence in 1995.[27] In the so-called Paris study, Auger et al. examined semen collected from 1351 men at a single sperm bank in Paris during the period between 1973 and 1992. They found that sperm counts declined independently of age.

The conclusions of these studies have been criticized. The study by Carlsen et al. has been questioned for biases in the studies they reviewed and the patients that were selected in each study. Bromwich et al. argue that observed declines can be explained by changing definitions of what is a "normal" sperm count, which may have resulted in the exclusion of some "normal" men from the study.[28]

The cause of the supposed decrease in sperm quality is still unknown, although endocrine disruptors are suspected as the cause.[29] It has been shown that alterations to the endocrine environment of the developing rat will result in changes in the number of Sertoli cells, the cells responsible for spermatogenesis. It is unclear, however, whether endocrine disruptors found in the environment can cause these effects in humans.

Scientific research to date has only scratched the surface of the topic of endocrine disruption. Scientists have some understanding of some of the possible mechanisms of action of endocrine disruption, but many of the details are still elusive. Effects can be mediated through a number of different mechanisms; even the well-studied pesticide DDT continues to surprise scientists in the complexity of ways that it can affect the endocrine system. Evidence from laboratory subjects, wildlife and DES-exposed humans suggests that endocrine disruptors in the environment can have deleterious impacts.

---

26 Carlsen, Elisabeth, Aleksander Giwercman, Niels Keiding, Niels Skakkebaek (1992). "Evidence for Decreasing Quality of Semen During the Past 50 Years," *British Medical Journal*, Vol. 305, September 12, pp. 609-613.

27 Auger, Jacques, Jean Marie Kunstmann, Francoise Czyglik, and Pierre Jouannet (1995). "Decline in Semen Quality Among Fertile Men in Paris During the Past Twenty Years," *New England Journal of Medicine*, Vol. 332, No. 5, February 2, pp. 281-285.

28 Bromwich, Peter, Jack Cohen, Ian Stewart, and Andrew Walker (1994). "Decline in Sperm Counts: An Artefact of Changed Reference Range of 'Normal'?" *British Medical Journal*, Vol. 309, July 2, pp. 19-22.

29 Sharpe, RM (1993). "Declining Sperm Counts in Men: Is There an Endocrine Cause?" *Journal of Endocrinology*, Vol. 136, pp. 357-360.

Studies of incidence, exposure, and the mechanisms involved in breast cancer and sperm quality continue to be pursued and debated.

Despite, or perhaps because of, the uncertainty in endocrine disruption, some scientists and policy makers have argued that chemicals should be tested for their endocrine disrupting capabilities. In the next chapter, we will examine one technology for identifying one kind of endocrine disruptor. The E-screen has shown some promise in identifying estrogen-mimicking chemicals. Whether the E-screen or other short term tests should be used will be the topic in the following chapters.

Chapter Three

An In Vitro Test for Estrogen-mimicking Chemicals: The E-screen

Characterizing the causes of the diverse set of effects discussed in the previous chapter will challenge scientists from many different disciplines. Our understanding of the causes, effects, and mechanisms is evolving quickly. It is clear that terms such as "estrogen-mimicking chemicals" do not encompass the whole set of observed and speculated phenomena. More inclusive terms such as "endocrine disruption" reflect a more complete understanding that the problem extends beyond a single hormone. Theo Colborn, in many respects the leader and catalyst in the growing concern about these phenomena, believes that even the term "disruption" is not adequate to describe the complexity and the ambiguity involved. She states that a more accurate term is "modulation." This term recognizes that the phenomena is not the result only of overt disruption of the endocrine system, but of more subtle changes in endocrine system signaling, the effects of which are difficult to predict and detect.[1]

It is fair to say, however, that scientists, regulators, and the public have focused their attention on the estrogenic properties of chemicals. Estrogen serves as the prototype for the larger issue of endocrine disruption. In the rest of this thesis, our attention will focus on estrogenic chemicals. While there are other important facets to endocrine disruption, estrogenicity is the most well developed portion in our understanding of endocrine disruption.

As we have seen, it has been suggested that the identification of estrogen-mimicking chemicals could be an important step in controlling risks to human health and the environment. In this chapter, we begin an in-depth examination of the E-screen, a short term in vitro bioassay being developed by Professor Ana Soto at Tufts University and her colleagues. First we will briefly

---

[1] Colborn, Theo (1995). Address to the Third Plenary Session of the EPA's Endocrine Disruptors Research Needs Workshop, April 10-13, Raleigh, North Carolina.

review some of the other methods and technologies that are being used to identify estrogen-mimicking chemicals. The E-screen will then be described in some detail.

Identification of estrogen-mimicking chemicals

The E-screen is just one of many methods that have been or are being used to identify estrogen-mimicking chemicals. One type of test uses whole organisms, typically rodents. In rodent bioassays, measurement of increases in uterotropic activity in the uterus and the presence of vaginal cornification are often used. However, increases in uterotropic activity, as indicated by uterine weight, is not a specific estrogen response; that is, factors besides estrogenic chemicals can cause such effects. It has been argued that rodent bioassays are not suitable for large-scale screening.[2]

In vitro bioassays have also been developed to identify estrogenic chemicals. One major type of bioassay looks at the degree to which a chemical will bind to the estrogen receptor. Although binding to the estrogen receptor may be required for estrogen-like action, expression of estrogen-controlled genes may or may not be induced. Consequently, this type of test may not provide information about whether a chemical that binds to the estrogen receptor is estrogenic (agonistic) or anti-estrogenic (antagonistic). The determination of whether a chemical is estrogenic or anti-estrogenic may be made using a second type of in vitro bioassay, one that uses a reporter gene. In this type of bioassay, human or animal cells are transfected with an estrogen-responsive reporter gene. This bioassay measures the expression of this reporter gene; the expression may be the production of a certain type of protein.[3] Production of this protein signals estrogen agonism. The reliability of bioassays using reporter genes has been questioned because gene expression may occur in the absence of estrogen.[4]

---

[2] Soto, Ana, Carlos Sonnenschein, Kerrie Chung, Mariana Fernandez, Nicolas Olea, and Fatima Olea Serrano (in press). "The E-screen Assay as a Tool to Identify Estrogens: An Update on Estrogenic Environmental Pollutants," *Environmental Health Perspectives.*

[3] Mayr, Ulrich, Annette Butsch, and Susanne Schneider (1992). "Validation of Two In Vitro Test Systems for Estrogenic Activities with Zearalenone, Phytoestrogens and Cereal Extracts," *Toxicology*, Vol. 74, p. 146.

[4] Soto et al. (in press).

The E-screen is a different type of in vitro test. It relies on the use of a human cell line that proliferates in the presence of estrogen. Soto et al. argue that the E-screen is "biologically equivalent to the increase of mitotic activity in the rodent endometrium." As discussed in the previous chapter, it has been argued that mitotic activity in the female genital tract is the hallmark of estrogenic activity. If this mitotic activity is the fundamental basis for estrogens, then the E-screen may provide more directly relevant information than the estrogen receptor binding assay and the reporter gene assays.

The E-screen

The E-screen has several advantages over other tests.[5] First, the E-screen uses human cells and serum. Tests based on animals or their cells are less applicable to humans because estrogen-mimicking chemicals may have different effects in different species. This problem is obviated in the E-screen. Second, the E-screen is highly sensitive. The E-screen can detect the equivalent of estradiol at a concentration of $3x10^{-11}$ M. Third, the test is simple to perform. A standard cell culture facility is adequate. Furthermore, results can be obtained quickly, in about 6 days.

Given these characteristics and the early results of the test, the E-screen could be a valuable tool. The E-screen has been discussed in scientific, regulatory, and industrial circles, as well as in the popular press. The E-screen was one of several tests that was highlighted at the third Estrogens in the Environment Conference sponsored by the National Institute of Environmental Health Sciences. The test was also discussed in a *Washington Post* feature on environmental estrogens.[6] And in *Chemical & Engineering News*, Soto was quoted as saying that the E-screen "should be adopted by industry and regulatory agencies to provide an inexpensive, rigorous screen for estrogenic compounds."[7]

---

5 Soto, Ana, Tien-Min Lin, Honorato Justicia, Renee Silvia, and Carlos Sonnenschein (1992a). "An 'In Culture' Bioassay to Assess the Estrogenicity of Xenobiotics (E-Screen)," in *Chemically-Induced Alterations in Sexual and Functional Development: The Wildlife/Human Connection*, Colborn and Clement, eds., Princeton, New Jersey: Princeton Scientific Publishing, pp. 305-307.

6 Weiss, Rick (1994). "Estrogen in the Environment: Are Some Pollutants a Threat to Fertility?" *The Washington Post*, January 25.

7 Hileman, Bette (1994). "Environmental Estrogens Linked to Reproductive Abnormalities, Cancer," *Chemical & Engineering News*, January 31, p. 23.

Basis and methodology

The E-screen developed out of the work that Soto and her laboratory at Tufts University conducted on the mechanisms of estrogen action on cell proliferation. Although it was well-known that application of estrogen to rodents with their ovaries removed would cause mitotic activity in the uterus, the mechanisms through which this cell proliferation occurred was poorly understood.

Soto et al. tested several different hypotheses of the mechanisms of estrogen.[8] According to the first hypothesis, the direct-positive hypothesis, estrogen itself triggers cell proliferation. Second, the indirect-positive hypothesis attributes cell proliferation to growth factors, the release of which are caused by estrogen. Finally, there is the indirect-negative hypothesis, which states that estrogen neutralizes the action of proliferation inhibitors found in the blood.

In their search for this inhibitor, Soto et al. used in their experiments MCF7 human breast tumor cells.[9] These cells have been observed to produce estrogen-sensitive tumors in athymic mice. In in vitro experiments, MCF7 cells placed in medium proliferate maximally, regardless of the presence of estrogen. This proliferation is inhibited when charcoal-dextran stripped human serum (CDHuS) is added to the medium; charcoal-dextran stripping removes sex steroids from the human serum. Soto et al. also observed that this inhibition is released in a dose-dependent fashion when estrogens are added. This evidence led them to accept the indirect-negative hypothesis of estrogen action. They called this inhibitory factor estrocolyone and characterized it as a non-steroid-like macromolecule.

The concept for the E-screen follows logically from this work. Serum extracted from human blood is treated with charcoal-dextran and added to MCF7 cells in vitro. A range of concentrations of a test chemical is applied, and after the sixth day, the cells are counted. If cell proliferation is observed,

---

[8] Soto, Ana and Carlos Sonnenschein (1985). "The Role of Estrogens on the Proliferation of Human Breast Tumor Cells (MCF-7)," *Journal of Steroid Biochemistry*, Vol. 23, No. 1, pp. 87-94.

[9] Soto, Ana, Renee Silvia, and Carlos Sonnenschein (1992b). "A Plasm-Borne Specific Inhibitor of the Proliferation of Human Estrogen-Sensitive Breast Tumor Cells (Estrocolyone-I)," *Journal of Steroid Biochemistry and Molercular Biology*, Vol. 43, No. 7, pp. 703-712.

then the inhibitory effect of the serum's estrocolyone was overcome. The chemical has estrogenic effects.

The potential for using a bioassay based on this work became apparent during the course of their experiments. MCF7 cells with CDHuS began proliferating regardless of whether estrogen had been added; these observations could not be explained. After 18 months of investigation, they discovered that the proliferation was being caused by a chemical found in their laboratory plastic ware.[10] Unknown to them, the manufacturer of the plastic ware had slightly modified the composition of its plastic, polystyrene. When heated, the plastic ware released p-nonyl-phenol. Further experiments by Soto et al. showed that this chemical caused the induction of the progesterone receptor in MCF7 cells, an expected effect of estrogens. In addition, they found that p-nonyl-phenol triggered mitotic activity in rat endometrium, the defining characteristic of estrogenic substances according to Hertz. The ability of their assay to identify p-nonyl-phenol bolstered their belief that the E-screen was a reliable assay to detect estrogenic chemicals.[11] The unexpected presence of an estrogenic chemical in plastic touched off concern that the ubiquitous use of plastic kitchenware could be exposing the public to estrogen-mimicking chemicals.[12] In addition, the E-screen has been used to show that canned vegetables can become estrogenic as a result of chemicals released from the inside of lacquer-coated cans.[13]

Early results

Scientists are currently engaged in validation studies of the E-screen. Some of the first results were presented at the Wingspread Conference in 1991,[14] and further testing continues.[15] At least 100 agricultural and industrial chemicals have been tested in the E-screen. The estrogenic

---

10 Soto, Ana (1995). Presentation given at the MIT Seminar on the Management of the Future Uses of Chlorine, April 18.

11 Soto, Ana, Honorato Justicia, Jonathan Wray, and Carlos Sonnenschein (1991). "p-Nonyl-Phenol: An Estrogenic Xenobiotic Released from 'Modified' Polystyrene," *Environmental Health Perspectives*, Vol. 92, pp. 167-173.

12 Weiss (1994).

13 Brotons, Jose, Maria Olea-Serrano, Mercedes Villalobos, Vincente Pedraza, and Nicolas Olea (1995). "Xenoestrogens Released from Lacquer Coatings in Food Cans," *Environmental Health Perspectives*, Vol. 103, No. 6, June, p. 608.

14 Soto et al. (1992a). pp. 295-309.

15 Soto et al. (in press).

characteristics of chemicals can be compared to the response in the E-screen of estradiol. One measure of response in the E-screen is relative proliferative effect (RPE). This is 100 times the ratio of the highest yield (proliferation) produced by the test chemical to the highest yield produced by estradiol. Another measure used is the relative proliferative potency (RPP). RPP is the ratio of the minimum concentration of estradiol needed to achieve maximum yield to the minimum concentration of the test chemical needed to achieve its maximum yield. These concepts are illustrated in Figure 3.1, which shows the RPE and the RPP for estradiol and two hypothetical chemicals, one a full agonist and the other a partial agonist.

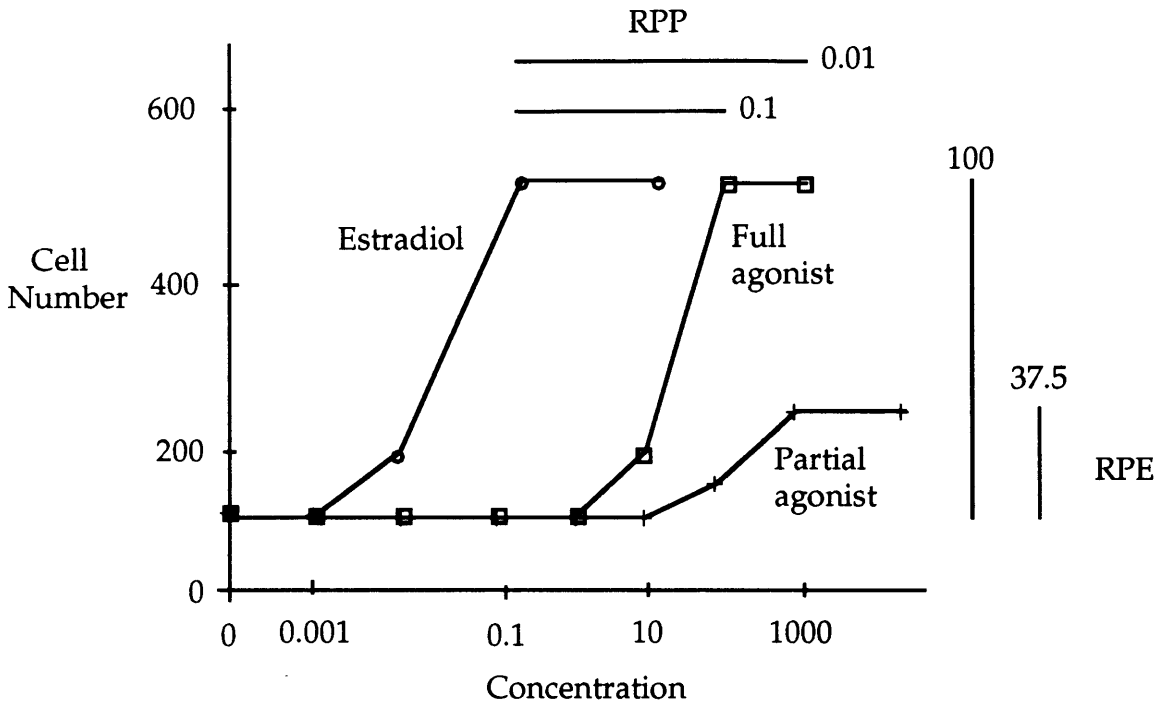
Soto et al. tested in the E-screen 22 chemicals reported to be estrogenic in animals. Table 3.1 shows the results. Although there were some exceptions, the relative potency measured by the E-screen correlated with both estrogen receptor binding assays and uterotropic assays.

The estrogenic characteristics of a number of agricultural and industrial chemicals are shown in Table 3.2. No herbicides or fungicides tested in the E-screen were shown to be estrogenic. Several pesticides, including DDT and its metabolites were estrogenic. Methoxychlor was also found to be estrogenic. Methoxychlor was not expected to be positive in the E-screen because it requires metabolic activation; it is hypothesized that MCF7 cells possess enzymes that can metabolize at least some chemicals into their estrogenically active forms. None of these agricultural chemicals have RPEs greater than that of estradiol. Higher RPE's were not observed because higher concentrations led to toxic effects. In addition to the industrial chemicals shown in Table 3.2, 28 PCB congeners and hydroxylated PCBs were tested. A little less than half were estrogenic, and their RPPs ranged from 0.01 to 0.0001.

The results of the E-screen have also been compared to other biomarkers for estrogenic activity, including levels of progesterone receptor, pS2, and relative binding affinities. Several human growth factors have been investigated, and it has been found that they did not affect proliferation and thus were not responsible for any of the results observed for agricultural and industrial chemicals.

Another line of investigation has found that estrogen-mimicking chemicals appear to act cumulatively. Soto et al. simultaneously applied to the E-screen ten different chemicals, each at doses that were one-tenth the

Figure 3.1 Schematic dose-response curves for estradiol, a full agonist, and a partial agonist



Source: Soto et al. (in press).

Table 3.1 E-screen results for chemicals known to be estrogenic in animal models

Compound	RPP, %
17b-Estradiol	100
17a-Estradiol	10
Estrone	1
Estriol	10
Moxestrol	1000
16-Hydroxyestrone	0.1
Diethylstilbestrol (DES)	1000
cis-Tamoxifen	0.001
Metabolite E (from Tamoxifen)	0.001
R26008	0.1
11b-chloromethylestradiol	1000
Indanestrol	10
Indenestrol-A	100
Indenestrol-B	10
Pseudo-DES	0.1
Pseudo-DES-e	10
Pseudo-DES-z	10
Zearalenol	1
Zearalenone	1
d-Equilenin	1
Coumestrol	0.001
Ethynyl-estradiol	100

Source: Soto et al. (in press).

Table 3.2 Agricultural and industrial chemicals tested in the E-screen

Estrogenic chemicals	Nonestrogenic chemicals	
Herbicides	Herbicides	
None	2,4-D	2,4-DB
	Alachlor	Atrazine
	Butylate	Cyanazine
	Dacthal	Dinoseb
	Hexazinone	Metolachlor
	Propazine	Picloram
	Trifluraline	Simazine
Insecticides	Insecticides	
p,p'-DDT	Bendiocarb	Carbofuran
o,p'-DDT	Chlordane	Chlordimeform
o,p'-DDE	Diazinon	Chlorpyrifos
Dieldrin	Heptachlor	Carbaryl
Chlordecone (kepone)	Kelthane	Lindane
Endosulfan	Malathion	Mirex
α-Endosulfan	Methoprene	Pyrethrum
β-Endosulfan	Parathion	
Methoxychlor	Rotenone	
Toxaphene		
Fungicides	Fungicides	
None	Chlorothalonil	Thiram
	Hexachlorobenzene	Zineb
	Maneb	Zirem
	Metiram	
Industrial chemicals	Industrial chemicals	
2,3,4-TCB	Butylated hydroxytoluene	2-CB
2,2',4,5-TCB	2,5-DCB	4-CB
2,3,4,5-TCB	2,6-DCB	2,3,6-TCB
2,4,4',6-TCB	3,5-DCB	2,3,5,6-TCB
2,2',3,3',6,6'-HCB	2,3',5-TCB	2,3,3',4,5'-PCB
2-OH-2',5'-DCB	2,3,4,4'-TCB	2,2',3,3',5,5'-HCB
3-OH-2',5'-DCB	2,3,4,5,6-PCB	DecaCB
4-OH-2',5'-DCB	2-OH-2',3',4',5,5'-PCB	Diamyl phthalate
4-OH-2,2',5'-TCB	2-OH-3,5-DCB	Dibutyl phthalate
4-OH-2',4',6'-TCB	4-OH-3,5-DCB	Dimethyl isophthalate
3-OH-2',3',4',5'-TCB	1,2-Dichloropropane	Dimethyl terephthalate
4-OH-2',3',4',5'-TCB	Irganox 1640	Dinonyl phthalate
4-OH-alkyl-phenols	2,3,7,8-TCDD	Octachlorostyrene
Bisphenol-A	Tetrachloroethylene	Styrene
4-OH-biphenyl		
t-Butylhydroxyanisole		
Benzylbutylphthalate		

Source: Soto et al. (in press).

minimal dose needed to produce a proliferative response individually. This mixture produced a response in the E-screen that was additive. This evidence suggests that measurement of the total body burden of estrogenic chemicals is more appropriate than the more common measurement of individual chemicals in the body.

As to the issue of validation, Soto et al. claim that the E-screen has not produced any false negatives or false positives among chemicals that have been tested for their estrogenicity by other methods. They argue that the E-screen could be established as a quantitative standard for estrogenic activity at the target organ level.

Problematic issues

The metabolic capabilities of MCF7 cells are uncertain. Metabolism in vivo can greatly affect estrogenic activity of chemicals, transforming them into more or less active compounds. Soto et al. point out that more complete testing would include the use of liver extracts to metabolize chemicals before they are tested in the E-screen.

Although they characterized and documented well the proliferative effect and proliferative potencies of the chemicals tested, Soto et al. did not give explicit or rigorous treatment to their assertion that no false positive or false negatives were encountered. Presumably, this implies that there is a standard of comparison to which E-screen results are being compared. They argue, however, that because there is a wide variety of animal models and endpoints used to measure estrogenicity, there is no "gold standard" for animal bioassays that measure estrogenicity. If this is the case, then it is not entirely clear what the results of the E-screen should be compared to. At the same time, they consider mitotic activity of the female genital tract the hallmark of estrogen activity.

Although the evidence presented by Soto et al. may be persuasive of the E-screen's utility and reliability, it is important to recognize the complexity and uncertainty in estrogen action. It is not clear whether the E-screen represents the *best* bioassay to identify estrogen-mimicking chemicals. It is possible, even likely, that there are other mechanisms of estrogen action that may not be detected by the E-screen.

For example, the work of Liu et al. suggests that a chemical may be estrogenic only under certain circumstances.[16] Their research examined the estrogenic and anti-estrogenic effects of indolo[3,2-b]carbazole (ICZ). ICZ is an acid-derived condensation product of indole-3-carbinol (IC3), a naturally occurring compound found in several vegetables such as broccoli and cabbage.

Rather than using cell proliferation as an indicator of estrogenicity, Liu et. al. used MCF7 cells transfected with a reporter gene. When estradiol and ICZ were applied separately to MCF7 cells in vitro, reporter gene activity was observed, indicating the estrogenic abilities of these two chemicals. Liu et al. discovered, however, that when ICZ and estradiol were applied to the cells at the same time, the level of reporter gene activity was reduced as compared to the application of estradiol alone, suggesting that ICZ is anti-estrogenic.

Liu et. al. considered several possible explanations for the anti-estrogenic effects of ICZ in combination with estradiol. First, ICZ could be competitively binding to estrogen receptors, preventing some estradiol from binding to the estrogen receptor. Since estradiol is a more potent estrogen than ICZ, the net effect would be to reduce reporter gene activity. Liu et. al. argue that this explanation is unlikely because concentrations of ICZ applied were too low to displace the estradiol. Second, ICZ induces CYP1A1, an enzyme that causes the hydroxylation of estradiol, thereby inactivating it. This would reduce the reporter gene activity. Liu et. al. also discount this possibility because the induction of CYP1A1 occurs too slowly and only at ICZ concentrations higher than those actually applied.

Liu et. al. hypothesize that the anti-estrogenic effects of ICZ in combination with estradiol are related to the estrogen receptor only indirectly. They suggest that ICZ binds to the Ah (aryl hydrocarbon) receptor, which then interacts with the estrogen receptor system. This anti-estrogenic interaction between the Ah and estrogen receptors is still unknown.

---

16 Liu, Hong, Mark Wormke, Stephen Safe, Leonard Bjeldanes (1994). "Indole [3,2-b]carbazole: a Dietary-Derived Factor That Exhibits Both Antiestrogenic and Estrogenic Activity," *Journal of the National Cancer Institute*, Vol. 86, No. 23, December 7, p. 758.

Conclusion

As concern for the potential risks posed by estrogen-mimicking chemicals grows, so too does the call for action to control these risks. One part of dealing with the risks is the identification of suspect causes. The E-screen and several other tests may play an important role in this respect. The E-screen possesses several advantages over other tests. Attractive are its simplicity, its quick turn around time, and its apparent ability to detect estrogenic chemicals.

There appears to be a great need for the E-screen, and the early results are promising. But to expect that the E-screen should or will be quickly embraced by regulators and industry is to ignore the history of environmental regulation and the context within which the E-screen might fit. It would be a disservice to regulators, industry, and scientists if this history was ignored. Reasonable expectations that reflect history, current circumstances, and future problems and opportunities will aid, in the long run, the development of screening technologies and environmental policy.

Predicting what will happen is no easy task. The concern for estrogen-mimicking chemicals is relatively new, and the E-screen is newer still. In the next chapter, a case study is presented for the purpose of shaping and adding texture to our expectations for the E-screen. The case study will illuminate some of the important issues that are likely to influence how the E-screen is used and what influence the test might have. The case study will examine the scientific and regulatory history of the Ames Test, the most extensively used short term test in cancer risk assessment.

Chapter Four

Cancer Risk Assessment and the Ames Test

One of the main goals of this thesis is to anticipate the future regulatory impacts of the E-screen. This task is no small undertaking; as Neils Bohr once said, "It is very difficult to make an accurate prediction, especially about the future." The E-screen is an emerging technology with which the scientific community has had little experience. It is still unclear how the E-screen will compare to other methods of identifying estrogen-mimicking chemicals, such as structure-activity relationships and animal bioassays. Although some estrogen-mimicking chemicals like diethylstilbestrol have been regulated on the basis of their carcinogenicity, concern for estrogen-mimicking chemicals is still in an embryonic stage in the regulatory arena.

In an attempt to provide a framework for understanding the regulatory impacts of short term tests, this chapter turns away from endocrine disruption and focuses on the use of short-term tests used in cancer risk assessment. In particular, the Ames Test for mutagenicity will be scrutinized to provide a point of reference with which to assess the E-screen.

The Ames Test provides a useful case study, for it has a rich history in scientific, industrial, and regulatory communities. In the first decade after its introduction in the early 1970s by University of California scientist Bruce Ames, the Ames Test became widely used. Compared to long-term animal bioassays, the Ames Test is less expensive and less time-consuming. Accompanying these advantages were high hopes that the Ames Test would accurately and efficiently identify carcinogenic chemicals. By 1979, the test was being used in more than 2000 governmental, industrial, and academic laboratories, and the test had been applied to some 2600 chemicals.[1] It is still

---

[1] Devoret, Raymond (1979). "Bacterial Tests for Potential Carcinogens," *Scientific American*, Vol. 241, No. 2, August, p. 45.

the most widely used short term test for cancer risk assessment.[2] There has been, however, a great deal of controversy surrounding the degree to which environmental regulations should rely on short term tests like the Ames Test. The cost and time advantages of the Ames Test must be balanced against the fact that the test is an imperfect identifier of carcinogens. The Ames Test can only identify mutagens, which *may* be accountable for causing cancer.

The high expectations for the influence of the Ames Test in cancer risk assessment have given way to a more realistic expectations under scientific and political scrutiny. Several environmental regulations, risk assessment guidelines, and chemical testing programs still recognize or incorporate the use the Ames Test in risk assessment, but its influence is somewhat limited. This chapter traces the scientific and political history of the Ames Test. The goal of this chapter is to highlight several key issues in the Ames Test and cancer policy which will later provide some lessons to be applied to understanding the impacts that the E-screen and other short term tests may have. First, society's strong concern for cancer and the establishment of a large and influential lobby for cancer research will be discussed. This chapter will then describe how expectations of the Ames Test developed and matured into its current use in environmental regulation. Ultimately, the goal of this chapter is to provide insight into the ways in which the Ames Test is or is not influential.

Cancer: the dread disease

The fear of cancer began long ago. The term "cancer" comes from the Latin word for crabs, which was used by Hippocrates to describe the disease: a slow, creeping consumption of the flesh. Ulysses S. Grant's well publicized death from cancer in the late 1880s, engendered a cancerphobia that led to the use of scientific and medical research for finding the causes and a cure for cancer.[3]

It was not until the 1940s that significant funding for cancer research was established. Government programs as large as that for cancer are the result of nothing less than tremendous effort. The roots of the enormous

---

2 OTA (1987). *Identifying and Regulating Carcinogens*, Washington, DC: US Government Printing Office, p. 164.

3 Patterson, James (1987). *The Dread Disease: Cancer and Modern American Culture*, Cambridge, MA: Harvard University Press.

public, professional, and political support for cancer prevention and cure can be traced largely to an energetic philanthropist, Mary Lasker, a successful businesswoman and wife of a millionaire who later died of intestinal cancer.[4] She worked tirelessly on several fronts to increase public awareness and support for cancer research. She arranged for a series of influential articles on cancer to appear in *Reader's Digest*, reorganized the American Cancer Society, was instrumental in forming a coalition of wealthy and influential businesspersons and politicians, and served on numerous presidential and congressional panels and commissions. Rettig attributes to Mary Lasker

> an unparalleled knowledge of how to publicize and popularize a major initiative....She has a keen understanding of the processes of government, a first-name acquaintance with most key officials, and a refined sense of timing. She is, without question, a remarkable political figure, and she and her friends, on more than one occasion, have attempted to redraw the map of medical research in the United States....Mrs. Lasker and her friends have displayed great capacity to overwhelm the opposition through a focused, highly publicized, over-simplified, dramatic appeal to the public and through the skillful tactical maneuver within the political process.[5]

In the late 1960s, Lasker embraced an initiative that helped launch the war on cancer. Physician Solomon Garb, in his 1968 book, *Cure for Cancer: A National Goal*, emphatically argued for a "national commitment to make the cure...of cancer a national goal, in the same way that putting a man into orbit around the earth was made a national goal, and then achieved."[6] Lasker and her associates embarked on a grass roots strategy to bring this concept to reality. In 1969, they created the Citizens Committee for the Conquest of Cancer. In a full page advertisement in the *New York Times*, the committee declared:

> MR. NIXON: YOU CAN CURE CANCER....
>
> This year, Mr. President, you have it in your power to begin to end this curse.
>
> As you agonize over the Budget, we beg you to remember the agony of those 318,000 Americans [who died of cancer in 1968]....

---

[4] Rettig, Richard (1977). *Cancer Crusade: The Story of the National Cancer Act of 1971*, Princeton, New Jersey: Princeton University Press, p. 18.

[5] Rettig (1977). pp. 23-24, 41.

[6] Rettig (1977). p. 79, Citing Solomon Garb, *Cure for Cancer: A National Goal*, New York: Springer Publishing Company, 1968.

> We ask a better...way to allocate our money to save hundreds of thousands of lives each year.
>
> America can do this. There is not a doubt in the minds of our top cancer researchers that the final answer to cancer can be found....
>
> Dr. Sidney Farber, Past President of the American Cancer Society believes: "We are so close to a cure for cancer. We lack only the will and the kind of money and comprehensive planning that went into putting a man on the moon."
>
> Why don't we try to conquer cancer by America's 200th birthday?...
>
> Our nation has the money on one hand and the skills on the other. We must, under your leadership, put our hands together and get this thing done.[7]

At the same time, Lasker and her associates were taking advantage of their insider's access to Congressional members in order to increase funding and public support for cancer research. The key office upon which Lasker focused her lobbying efforts belonged to Senator Ralph Yarborough of Texas, who was the chairman of the Labor and Public Welfare's subcommittee on health. Mrs. Lasker suggested to Senator Yarborough in 1969 that he should create an outside citizen's committee to advise the full Senate Committee on Labor and Public Welfare.[8]

The wheels of Congress began to move. By the end of April of 1970, Senator Yarborough secured unanimous Senate support for a resolution to establish a Panel of Consultants on the Conquest of Cancer. In a press release, the Senator Yarborough stated that the Panel of Consultants was to study cancer research needs and

> to recommend to Congress and to the American people what must be done to achieve cures for the major forms of cancer by 1976—the 200th anniversary of the founding of this great Republic....[Cancer] is a vicious disease...and mankind's most relentless foe....[They should strive to guarantee] that the conquest of cancer become a highly visible national goal.[9]

Similarly, the House of Representatives passed a concurrent resolution by unanimous vote. The resolution declared

---

7 Rettig (1977). p. 79. Citing *New York Times*, December 9, 1969, p. 61.

8 Rettig (1977). p. 80.

9 Strickland, Stephen (1972). *Politics, Science, and the Dread Disease*. Cambridge, Massachusetts: Harvard University Press. p. 260.

> That the Congress appropriate the funds necessary for a massive program of cancer research and for the buildings and equipment with which to conduct the research and for whatever other purposes are necessary to the crusade so that the citizens of this land and of all other lands may be delivered from the greatest scourge in history.[10]

Rettig observes that these actions indicate not only the congressional attitude of resolve, but also their "sometimes uncritical" optimism. Although the exact form by which this effort was to be implemented was the subject of some debate, the goal and hope of a cure in only six years was not scrutinized. Rettig argues that "no sober voice" of contradiction to these claims was heard in Congress, the National Cancer Institute, the American Cancer Society, the Panel of Consultants, or the scientific community.[11]

Just eight months after its establishment, the Panel of Consultants made their report at a congressional hearing. The Panel of Consultants consisted of prestigious members of the medical research community, a Nobel laureate geneticist, and "lay people" like the president of the United Steel Workers, large contributors to political parties, wealthy philanthropists, and journalists.[12] The Panel's voice could not be expected to be ignored, or, for that matter, entirely unbiased.

The recommendations were significant. Senator Jacob Javits prefaced the report of Panel by pledging the adoption of the Panel's recommendations with that hope that "we can do for cancer what the Salk vaccine did for polio."[13] The Panel recommended the creation of a new government agency with the mission to conquer cancer and the infusion of greater resources to the scientific research. For the next year, Senator Ted Kennedy, who replaced Senator Yarborough as subcommittee chairman, worked to implement the Panel's recommendations as incorporated in the Conquest of Cancer Act. President Nixon, wary of Senator Kennedy's aggressive moves in the emerging political arena of health and attempting to take the initiative from his political foe, responded by stating in his 1971 State of the Union speech that "The time has come in America when the same kind of concentrated effort that split the atom and took man to the moon should be turned toward conquering this dread

10 Rettig (1977). p. 82
11 Rettig (1977). p. 83.
12 Rettig (1977). p. 85.
13 Rettig (1977). p. 103.

disease. Let us make a total national commitment to achieve this goal."[14] Much of the debate between Kennedy and Nixon centered around whether a new cancer agency should be created or whether the national program would be based in the existing National Institutes of Health, but both Kennedy and Nixon supported greatly increasing resources for cancer research and placing it at the top of their agendas.[15]

The stakes were high, and Lasker assured this by increasing publicity for the measures that were being debated. Lasker convinced the highly read columnist Ann Landers to tell some 50 million readers about the importance of the Conquest of Cancer Act.[16] Ann Landers painted a stark picture, stating that more Americans died from cancer in 1969 than in World War II, that one of four Americans will eventually get cancer, and two of those three will die from it. Landers wrote:

> Dear Readers, If you're looking for a laugh today, you had better skip Ann Landers....[But] if you want to be a part of an effort that might save millions of lives—maybe your own—please stay with me....How many of us have asked...if this great country of ours can put a man on the moon, why can't we find a cure for cancer?...Today you have the opportunity to be a part of the mightiest offensive against a single disease in the history of our country. If enough citizens let their senators know that they want the bill—S. 34—passed, it will pass.

The response to the column was overwhelming. Senator Percy of Illinois had "never had such a tremendous response from any other issue." Similarly, Senator Cranston of California said that no other issue of 1971 had created so much mail. Four weeks after the column, the Senate had received 250,000 to 300,000 letters in support of the Conquest of Cancer Act.[17] With an election just around the corner, there was an urgency in Congress and the White House to get behind the public's outpouring of support for a national mission to cure cancer.

On December 23, 1971, President Nixon signed the Cancer Act. Although a new agency was not created to orchestrate the national effort, the

14 Rettig (1977). p. 77.
15 Strickland (1972). pp. 260-291.
16 Strickland (1972). p. 271.
17 Rettig (1977). 176-177.

Cancer Act authorized $1.6 billion in cancer research over a three-year period.[18] The "great crusade against cancer" had been launched.

This account of the events leading up to the beginning of the war on cancer reveals several themes. First, the public had raised a loud and consistent voice that cancer was among the most feared diseases and its cure was a national priority. Much of the public's response can be attributed to the masterful ability of Mary Lasker to engage and rally the public for the cancer crusade. Second, the public and political leadership expected that the control of cancer by 1976 was a realistic goal. Coming after the heels of a successful national commitment to the space program, the application of massive resources to the war on cancer seemed quite plausible to the public. In the face of overwhelming support for the campaign against cancer, where opponents to the Cancer Act were lambasted for being in favor of cancer, the scientific community was not vocal. One scientist at the ceremonial signing of the Cancer Act later recalled that the politicians were smiling, but the scientists were not; they were worried. "The hoopla surrounding the Cancer Act implied the conquest of cancer in the near future because [of increased research funding]. Those of us who knew the 'state of the art' had cause to worry."[19] Nevertheless, the pressure for finding the cure defined the atmosphere in cancer research.

Development of the Ames Test

Amid these political developments and societal attitudes about cancer, scientists were working to cure cancer and to understand its causes and mechanisms. Science does not proceed in a vacuum, but is affected by, among other things, societal attitudes and government funding. This section describes the scientific development of the search of methods to identify chemical carcinogens. The purpose of this section is to understand how the societal concern for cancer affected research and how society accepted the science.

---

18 Strickland (1972). pp. 288-289.

19 Rettig (1977). p. 277.

The need for short term tests

One of the great efforts in cancer research was (and is) the search for the causes and mechanisms of carcinogenesis. Research into chemical carcinogenesis has a long history.[20] Observations of tumors in patients exposed to large quantities of crude organic mixtures were made at least two centuries ago. In possibly the first epidemiological link between cancer and chemicals, Sir Percival Pott observed in the late 1700s that many of his patients with cancer of the scrotum were also chimney sweeps, suggesting to him that soot and coal tars were responsible for the disease. Laboratory research recorded in 1915 its first success in reproducibly inducing cancer in animals; Yamagiwa and Ichikawa induced tumors in rabbits to which coal tar had been repeatedly applied.

These two approaches to the identification of chemical carcinogens, epidemiology and animal bioassays, have several shortcomings.[21] Epidemiological studies are investigations of patterns of disease found in human populations and risk agents (e.g., chemicals) that influence these patterns. Well conducted epidemiological studies are generally viewed as the most relevant information regarding human health risks. Epidemiology is well suited in only limited circumstances: where exposure to the risk agent is high and the adverse health effect is unusual. Controlling for confounding factors such as lifestyle, different exposure patterns, and other unknown risk agents is a formidable obstacle. Epidemiological studies cannot be used to predict the effects of new risk agents such as new chemicals.

Due to these problematic aspects of epidemiological studies, long-term animal bioassays have been a mainstay in cancer research. Animals such as rats and mice can be exposed to chemicals and observations made of the effects. There is controversy, however, in making inferences from animal data to human outcomes. In order to observe effects, the animals are exposed to chemical doses that may be higher than the levels to which humans are exposed, introducing another controversial extrapolation from high doses to the low doses that may be more relevant to human risks. These tests require

---

20 Miller, James (1994). "Brief History of Chemical Carcinogenesis," *Cancer Letters*, Vol. 83, pp. 9-14.

21 Cohrssen, John and Vincent Covello (1989). *Risk Analysis: A Guide to Principles and Methods for Analyzing Health and Environmental Risks.* Prepared for the U.S. Council on Environmental Quality. pp. 27-43.

several years to perform and cost hundreds of thousands to several millions of dollars.

Identification of chemical carcinogens using these two methods appears to be an enormous task, considering the existence of 50,000 to 70,000 chemicals and the introduction of thousands more each year. Under these circumstances, clearly attractive are tests that can identify chemical carcinogens quickly and inexpensively.

The Ames Test

The initial search for a relatively quick and inexpensive method to identify carcinogens was driven from the start by what is called the somatic mutation theory of cancer.[22] According to this theory, the DNA of cells mutate to a form that prevents a normal cell from responding to signals to stop reproducing. These mutated cells grow uncontrollably, leading to cancer.

Given the hypothesis that at least some cancers are caused by mutation of cellular DNA, Bruce Ames developed a test to identify chemicals that caused mutation. The task of finding chemicals that cause mutations in DNA is far simpler than identifying chemicals that cause cancer, which is a much more complex process.

The Ames Test is conceptually very simple.[23] In the late 1960s, Bruce Ames developed a strain of Salmonella bacteria that is unable to synthesize histidine, an amino acid necessary for cellular reproduction. These Salmonella cells are placed in a petri plate; normally, these cells will eventually die. The only way that the bacteria can grow is if a mutation occurs that allows it to once again synthesize histidine. Therefore, if a chemical is added to the petri plate and large colonies of bacteria are observed (compared to controls), then the chemical can be considered a mutagen. The test requires only several days to perform and costs several hundred dollars.

Bruce Ames made a major breakthrough for the Ames Test when he added a mixture of liver cells to the petri plate. There was evidence that some cancer was caused by the metabolites of a chemical, rather than the chemical

---

22 James, Robert and Christopher Teaf (1985). "Carcinogenesis," in *Industrial Toxicology*, Williams and Burson, eds., New York: Van Nostrand Reinhold, p. 307.
23 Teaf, Christopher (1985). "Mutagenesis," in *Industrial Toxicology*, Williams and Burson, eds., New York: Van Nostrand Reinhold, pp. 286-287.

itself. The added liver cells allow for metabolism of chemicals to their carcinogenic form, making the Ames Test more realistic and relevant to human carcinogenesis.

Early results

The early results of the Ames Test were spectacular. In 1973, Ames and his colleagues completed a study of eighteen carcinogens and declared that "carcinogens are mutagens."[24] The Ames Test was applied to twenty aromatic type chemicals that had previously been shown to be carcinogenic in humans or animals; included were aflotoxin B_1 and benzo(a)pyrene. Eighteen were shown to be positive (mutagenic) in the Ames Test, and two were negative. Ames pointed out that negative results did not necessarily imply that the chemicals were not mutagenic, but left open the possibility that the chemicals caused a different type of mutation not yet detectable in the Ames Test.

Ames clearly believed that these results bolstered the somatic mutation theory of cancer, proposing that "those carcinogens that are mutagens cause cancer by somatic mutation. This hypothesis, which of course is not new, seems compelling in view of this correspondence between carcinogens and mutagens." Ames also hinted at the possible regulatory implications of his new technology for human cancer: "It is quite reasonable to use bacteria as a test system for carcinogen detection, because so many carcinogens appear to be mutagens acting on DNA and all DNA is basically the same." Ames continued, saying that the Ames Test

> should be used in screening of suspected carcinogens....The [test] system provides a rapid, simple, sensitive, and economical method for detecting those carcinogens that cause point mutations. This class, though it includes an impressive array of chemical carcinogens and radiations, probably will not include all carcinogens: e.g., those that cause mutations indirectly by inhibiting mammalian repair.

Ames had clearly anticipated that need for such a short term test with these attractive cost and time attributes. Several points deserve attention

---

[24] Ames, Bruce, William Durston, Edith Yamasaki, and Frank Lee (1973). "Carcinogens are Mutagens: A Simple Test System Combining Liver Homogenates for Activation and Bacteria for Detection," *Proceedings of the National Academy of Sciences USA*, Vol. 70, No. 8, August, pp. 2281.

here. First, it is not clear what Ames meant by arguing that his test should be used in *screening*. Screening could mean that test results could be the basis for decisions to prevent new chemicals from being developed and manufactured; or it could mean that test results could be used as a means to identify chemicals that require further testing in more sophisticated, time-consuming, and costly methods. It is not clear that the Ames Test would be equally useful in these two applications. Second, the ability to detect mutagens is only useful to the extent that the somatic mutation theory of cancer is correct. It must be emphasized that the Ames Test detects *mutagens*, not *carcinogens*. If mutagens do in fact cause cancer, than the detection of mutagens will be very useful in regulating carcinogens. While Ames qualifies his statement about the usefulness of his test by saying that it provides a method for detecting *those carcinogens that cause point mutations*, he nevertheless declares that *carcinogens are mutagens*. The accuracy of this latter statement was to become a controversial matter.

The first major study applying the Ames Test to a large number of chemicals was published by McCann, Choi, Yamasaki, and Ames in a series of papers in 1975[25] and 1976.[26] In this study, 300 chemicals were investigated in the Ames Test. Most chemicals were chosen to represent several classes of chemicals and because they had been subjected to carcinogenicity tests in animals or were known to be human carcinogens. This group of chemicals also included chemicals thought to be "non-carcinogens." In addition, the study examined chemicals for which their carcinogenicity was uncertain. The author's motivation for undertaking the study and their conclusions were stated clearly:

> A program of cancer prevention aimed at identifying and eliminating human exposure to hazardous chemicals requires the development of rapid, inexpensive, screening methods as complements to expensive, long-term animal tests, to pinpoint dangerous chemicals among the thousands to which humans are exposed. The [Ames] mutagenicity test

---

25 McCann, Joyce, Edmund Choi, Edith Yamasaki, and Bruce Ames (1975). "Detection of Carcinogens as Mutagens in the Salmonella/Microsome Test: Assay of 300 Chemicals," *Proceedings of the National Academy of Sciences USA*, Vol. 72, No. 12, December, pp. 5135-5139.

26 McCann, Joyce, Edmund Choi, Edith Yamasaki, and Bruce Ames (1976). "Detection of Carcinogens as Mutagens in the Salmonella/Microsome Test: Assay of 300 Chemicals: Discussion," *Proceedings of the National Academy of Sciences USA*, Vol. 73, No. 3, March, pp. 950-954.

> has been sufficiently developed and validated to be seriously considered for widespread use in this way.

This is a particularly insightful paper, for McCann et al. highlight several problems with the test that may also be applicable to tests for estrogen-mimicking chemicals. We will discuss this paper is some detail.

Several measures of the correlation between mutagenicity and carcinogenicity can be used to characterize the Ames Test.[27] First, the sensitivity of the test refers to the frequency with which carcinogens are correctly identified by the test, i.e., the frequency that a carcinogen tests positively (a mutagen). Second, specificity refers to the frequency that non-carcinogens are correctly identified, i.e., the frequency that non-carcinogens test negatively (not a mutagen). Ideally, both the sensitivity and specificity would be 100 percent, implying a perfect correlation between mutagenicity and carcinogenicity.

In the 1975-76 study, the Ames Test had a sensitivity of 90 percent; 175 carcinogens were tested and 157 produced positive results in the Ames Test. The other 18 carcinogens were not mutagenic in the Ames Test, resulting in a 10 percent false-negative rate. McCann et al. proposed several reasons for these false-positives. First, some chemicals may require additional metabolic activity before they exert mutagenic effects in the Ames Test. In particular, some chlorinated hydrocarbons shown to be carcinogenic were not mutagenic in the Ames Test. McCann et al. believe that false negative results could be corrected by using other short term tests or with improvements in the Ames Test. Second, the presumed carcinogenicity of the chemicals were open to question. Some chemicals were identified as carcinogens on the basis of limited studies or were considered to be only weak carcinogens.

In addition, the Ames Test exhibited a specificity of 87 percent; that is, out of 108 non-carcinogens, 94 were not mutagenic in the Ames Test. Thirteen percent of the non-carcinogens showed mutagenic activity (false positives). McCann et al. explained that it is possible that these false-positives are the result of incorrect determinations from animal bioassays that a chemical is a non-carcinogen when it is in fact carcinogenic. While the Ames Test can detect chemicals that may be very weakly mutagenic, animal bioassays cannot be

---

[27] OTA (1981). *Assessment of Technologies for Determining Cancer Risks from the Environment*. Washington, DC: US Government Printing Office, pp. 116-117.

used to detect chemicals that are only weak carcinogens. McCann et al. stated that the "statistical limitations inherent in animal carcinogenicity tests limit their usefulness for the detection of weak carcinogens. More extensive animal carcinogenicity tests may therefore be required to determine if the 'false positives' showing weak mutagenic activity are really non-carcinogens."[28]

This discussion on the sensitivity and specificity of the Ames Test highlights a key issue in the validation of a short term test: the validity of the test must have a meaningful standard of comparison. Animal carcinogenicity tests that are typically used as the standard of comparison may not be accurate. Not only is there uncertainty in the whether or not animal bioassays can detect whether a chemical is carcinogenic in animals, but there is also uncertainty in whether a chemical that is carcinogenic in animals is also carcinogenic in humans.

Maturation of the Ames Test

The advent of the Ames Test was just the first of many short term tests that were to be developed, although none were so widely used. By the 1979, there were at least 100 short term tests for detecting potential carcinogens and mutagens.[29] The expectations of the Ames Test evolved as more chemicals were tested for mutagenicity and carcinogenicity and as more scientists gained experience with the test.[30] While many studies stated that the Ames Test is a valuable tool that can complement long-term animal bioassays in the identification of carcinogens, it became more apparent that Ames' optimistic statement in 1973 that "carcinogens are mutagens" did not reflect the complexity of carcinogenesis.

Several investigators suggested that the correlation was not as high as initially believed. Bartsch et al. tested 89 chemicals for which sufficient data existed to classify each of them as carcinogenic or non-carcinogenic in animals.

---

28 McCann et al. (1976). p. 952.

29 Hollstein, Monica and Joyce McCann (1979). "Short Term Tests for Carcinogens and Mutagens," *Mutation Research*, Vol. 65, p. 135.

30 See for example: Hollstein and McCann (1979); Purchase, IFH, E Longstaff, John Ashby, JA Styles, Diana Anderson, PA Lefevre, and RF Westwood (1976). "Evaluation of Six Short Term Tests for Detecting Organic Chemical Carcinogens and Recommendations for the Use," *Nature*, Vol. 264, December 16, pp. 624-627; Bridges, Bryn (1976). "Short Term Screening Tests for Carcinogens," *Nature*, Vol. 261, May 20, pp. 195-200.

This study showed that the Ames Test had a concordance[31] of 74 percent.[32] Legator tested 274 chemicals and found that the concordance between the Ames Test and animal carcinogenicity tests was about 76 percent.[33] Legator observed that the concordance of different chemical classes varies. Lijinsky corroborates this assertion in his study that showed that the concordance for polycyclic hydrocarbons was 55 percent, while nitrosamines exhibited a concordance of 70 percent.[34] Hollstein notes that the Ames Test does not provide meaningful results when used to test metals.[35]

The Ames Test's underlying premise—the somatic mutation theory of cancer—was also the target of criticism. In 1976, Rubin argued that this theory deserved further scrutiny because it could not explain the origins of several cancers.[36] There was evidence that carcinogenesis could proceed by means other than genetic mutation. While acknowledging the desirability of accurate short term tests, Rubin argued that reliance on the Ames Test is misguided:

> [Using the Ames Test to identify carcinogens] is a bit like looking under a lamppost for the coin lost a block away because of the availability of light. For the present, we must still assume the hard and expensive task of looking for carcinogens by determining a compound's carcinogenic action because that is the only way we can know what we have found.

By 1979, with several years of experience with the test, Bruce Ames had developed a more realistic picture of the test.[37] Ames stated that the test is much less successful in identifying certain classes of carcinogens, such as hydrazines and heavily chlorinated chemicals. Furthermore, he acknowledged that the Ames Test can only detect genotoxic carcinogens (initiators), but not epigenetic carcinogens (promoters) since the latter probably do not act through direct interaction with DNA. These exceptions are quite important, since they

---

31 Concordance is the ratio of the number of results that are correctly positive and correctly negative to the number of chemicals tested.

32 OTA (1981). p. 118.

33 Maugh, Thomas (1978). "Chemical Carcinogens: the Scientific Basis for Regulation," *Science*, Vol. 201, Sept. 29, p. 1205

34 Maugh (1978). p. 1205.

35 Hollstein and McCann (1979). p. 140.

36 Rubin, Harry (1976). "Carcinogenicity Tests," *Science*, Vol. 191, January 23, p. 241.

37 Ames, Bruce (1979). "Identifying Environmental Chemicals Causing Mutations and Cancer," *Science*, Vol. 204, May 11, p. 589.

represent classes of chemicals that have been regulated as carcinogens. For example, heavily chlorinated chemicals have been regulated pesticides.

Nevertheless, the Ames Test could still play an important role in the identification of carcinogens, argued Ames. Because a single short term test cannot detect all classes of carcinogens, use of a battery of complementary short term tests would provide more meaningful results. In addition, such a battery of short term tests could be used to prioritize chemicals for further testing in traditional long term animal bioassays.

Work continued on examining the correlation between the Ames Test and carcinogenicity. In 1987, Tennant et al. published an influential study that argued that previous correlation studies suffered in two important ways.[38] First, standard protocols were not yet established; Tennant and his colleagues relied on a standard protocol developed by the National Toxicology Program. Such a protocol was shown to provide reproducible results in inter-laboratory trials. Second, correlation studies based on evaluation of published literature reflect a bias because the literature tends to focus on strongly positive mutagens and mammalian carcinogens. In addition, there were few results documenting non-carcinogens. Correcting for this bias was one goal in the selection of chemicals to be examined in their study, which was initiated by the National Toxicology Program in 1984. The results are shown in Table 4.1. The correlation between the Ames Test and carcinogenicity was lower than previously indicated.

Tennant et al. also examined whether the performance of the Ames Test would be affected by refocusing the correlation study on only certain groupings of carcinogens. This aspect of the study was motivated by their concern that simplifying complex animal carcinogenicity data into a only a positive or negative result. Consequently, Tennant et al. reconsidered the performance of the Ames Test when applied only to a certain group of chemicals, including the 21 chemicals that were judged to be high potency carcinogens, the 32 chemicals that were carcinogenic in more than one sex or animal species, and the 20 chemicals exhibiting animal carcinogenicity in multiple organs. The concordance was found to improve, but it did not exceed 74 percent.

---

38 Tennant, Raymond, Barry Margolin, and Michael Shelby (1987). "Prediction of Chemical Carcinogenicity in Rodents from in Vitro Genetic Toxicity Assays," *Science*, Vol. 236, May 22, pp. 933-941.

Table 4.1 Correlation between results of the Ames Test and carcinogenicity.

	Ames Test	
Carcinogenicity	+	—
+	20	24
—	4	25

Sensitivity, %	45	(% of carcinogens yielding positive Ames Test)
Specificity, %	86	(% of noncarcinogens yielding negative Ames Test)
Positive predictivity, %	83	(% of Ames Test positives that are carcinogens)
Negative predictivity, %	51	(% of Ames Test negatives that are noncarcinogens)
Concordance, %	62	(% of qualitative agreements between Ames Test and rodent carcinogenicity results)

Source: Tennant et al. (1987).

The Ames Test in the regulatory arena

The concerns and controversies expressed about the Ames Test in scientific journals were only amplified when discussed in the regulatory arena. In this section, we examine the use of the Ames Test in environmental decision making. The term regulatory arena is meant to include not only the decisions and rules made by agencies like the EPA, but also the internal decisions made by companies in response to or in anticipation of environmental concerns. We discuss below the use of the Ames Test in several circumstances: risk assessment guidelines, the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), and the Toxic Substances Control Act (TSCA).

Risk assessment guidelines

The first place that we can examine to understand the effect of the Ames Test in the regulatory arena is the policies that have been established

by several regulatory bodies to guide the practice of risk assessment. Since the evidence on which environmental regulations are based cannot be characterized as rigorous scientific proof, subjective judgment is often required. Recognizing this in its influential guide for risk assessment published in 1983 (the "Red Book"), the National Research Council asserted that regulatory bodies should establish risk assessment guidelines.[39]

Because of the ambiguity in the data and gaps in scientific theory that are characteristically encountered in environmental regulation, judgments must be made often between several scientifically plausible option, which the NRC called "inference options." Rather than engaging in prolonged debate in making these inference options on a substance by substance basis, the NRC argued that risk assessment guidelines would lay out general principles about what inference options should be made. By applying general guidelines to a series of chemical risk assessments, the regulatory agency expedites the risk assessment process. Furthermore, if decisions were made entirely on an ad hoc basis, the regulatory agency's decision could more easily be criticized for being influenced by special interests.

One issue that guidelines have discussed is the kinds of evidence that are sufficient to make a determination about the likelihood that a substance can be considered harmful. The NRC discusses the four main types of evidence—epidemiology, animal bioassays, short term tests, and comparisons of molecular structure. Epidemiological and animal bioassay studies have provided the most persuasive evidence. On the issue of short term tests, the NRC stated that there is "considerable evidence" that supports the correlation between mutagenicity and carcinogenicity. However, the NRC noted that short term mutagenicity tests, "in the absence of a positive animal bioassay, are rarely, if ever, sufficient to support a conclusion that an agent is carcinogenic. Because short term tests are rapid and inexpensive, they are valuable for screening chemicals for potential carcinogenicity and lending additional support to observations from animal and epidemiological data."[40]

In this section, we will examine the risk assessment guidelines of two regulatory agencies, the Occupational Safety and Health Administration (OSHA) and the EPA. The OSHA guidelines of 1980 created a great deal of

---

39 NRC (1983). *Risk Assessment in the Federal Government: Managing the Process.* Washington, DC: National Academy Press.

40 NRC (1983). pp. 20-23.

controversy, partly due to the fact that the guidelines were promulgated as binding regulations. Not surprisingly, the NRC risk assessment guide of 1983 advocated the use of non-binding, flexible guidelines. The 1986 EPA risk assessment guidelines for cancer risk assessment are more typical of the various agencies' guidelines.

OSHA risk assessment guidelines

Nearly three years were required to obtain testimony and respond to public comments before OSHA promulgated in 1980 a rulemaking outlining the agency's risk assessment guidelines. In regard to the use of short term tests like the Ames Test, the agency stated that short term tests should not be used as the "*sole* basis for the identification of potential carcinogens....OSHA was...primarily concerned with using short term tests to *confirm* results provided by animal bioassays and not as a *substitute* for bioassay data."[41]

OSHA took a particularly controversial stance in defining the types of evidence that are required to designate a chemical as a Category I Potential Carcinogen, the category of highest regulatory concern. A chemical could be placed into Category I on the evidence of 1) positive human epidemiology studies or 2) a positive result in a single long term animal bioassay that is in concordance with other evidence, including another positive long term animal bioassay or positive results in short term tests.[42]

To accept as equally convincing evidence positive results from two animal bioassays or a single animal bioassay plus two or more short term tests was to give short term tests a large degree of validity. Compared to other risk assessment guidelines drawn up by other regulatory bodies, OSHA's use of short term tests was unique in that the weight given to short term tests was substantial.[43] The response from industry came in no uncertain terms. The American Industrial Health Council, comprised of over 90 companies and 30 trade associations, argued that short term tests are so unreliable as predictors

41 OSHA (1980). "Identification, Classification and Regulation of Potential Occupational Carcinogens," *Federal Register*, Vol. 45, No. 15, January 22, p. 5160.
42 OSHA (1980). p. 5284.
43 Rushefsky, Mark (1986). *Making Cancer Policy*. Albany, NY: State University of New York Press. p. 45.

of human carcinogenicity that they are unsuitable as a basis for regulatory action.[44]

In addition to the potentially large role given to short term tests in the determination of carcinogenicity, the OSHA policy was also unique in its general approach to dealing with uncertainty. Rushefsky characterizes the OSHA as employing a "presumption-rebuttal" approach.[45] OSHA prescribed certain inference options and would allow deviations from them only in certain conditions—conditions that were so strict that they were unlikely to be met.

OSHA's policy has had a troubled history since its promulgation in 1980. In addition to its discussion of the kinds of evidence needed to classify a chemical as a Class I Carcinogen, the policy required that exposure to such carcinogens in the workplace were to be reduced to the maximum extent feasible. In several court cases, challenges were made to OSHA's action to reduce exposure, the level of control required, and economic and technological feasibility. Despite several attempts of revision, the policy is not used by the agency.[46]

The purpose of this section is not to imply that the lackluster results of OSHA's policy has been caused by the agency's use of short term tests in identifying Class I Carcinogens. Much more important were the rigid and stringent measures that OSHA placed on companies where workers were exposed to Class I Carcinogens; the setting of exposure limits and considerations of economic and technological feasibility were the main areas of contention. The use of short term tests reflects the relatively risk-averse stance taken by OSHA, and this was a highly controversial step.

EPA cancer risk assessment guidelines

Perhaps more relevant to our discussion is the cancer risk assessment guidelines of EPA. This agency's guidelines are more in line with the 1983 recommendations of the National Research Council. In contrast to OSHA's policy on risk assessment, the EPA's 1986 final guidelines for cancer risk assessment were not binding regulations, were extremely flexible and general,

44 Maugh (1978). p. 1203.
45 Rushefsky (1986). pp. 56-58.
46 Ashford, Nicholas and Charles Caldart (1991). *Technology, Law, and the Working Environment*. New York: Van Nostrand Reinhold, p. 102.

and did not commit the agency to any particular stance or regulatory action. In addition, the guidelines use a weight of evidence approach in dealing with uncertainty. In this approach, decision makers are to weigh the "entire body of scientific evidence in any specific case."[47] For these reasons, EPA's risk assessment guidelines have been less controversial than OSHA's experience.

As outlined in the guidelines, EPA categorizes the carcinogenic potential of a chemical into several classes. Shown in Table 4.2 is an illustrative categorization of evidence and how that evidence might be used to determine the degree to which a chemical may be considered as a carcinogen. Of primary importance in the categorization are epidemiological studies and long-term animal bioassays. Categorization takes place in three major steps: 1) evidence from epidemiological and animal bioassay studies should be evaluated individually, 2) the "combination of the characterizations of these two types of data into an indication of the overall weight of evidence for human carcinogenicity," and 3) supporting information, including short term tests and chemical structure should be evaluated "to determine if the overall weight of evidence should be modified."[48] The EPA guidelines show that, as a general matter, short term tests have limited influence in the determination of whether a substance is carcinogenic to humans.

Environmental statutes and regulations

The past 25 years have spawned a large number of major environmental laws at the federal level. Much of the recent focus has been on toxic chemicals. Toxic chemicals are the primary focus of two environmental laws, the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) and the Toxic Substances Control Act (TSCA). FIFRA is concerned with the risks posed by pesticides (defined broadly) used in agricultural and household uses. TSCA's scope is more expansive, covering most industrial chemical manufacturing and processing that is not already under the jurisdiction of other statutes. FIFRA's jurisdiction includes some of the most notorious chemicals that have been identified as environmental bad actors, such as the pesticides DDT and chlordecone. TSCA has been the basis for regulations

---

47 Rushefsky (1986). pp. 80-83.

48 EPA (1986). "Guidelines for Carcinogen Risk Assessment," *Federal Register*, Vol. 51, September 24, p. 33996.

Table 4.2 Illustrative categorization of evidence for carcinogenicity

Human evidence	Animal data				
	Sufficient	Limited	Inadequate	No data	No evidence
Sufficient	A	A	A	A	A
Limited	B1	B1	B1	B1	B1
Inadequate	B2	C	D	D	D
No data	B2	C	D	D	E
No evidence	B2	C	D	D	E

Group A	Human carcinogen
Group B1, B2	Probable human carcinogen
Group C	Possible human carcinogen
Group D	Not classifiable as to human carcinogenicity
Group E	Evidence of noncarcinogenicity

Source: EPA (1986).

covering products and chemicals such as asbestos and PCBs. Of particular interest here are the ways in which these two statutes are used to regulate both old and new chemicals and how they rely on scientific data to evaluate the risks posed by chemicals.

FIFRA

EPA has implemented a number of programs under FIFRA, including pesticide registration, setting tolerances for pesticide residues on food, and setting standards to protect workers exposed to pesticides. No pesticide may legally be sold or used in the United States unless it has been registered by the EPA, indicating that the pesticide does not pose unreasonable risks to human health or the environment. Currently, EPA has registered some 25,000

pesticide formulations; however, EPA basis its regulations on these pesticides on their active ingredients, which number less than 750.[49]

EPA regulates new and existing pesticides through different programs. Prior to registration of new pesticides, the registrant must submit to EPA a number of data. Table 4.3 lists the minimum data required. EPA estimates that the manufacturer of a major food-use pesticide could spend up to $10 million for these tests. Furthermore, it can be expected to take six to nine years for a new active ingredient to move from development to retail; two or three of those years are needed to step through EPA's registration requirements. The purpose of this long and costly registration procedure is to document the possible health and environmental effects of a new pesticide, establish residue tolerances for pesticides used on food or feed crops, and, if necessary, restrict its use. Between 1978 and 1991, 130 new active ingredients were registered.[50]

Table 4.3 Minimum data required for FIFRA registration

Chemistry	Toxicology
list of ingredients	acute oral
description of manufacturing process	acute dermal
discussion of formation of impurities	acute respiratory
physico-chemical properties	eye irritation
residue studies	chronic toxicity
metabolic studies	subchronic oral toxicity
analytical methods	reproduction and fertility
results of analytical procedures	metabolism
	mutagenicity
	birth defects
	carcinogenicity
Environmental fate	**Ecological effects**
hydrolysis	aquatic, acute toxicity
leaching	avian, dietary & acute oral
terrestrial dissipation	
photodegradation	
soil metabolism	
rotational crop study	

Source: EPA (1991).

49 EPA (1991). *EPA's Pesticide Programs.* 21T-1005, May, p. 2.
50 EPA (1991). pp. 3-4.

FIFRA also requires that EPA reregister existing pesticides that were registered before the current regulatory standards were in place. Like the new pesticide registration, a large number of tests are required to be submitted; sometimes over 100 studies are necessary. Over 600 pesticides must be reregistered, a process that the EPA estimates will not be completed until 2006 at a cost of several hundred million dollars.[51]

For pesticides that are used on food and feed crops, EPA must set pesticide residue tolerances that will ensure public safety. The data used to determine residue levels are typically based on animal studies for non-cancer and cancer endpoints. The non-cancer "reference dose" is based on the no observed adverse effects level of pesticide dose. Tolerances for cancer risk are not to exceed a 10^{-6} additional risk from lifetime exposure.[52]

It is clear that FIFRA requires the registrant to submit a great deal of scientific data for the EPA to register (or reregister) a pesticide and/or set residue tolerances. When weighing this large set of data, which includes long term animal bioassays—the toxicological "gold standard"—it is unlikely that results from the Ames Test would hold much sway.

This does not mean, however, that the Ames Test has had little impact on the regulation of chemicals under FIFRA. Interviews with scientists in industrial laboratories suggest that the Ames Test is in fact influential. If a new chemical gives a positive result in the Ames Test, the company can choose to continue to perform more sophisticated and more expensive tests, or it can choose to halt development of the chemical. Often, the company will choose not to spend additional resources on testing given the possibility that the other tests required for FIFRA registration will result in implicating the chemical as harmful.

An important distinction emerges here. The influence of the Ames Test can be assessed from either the perspective of EPA or of industry; our conclusions about influence will depend on the perspective taken. From EPA's perspective, the Ames Test has little influence when weighed with a large set of other data in hand. Results from long term animal bioassays will take precedence over Ames Test results in assessing cancer risk. When a company uses the Ames Test, it is usually early on in the testing process, so the results

---

51 GAO (1993). *Pesticide Reregistration May Not Be Completed Until 2006*, GAO/RCED-93-94, May, pp. 2-7.
52 EPA (1991). p. 8.

are usually judged *without* the benefit of a large set of data. The company will choose whether to perform additional testing with what little information it has, perhaps little more than the Ames Test. A positive result may lead the company to halt development or to pursue additional testing. As would be expected, the results of the Ames Test would have relatively less influence later on in the chemical testing process after additional tests have been performed. So we see that the Ames Test probably will not affect EPA's decision making, while the Ames Test may have a large impact during some parts of the chemical testing process within a company.

TSCA

While FIFRA is used to regulate pesticides, TSCA's jurisdiction includes the broader range of industrial chemicals. Given the extremely large number of chemicals under TSCA's purview, the need for quick, accurate, and inexpensive tests would be expected to be particularly useful. In this section we examine the role of the Ames Test in this context.

The Toxic Substance Control Act was approved in 1976 "to prevent unreasonable risks of injury to health or the environment associated with the manufacture, processing, distribution in commerce, use, or disposal of chemical substances." Because of its emphasis on preventing the introduction of new chemicals that are harmful, some have observed that "TSCA was ahead of its time," particularly with respect to the EPA's current focus on pollution prevention. In this brief introduction, we highlight several major TSCA provisions.[53]

First, TSCA §8(b) required EPA to establish an inventory of all chemical substances that are manufactured or processed in the United States. This inventory lists over 60,000 chemicals. This inventory serves as the reference point for an important distinction made throughout TSCA: existing chemicals and new chemicals. Chemicals on the inventory are treated as existing chemicals and are subject to different requirements.

Second, TSCA §5 requires that the manufacture or importation of new chemicals be preceded by the submission of a premanufacturing notice.

---

53 Hathaway, Carolyne, David Hayes, and William Rawson (1994). "A Practitioner's Guide to the Toxic Substances Control Act: Part I," *Environmental Law Reporter News and Analysis*, Vol. 24, No. 5, May, p.10208-10209.

Existing chemicals that are going to be applied in a "significant new use" are also subject to this notification requirement. EPA may require industry to submit additional information. EPA may also restrict or ban the new chemical.

Third, TSCA §4 and §6 allows EPA to require industry to test existing chemicals and to regulate those existing chemicals that pose an unreasonable risk.

In spite of these far-reaching provisions, TSCA has been criticized for its unfulfilled potential. In a 1992 hearing before the U.S. House of Representatives' Subcommittee on Environment, Energy and Natural Resources, Congressman Synar observed in reference to the disappointing implementation of the chemical testing provisions that "never has a law with so much promise gone aground so fast." Congressman Clinger added that "we must focus our attention and very limited resources on those chemicals that present the most risk...."[54] In addition, the U.S. General Accounting Office completed a critical assessment of TSCA in 1994 and found that in the 18 years since TSCA was enacted, only five existing chemicals and four new chemicals have been regulated; the EPA has been relatively more effective in entering into individual agreements with chemical companies to implement exposure reduction measures while more test data is gathered for new chemicals.[55] In the following sections, we examine the implementation of some key provisions and how short term tests like the Ames Test have affected the regulation of toxic industrial chemicals.

Testing existing chemicals

The EPA's existing chemical testing program is responsible for reviewing the risks of over 60,000 chemicals on the TSCA inventory. TSCA established the Inter-agency Testing Committee to recommend to EPA which chemicals should be tested. Chemicals to be tested are chosen on the basis of whether there is sufficient data to determine whether it poses an unreasonable risk and whether testing is necessary to make that determination. Consideration is

54 US Congress (1992). "Toxic Substances Control: Still Waiting All These Years," hearing before the House of Representatives, Committee on Government Operations, Subcommittee on Environment, Energy, and Natural Resources, March 18, pp. 2-3.

55 GAO (1994). *Toxic Substances Control Act: Legislative Changes Could Make the Act More Effective.* GAO/RCED-94-103, p. 15.

also given to those chemicals that are produced in large volumes or may enter the environment and lead to substantial human exposure. EPA has the authority to promulgate chemical test rules that require the manufacturers or processors of the chemical to perform and submit tests.

A limited number of existing chemicals have been tested in either animal bioassays or short term in vitro assays. A study completed in 1984 by the National Research Council characterized the kinds of testing that had been performed for the chemicals on the TSCA inventory. This was done by examining the test data that existed for a small, representative subsample of TSCA inventory chemicals. The NRC estimated that nearly 80 percent of the chemicals had no toxicity information. In addition, mutagenicity tests like the Ames Test had been performed for approximately 10 percent of the chemicals.[56] Table 4.4 summarizes some of their findings. None of these chemicals have enough data to perform a complete health-hazard assessment. The NRC, however, pointed out that a judgment about whether the existence of adequate testing data for a particular chemical may depend on the chemical's intended use.[57] For example, chemicals used in the pharmaceuticals industry have relatively more data due in part to the fact that those chemicals are meant to be ingested. In contrast, humans may be exposed less to the industrial chemicals in the TSCA inventory.

Table 4.4 Test data for existing TSCA chemicals

Production volume	Percent with prescibed test						
	Acute	Subchronic	Chronic	Repro or Develop-mental Biology	Muta-genicity	Percent with minimal tox info	Percent with no tox info
≥ 1 million lb/yr	20	10	4	6	9	22	78
< 1 million lb/yr	15	7	3	7	8	18	82

Source: NRC (1984).

56 NRC (1984). *Toxicity Testing: Strategies to Determine Needs and Priorities.* Washington, DC: National Academy Press, p. 84.
57 NRC (1984). p. 125.

The GAO found that EPA has had limited success in its existing chemical testing program. EPA is responsible for obtaining all available data on chemical toxicity and exposure, a resource-intensive process. EPA has reviewed the risks of about 2 percent of the chemicals that were on the TSCA inventory prior to 1979. With the subsequent addition to the inventory of new chemicals subject to the testing program for new chemicals, about 16 percent of all chemicals currently on the inventory have been reviewed. If EPA finds that additional testing by industry is required, a formal rule must be promulgated. Despite the dearth of both toxicity and exposure-related data with which to determine risk, EPA has not frequently used its authority to require testing. EPA has issued 30 test rules covering 121 chemicals. An additional 59 chemicals were subject to negotiated agreements between EPA and industry, and 230 chemicals have been tested voluntarily by industry. Part of the reason why EPA has not required industry to test more chemicals is that the issuance of a test rule is difficult; it can take over two years and costs between $70,000 and $234,000.[58]

Short term tests have had limited impact in EPA's existing chemical testing program. EPA has attempted to target those chemicals that may present the greatest risk, which will depend on the chemical's toxicity and production volume. The chemicals produced in large volumes will typically have large capital resources invested in them; industry can be expected to resist restrictions on these chemicals. It is not surprising, then, that EPA has had difficulty to promulgating test rules that may implicate chemicals as harmful. EPA would be hard pressed to justify a test rule on the basis of results from short term tests. Even if short term tests cannot form the sole basis for rulemaking, we might expect that they might at least be useful in screening large numbers of chemicals to identify those that warrant more scrutiny in animal bioassays. As the 1984 NRC study showed, however, mutagenicity data exist for few chemicals. Few existing chemicals have been screened by the Ames Test or other short term tests.

58 GAO (1994). pp. 44-55.

Testing new chemicals

Since the initial compilation of the TSCA inventory in 1979, EPA's new chemical testing program has reviewed over 20,000 new chemicals submitted as part of the pre-manufacturing notice (PMN) provisions. The manufacturer or processor of a new chemical must submit a PMN, which should include the chemical identity and structure and exposure-related information such as categories of use, volume of production, and methods of disposal. In addition, the company must submit available test data on health and environmental effects. There is, however, no minimum requirement for the amounts and types of testing that must be performed.

Upon receipt of the PMN, EPA attempts to ascertain the toxicity and exposure potential of the new chemical. EPA has several options. First, it may find that no further action is required and then place the chemical on the TSCA inventory; the chemical is then cleared for use. Second, if EPA believes that there may be some risk, it may allow the company to use the chemical with the stipulation that exposure controls be implemented. If and when additional test data is provided by the company, EPA may find that exposure controls are no longer necessary. Finally, the new chemical could be banned pending receipt of further information. EPA is required to complete its review and arrive at a decision on each new chemical within 90 days after the PMN is received.

EPA's review of PMNs is made difficult by the fact that very little data on health effects and exposure are submitted by the registrant. Table 4.5 shows that PMNs are often submitted with very little or no data. The table distinguishes polymer and nonpolymer chemicals because polymer chemicals are generally less toxic; 52 percent of PMNs are for nonpolymers. Nearly 40 percent of PMNs for nonpolymers are submitted with no test data at all. The data that is included come from short term tests; long-term bioassays are not performed for PMN submissions. Bioassays for mutagenicity, such as the Ames Test are performed for less than one-quarter of new nonpolymer chemicals.

Several reasons for the lack of data in PMN submissions are cited by officials in EPA and industrial laboratories. EPA officials state that companies often are not willing to spend money to conduct testing when such tests are not required for PMN submission. They also state that when test data is

Table 4.5 Test data submitted with PMNs

Type of test data	Percent of PMNs		
	All	Nonpolymer	Polymer
No test data of any type	51	38	68
Health data			
Acute toxicity			
Oral	38	50	22
Dermal	23	29	14
Inhalation	11	14	7
Local toxicity			
Eye irritation	36	47	21
Dermal irritation	38	50	22
Sensitization	11	17	5
Mutagenicity	15	23	6
Other (eg, developmental tox, neurotox, phototox)	11	16	4
Ecotoxicologal data			
Acute lethal vertebrate	6	9	3
Acute lethal invertebrate	3	3	2
Environmental fate data			
Biodegradation	6	8	2
Log P	4	5	1

Source: Auer et al. (1990).

submitted, the tests were performed for reasons incidental to PMN provisions. For example, companies marketing new chemicals in Europe are required to submit a minimum data set of test results. When these chemicals are to be used in the United States, these results will be submitted with the PMN, since TSCA requires that companies submit to EPA all test data that they have available. In addition, testing may have been performed for other statutes such as FIFRA. As one EPA official put it, companies don't perform testing for PMN submissions "out of the goodness of their hearts." One EPA analyst was more critical of industry, going so far as to say that companies simply do not

desire to know anything. They are most concerned with getting the chemical to market quickly and cheaply. Interviews with industrial laboratories also suggest that testing is not performed because they cost money and are not required. Furthermore, chemical testing requires time, and for competitive reasons, companies are reluctant to delay the time it takes to bring the chemical to market.

Despite the paucity of test data submitted voluntarily by companies, EPA must nevertheless assess the likelihood that a new chemical will pose a risk and decide what action must be taken. Table 4.6 shows the frequency with which different actions are taken by EPA. Nearly all new chemicals require no action; they are added to the TSCA inventory. PMNs very rarely result in prohibitions or restrictions. For a small number of new chemicals, more data is gathered or controls are implemented either voluntarily by the company or as required by the EPA. In making these decisions, EPA relies most heavily on two items: exposure and EPA's own analyses of structure-activity relationships (SAR). About 5 percent of all PMNs are flagged for more rigorous review based on these data. EPA officials state that the agency will typically require companies developing chemicals with high exposure to perform a battery of short term tests. This battery usually will consist of several mutagenicity tests (most often the Ames Test), ecological tests, acute toxicity tests, and in some cases a 28-day test. In addition to exposure data, EPA relies heavily on SAR to decide whether companies should perform testing.[59] One EPA official stated that 99% of all PMNs go through some sort of SAR analysis.

It appears, then, that the Ames Test has had a limited influence in EPA's decisions on new chemicals submitted through PMN provisions. First of all, most PMNs do not include Ames Test results. Because of lack of most kinds of test data, EPA relies mostly on exposure data and SAR to decide what actions should be taken on each new chemical. In the small number of cases in which EPA does require further testing, the Ames Test is generally one of the tests required. Although there are some situations where the Ames Test is used, it is not used by EPA as the main factor in its decision making process.

---

59 Auer, Charles, Vincent Nabholz, and Karl Baetcke (1990). "Mode of Action and the Assessment of Chemical Hazards in the Presence of Limited Data: Use of Structure Activity Relationships (SAR) under TSCA, Section 5," *Environmental Health Perspectives*, Vol. 87, pp. 183-197.

Table 4.6 Actions on PMNs, 1979-1987

Actions	Aggregated total to date
Total new submissions received	10842
Valid PMNs received	9132
PMNs requiring no further action	7166
Voluntary testing in response to EPA concerns	149
Voluntary control actions by submitters	45
PMNs withdrawn in face of regulatory action	183
PMNs subject to control, pending data	349
PMNs resulting in prohibition or restrictions	4
Final action taken	
Granted	1473
Withdrawn	147
Denied	12

Source: Ashford et al. (1991).

Most EPA officials interviewed shared this sentiment. Most say that the Ames Test is sometimes useful but not determinative. In judging the influence of the Ames Test at EPA, one official points out that a historical perspective sheds additional light on the subject. She states that in the early 1980s, EPA's office handling PMNs did not recognize the Ames Test as a useful predictor of carcinogenicity. When EPA suggested or required companies to submit additional test data, long term animal bioassays were often requested because of the lack of other informative tests. Often, because of the high time and money costs for long term animal bioassays, the company would decide to halt further development of the new chemical.

She recalls that in the mid 1980s, EPA began using the Ames Test in assessing cancer risks. EPA believed that the test was sufficiently informative to use as a screen. A positive result in the Ames Test would often lead to requests for additional tests from industry. If a chemical had a negative result in the Ames Test, EPA would take "a leap of faith" and accept the PMN submission. In this instance, the Ames Test is influential in the sense that EPA used the test to decide whether further testing was needed or if the chemical required no further action.

By the late 1980s, studies had shown that the Ames Test was less predictive for carcinogenicity for some classes of chemicals. She states that

as a consequence, EPA began using other tests besides the Ames Test. The results of the Ames Test are considered only with other data, never alone. Although the Ames Test is not a driving factor in its decision making, she argued that the PMN process has benefited from the fact that neither EPA nor industry has to rely exclusively on tests that are much more expensive and time consuming. The PMN process is more efficient.

Several EPA officials state that although testing is not required for PMNs, chemical companies perform more testing internally than they would have without TSCA. For example, EPA outlines the chemical classes EPA may be concerned about, why the agency has these concerns, and what tests EPA might wish to have to assess new chemical risks.[60] As one EPA official states, this gives chemical companies an "advanced warning" about EPA's concerns. In this way, EPA has a "silent impact" on the development of new chemicals. This impact may not be apparent from looking solely at PMN submissions.

An important point emerges here. By looking only at EPA's decision making process, we may not be getting the whole story about the influence that the Ames Test has. We should look beyond statutes and regulations and into the use of the Ames Test from the perspective of the decision making within industrial laboratories. A number of interviews were conducted with scientists at industrial laboratories responsible for testing new chemicals. The goal was to ascertain whether the Ames Test had a "silent impact" within industry.

A number of company officials state that one way that the Ames Test may have a silent impact is in comparing Ames Test results of several substitutable chemicals. For example, if there were two new chemicals that could be substituted for each other, but one is positive in the Ames Test and the other is not, then the test results could be used as the basis for halting the development of the mutagenic substitute and developing the other. Decisions such as this occur internally to the company and are not typically observable by EPA or the public. Although this certainly would be a clear way in which the test would be influential in decision making, industry officials say that this does not happen very often. In fact, this possibility was typically described in abstract terms rather than with any specific examples.

---

60 EPA (1995). *New Chemicals Program Chemical Categories*, produced by the Office of Prevention, Pesticides and Toxic substances.

Judging the influence of the Ames Test in a company's decision making is more difficult than might be suggested by the example of the two substitutable chemicals. Rarely are decisions cut-and-dried. Some company officials state that the influence of the Ames Test should not be inferred from the fraction of new chemicals that are tested for mutagenicity. The Ames Test is not applied mechanically to all new chemicals. There are three main factors that appear to determine whether the Ames Test is used for a specific chemical. First, companies often use SAR analysis (sometimes qualitatively) to determine whether the Ames Test might provide informative results. The test is more predictive for some chemical classes than others. Second, companies consider the estimated exposure potential of the chemical. Chemicals that will have low exposure potential, perhaps because it will be used only as an intermediate in processing and not incorporated into products that may enter the environment. The third and probably most important factor that a company considers is the likely actions of EPA. Before performing tests, a company will try to get an initial "read" on what the agency might do. Will testing be required? Which tests will EPA want? The goal is for a company to avoid any testing that the EPA might not ask for. As one scientist from an industrial lab stated, they "want to do the least possible testing" that will satisfy EPA.

The decisions about what should be done with Ames Test results are not made in a mechanical fashion as suggested by the previous example of two substitutable chemicals. In general, most people interviewed stated that a negative result from the Ames Test typically would not be sufficient grounds for halting the development of an industrial chemical. On the other hand, a positive result may or may not lead to additional testing. The most important factor in decisions to perform more tests is whether EPA requests or formally requires additional testing. Companies use what can be described as a "wait and see" approach. Rather than automatically performing additional testing after a positive result in the Ames Test, a company is likely to be more inclined to wait and see if EPA will require more testing. And as was discussed above, EPA considers most heavily in PMN review factors such as exposure potential and SAR. This wait and see approach is manifested in the disposition of PMNs as described in Table 4.6. For the small number of chemicals that require further action beyond the submission of the PMN, EPA must ask for additional

information. Companies will either perform more testing, implement control measures, or withdraw the chemical from further review.

Conclusion

This chapter has traced the history of the Ames Test, from the early development of the public's concern about cancer to its current use in EPA and industrial decision making. More than two decades have passed since the War on Cancer was begun. In that time span, the public's attitude about environmental and health problems has developed and matured. It may be difficult to grasp the magnitude of the public's concern about cancer during the late 1960s and early 1970s.

It is clear, however, that the national mood for an intensive campaign to fight cancer was strong. Encouraged in no small part by full page advertisements in newspapers and popular columnists like Ann Landers, the public exerted great pressure on policy makers to cure cancer. The similarity, in both effort and expected results, was explicitly made between America's mission to put a man on the moon and the fight against cancer.

It is no surprise, then, that great hopes were pinned on the early development of the Ames Test. Here was a simple and inexpensive way to find out which chemicals were causing cancer. By the mid-1970s, scientists argued that the Ames Test had "been sufficiently developed and validated to be seriously considered" for use a screening test to complement long term animal bioassays. At the same time, others were more critical of the relevance of the Ames Test to carcinogenicity. Much of the debate about how to use the Ames Test in the regulatory setting, such as OSHA's cancer policy, revolved around validation studies of the test.

With time and use, and a great deal of controversy, the Ames Test has matured. While it is useful in some circumstances, it does not appear to be as influential as first hoped. It is not used by regulators or industry as an automatic test to screen all chemicals. It has a limited role in EPA's cancer risk assessment guidelines. It is used in FIFRA and TSCA regulations, but it is used only in specific situations. By looking at PMN submissions and industrial testing programs, one questions whether the Ames Test has had much impact. What, then, has the influence of the Ames Test been?

The meaning of influence

In assessing the influence of the Ames Test one is likely to point to the experience of furylfuramide (AF-2), a food additive used extensively in Japan.[61] AF-2 showed no carcinogenic activity in rodents in 1962 and 1971. In 1972, it was shown to be mutagenic in cultured human cells and was later found to be mutagenic in the Ames Test. More thorough animal testing was done, showing the carinogenicity of AF-2, resulting in a prohibition of its use in Japan. This is a dramatic instance of how the short term mutagenicity tests influenced decision making. This probably does not happen frequently, however. Narrowly focusing on examples like this does not fully contemplate the different types of influence that the Ames Test has had. This chapter has brought to light several issues that make answering this question difficult.

The first important issue is that the influence of the Ames Test depends on the perspective. As noted in the discussion about FIFRA, when EPA weighs the volumes of test data for a new pesticide, the agency will have in hand information on mutagenicity, multigenerational tests, long term animal bioassays, and a number of other data. It is unlikely that results of an Ames Test would have much influence. But from the perspective of a company, it can be argued that the Ames Test is more influential. Preliminary testing will often include an Ames Test. At the early stages of testing, the company will have to decide whether or not to continue development of the pesticide based on whatever little information it has at that point in time. In these circumstances, the Ames Test can have a large impact on development of new products.

Secondly, assessment of the Ames Test takes place in a historical context. As pointed out by one EPA official, EPA's new chemical testing program under TSCA would be far different today if the Ames Test and other short term tests were not available or not used. The alternatives to using the Ames Test include test systems that are much more expensive and time-consuming. These other test systems would make the testing process slower and may inhibit the development of new chemicals if industry is reluctant to perform these other tests. Compared to the past alternatives, the Ames Test has changed the way that chemical testing is performed.

---

61 Ames (1979). p. 590.

The third important issue is that the influence of the Ames Test is not necessarily or exclusively defined by the characteristics of the test per se. While the predictive power of the test is an important element of how the test results will influence decision making, it is by no means the only factor that affects decision making. In the PMN provisions of TSCA, EPA has very little data with which to judge the risk associated with new chemicals. We might expect that short term tests like the Ames Test would be of great importance to both EPA and companies. As we have seen, however, companies perform the Ames Test on only a fraction of new chemicals, and EPA's decision making relies most heavily on SAR and exposure potential.

This does not mean that the Ames Test has no influence in TSCA. Measures of influence extend beyond the frequency with which the test is used, the number of times that chemical development has stopped, or the number of times that EPA prohibits a new chemical. It is necessary to understand that many chemicals submitted as PMNs are chemical process intermediaries that are not incorporated into a product that is shipped to consumers. Since risk is a function of both hazard (e.g., toxicity) and exposure, companies can reduce exposure, and thus risk, by designing closed chemical processes. That is one reason why EPA's decision making relies heavily on exposure estimates. Dropping a new chemical from further development is not the only way to control risk, and in the case of PMNs, it may not even be the primary way. The influence of the Ames Test is subtle in these circumstances.

The fourth issue is the most obvious. Both company and EPA officials typically said that the Ames Test is useful, but the test by itself is not enough to be determinative. That does not mean that the Ames Test is not influential. Searching for direct cause-and-effect relationships between the Ames Test and a judgment about a chemical's risk is likely to be frustrating. The Ames Test is a contributing factor to decision making.

Chapter Five

Lessons for the E-screen

Given the suggestive evidence of the deleterious effects of estrogen-mimicking chemicals in the environment, preventing the release of these and other endocrine disrupting chemicals could be an important step in improving human health and the environment. The identification of endocrine disruptors using short term tests could have significant implications. Empirical evidence of the test's performance appears promising. Consequently, regulators and industry are examining the possibility of using them routinely in a testing strategy.

What role will short term tests play? Using the E-screen as an example, this research attempts to understand the factors that are likely to affect whether and how short term tests for endocrine disruptors will be used by government and/or industry. The case study on the Ames Test presented in the previous chapter clearly illustrates some of the problematic aspects of developing and using short term tests. At the time of its inception, the Ames Test was believed to hold great promise for identifying carcinogens. The extensive scrutiny given to the test later resulted in a maturation of the expectations that regulators and industry had for it. No longer did anyone argue that "carcinogens are mutagens." Currently, the Ames Test holds a position of limited influence as a screening tool in the identification of carcinogens.

In this chapter we will discuses in turn three major themes that stand out from the experience of the Ames Test. First, the public attitudes o other testsablished a great policy need for the development of a short term test like the Ames Test. In retrospect, we see that the pressure from this policy need led, at least initially, to expectations that were over optimistic. Consequently, it is not surprising that the subsequent history of the Ames Test shows disappointing results. In regard to the E-screen, the evidence for a policy need

is not as clear. Second, the meaning and determination of the validity of the Ames Test was a crucial aspect of its use. Validation studies came to different conclusions about its ability to identify carcinogens. These validation studies were influenced greatly by their design and purposes. The validation of the E-screen will also be of fundamental importance to the way and extent that it is used. Third, implementation of the Ames Test was affected by the existing statutory and regulatory framework for dealing with chemical carcinogens. The use of the Ames Test did not evolve in a vacuum, but rather it found its application in certain niches defined by existing laws and regulations. So too will similar factors affect how the E-screen fits into the current framework for risk assessment.

Estrogen-mimicking chemicals as a policy problem

The potentially adverse effects of estrogen-mimicking chemicals are increasingly identified as a key threat to public health and the environment. Several groups of people are attempting to increase the awareness of the issue, and hence its importance in policy. An influential lobby supporting research and prevention of breast cancer has identified this as an issue of concern and is pushing for the control of estrogen-mimicking chemicals. In addition, scientists in diverse fields such as wildlife biology, toxicology, and endocrinology are advocating greater scientific, political, and regulatory emphasis on the issue. Despite a growing number of reports in the media and the scientific community, it does not appear that estrogen-mimicking chemicals have a prominent place in the current agenda of national policy. Concern for estrogen-mimicking chemicals has not been of the same magnitude as seen in the growth of the powerful cancer establishment that Mary Lasker helped to create in the late 1960s.

Public concern for estrogen-mimicking chemicals

The public is being exposed to increasing numbers of scientific and media reports about estrogen-mimicking chemicals. No longer is it a subject exclusively found in esoteric technical journals. Articles have appeared in

magazines such as *Science*,[1] *Scientific American*,[2] and *Chemical & Engineering News*.[3] Even in publications such as *Newsweek*[4] and *The Washington Post*,[5] the public is receiving news that estrogen-mimicking chemicals could be responsible for cancer and reproductive and developmental defects. While most reports suggest that very little about estrogen-mimicking chemicals is known for certain, the potential for great danger exists. For example, the article that appeared in *The Washington Post* quotes a "pioneer in the field of estrogen chemistry" who says that "scientifically, the potential for harm from these [environmental] estrogens is still really not known. But the potential is so huge for all kinds of reproductive problems that, by God, we have to get after this problem."

The public's response to these media reports is difficult to predict. The public attitude towards environmental risks is notoriously fickle. In some instances, the public appears jaded when confronted with media reports about a new addition to the "carcinogen-of-the-month club."[6] The public may react with indifference, assuming that the environmental scare will soon be contradicted by another scientific study. People may also choose to ignore some reports because if they stopped every risky activity or stopped eating every suspected hazard they would end up not doing or eating anything at all.

The public's desensitization to some environmental risks cannot be dismissed simply as foolish behavior. There is some evidence that cancer risks from industrial chemicals found in the environment are very small compared to naturally occurring carcinogens found in nature. Michael Gough argues that since the 1970s, scientists have believed incorrectly that 80 to 90 percent of all cancer cases could be attributed to environmental causes. He believes that environmental pollution causes only 2 to 3 percent of all cancers in the United States; the current regulatory system that focuses on environmental

---

1 Stone, Richard (1994). "Environmental Estrogens Stir Debate," *Science*, Vol. 265, July 15, pp. 308-310.

2 Rennie, John (1993). "Malignant Mimicry," *Scientific American*, September, pp. 34-35.

3 Hileman, Bette (1994). "Environmental Estrogens Linked to Reproductive Abnormalities, Cancer," *Chemical & Engineering News*, January 31, pp. 19-23.

4 Begley, Sharon (1994). "The Estrogen Complex," *Newsweek*, March 21, pp. 76-77.

5 Weiss, Rick (1994). "Estrogen in the Environment: Are Some Pollutants a Threat to Fertility?" *The Washington Post*, January 25.

6 Wildavsky, Aaron (1990). "No Risk is the Highest Risk of All," in *Readings in Risk*, Glickman and Gough, eds., Washington, DC: Resources for the Future, p. 121.

carcinogens will do little to control cancer.[7] Interestingly, another advocate of this position is the developer of the Ames Test, Bruce Ames. Based on epidemiological evidence, he argues that the public incorrectly thinks that synthetic pollutants are a major cause of cancer; more important factors include diet, tobacco smoking, and oxidants.[8] In *The New York Times*, Ames bluntly stated that "pollution [is] mostly a red herring as a cause of cancer."[9]

This should not suggest, however, that the public no longer cares about environmental risks. When the media reported that the use of the synthetic pesticide Alar on apples led to cancer, the public reacted swiftly and strongly against its use. Public outcry came on the heels of an effective media campaign of the Natural Resources Defense Council (and the public relations firm it hired for this issue), an alarming report on the television program *60 Minutes*, and highly publicized statements by well known individuals (such as actress Meryl Streep). Within four months of the *60 Minutes* report, many school systems removed apple products from their menus, consumers demanded that grocery stores sell organic produce, the apple growing industry reported losses of $100 million, and the manufacturer of Alar suspended sales of the pesticide.[10] Some scientists have argued that the risks of Alar were far overstated.[11] The important point here is not whether Alar indeed posed a serious risk, or whether the media reported the case responsibly, but rather that the public still sometimes shows great concern for chemical risks found in the environment.

The E-screen should not be expected to be accepted and implemented overnight, or even as quickly as the Ames Test. If we compare the current concern for estrogen-mimicking chemicals with the concern shown for cancer during the time that the Ames Test was being developed, public concern of the magnitude shown for cancer during the 1970s—the height of the War on Cancer—has yet to materialize for estrogen-mimicking chemicals. As was

---

7 Gough, Michael (1990). "How Much Cancer Can EPA Regulate Away?" *Risk Analysis*, Vol. 10, No. 1, pp. 1-6.

8 Ames, Bruce, Lois Gold, and Walter Willett (1995). "The Causes and Prevention of Cancer," *Proceedings of the National Academy of Sciences USA*, Vol. 92, pp. 5259-5262.

9 Brody, Jane (1994). "Strong Views on Origins of Cancer," *The New York Times*, July 5, pp. C1, C10.

10 Rosen, Joseph (1990). "Much Ado About Alar," *Issues in Science and Technology*, Fall, pp. 85-90.

11 Rosen (1990). pp. 85-90.

discussed in Chapter 4, the public outpouring of support for cancer research during the 1970s was massive enough to help mobilize major legislative action and scientific research. Without public pressure for regulators and industry to identify estrogen-mimicking chemicals, it is unlikely that the E-screen would be implemented quickly and used widely. We should not expect the E-screen to be implemented on a large scale just because the assay exists. Although the E-screen may in fact become an integral part of chemical testing, this should by no means be expected to happen quickly without the presence of public concern for estrogen-mimicking chemicals.

It is difficult to predict how the public will respond to scientific findings about estrogen-mimicking chemicals. It is unclear whether the public will react with skepticism, indifference, or with grave concern. It cannot be taken for granted that the public will demand legislative and scientific action on estrogenic chemicals.

It should not be assumed, however, that anything less than strong public pressure will preclude the adoption of the E-screen. In fact, less public pressure could be a positive factor in its implementation. Given the national fervor for the cure and prevention of cancer during the 1970s, it is not surprising that the Ames Test was quickly adopted or that Bruce Ames would optimistically state that "Carcinogens are Mutagens." The Ames Test certainly had to live up to high expectations. With less public pressure to quickly find the "silver bullet" to the problem of estrogen-mimicking chemicals, the E-screen may be developed more slowly but perhaps in a more reasoned and methodical way.

Regulatory concern for reproductive and developmental effects

The suspected effects and targets of estrogen-mimicking chemicals (and endocrine-modulators more generally) span a wide range, from wildlife to humans, from infants to mature adults, from cancer to reduced reproductive success, and immunological and neurobehavioral effects. Besides cancer, particularly breast cancer, most of the concern for estrogen-mimicking chemicals stems from their potential reproductive and developmental effects.

While some people in the scientific community and the media have expressed great concern for the issue of reproductive and developmental effects of estrogen-mimicking chemicals, the regulatory arena has not looked

towards those effects as the basis for regulations. The development and use of the E-screen is likely to be affected by the degree to which regulators recognize estrogen-mimicking chemicals as a legitimate policy problem. If the regulatory concern for the suspected reproductive and developmental effects of estrogen-mimicking chemicals fails to materialize, the use of the E-screen may not be embraced.

Several environmental statutes authorize regulatory measures based on reproductive and developmental health. For example, such language can be found in TSCA, the Clean Water Act, and the Comprehensive Environmental Response, Compensation, and Liability Act (Superfund), while other statutes like FIFRA do not explicitly state these health effects but have a broad charge to protect human health. Nevertheless, a GAO study found that reproductive and developmental health generally has not formed the basis for most regulation; concerns for cancer and acute toxicity have historically been the most important factors. And while it is sometimes assumed that the control of cancer risks will at the same time protect the public from reproductive and developmental effects, the GAO says that this may not be the case.[12]

Through literature reviews and surveys of expert opinion, the GAO identified environmental chemicals that are generally believed to be of highest concern for their effects on reproduction and development. A list of 30 substances was compiled, and it included chemicals such as chlordecone, DDT, DES, mirex, PCBs, and dioxin. In their analysis of the regulations for these 30 chemicals of high concern, the GAO concluded that cancer and acute toxicity played the major role in their promulgation. Less than one-third of the regulations for this set of chemicals was based to any extent on reproductive or developmental health. Furthermore, even in these cases, the GAO noted that the focus is still often on cancer. In fact, one agency official stated that his agency misled the GAO in implying that reproductive and developmental effects were considered at all.[13]

Despite the relatively low regulatory focus on reproductive and developmental effects, the issue of estrogen-mimicking chemicals is gaining more attention. In April 1995, the EPA sponsored the Endocrine Disruptors

---

12 GAO (1991). *Reproductive and Developmental Toxicants: Regulatory Actions Provide Uncertain Protection.* GAO/PEMD-92-3, October.
13 GAO (1991). p. 44.

Research Needs Workshop.[14] Although the workshop touched on many facets of the endocrine system, much of the focus was on estrogen in particular. Nearly 100 panelists were invited to discuss research needs in assessing the risks of endocrine disruptors. Approximately 200 observers also were present. In attendance were scientists from academia, government, industry, and public interest groups. The stated goals of the workshop were 1) to provide a forum for communication of information on endocrine disruptors among a diverse assembly of organizations and scientific specialties, and 2) to develop a national research strategy that delineates short term and long term projects necessary for understanding the magnitude and nature of endocrine disruption.

At this workshop, many researchers asserted that very little money is available to study the issue. Some speakers noted that the issue spans a great number different disciplines, from wildlife biology to microbiology. Within each discipline scientists are finding it difficult to obtain funding for research. The importance of the issue cannot be appreciated within individual disciplines; the forest for the trees has not been noticed by organizations funding research. EPA officials at the workshop stated that the agency is in the process of securing more funding for scientific research on endocrine disruption issues. In the current climate of budget-cutting in the federal government, there is great uncertainty in the level of funding that might be available.

At EPA headquarters in Washington, D.C. an agency-wide working group has been established to coordinate agency efforts relating to endocrine disruption. The chairman of this working group, Steven Devito, stated that the EPA is still at the very early stages of understanding the issues. Currently, the agency does not have sufficient rationale to regulate chemicals based on concerns about endocrine disruption. EPA is attempting to ascertain whether adverse health effects in humans and wildlife are caused by endocrine disruptors.[15]

The EPA, along with the Department of Interior, has also requested the National Research Council (NRC) to conduct a thorough, two-year study of endocrine disruptors. The NRC plans to review critically the literature, identify

[14] EPA Endocrine Disruptor Research Needs Workshop, April 10-13, 1995, sponsored by EPA's Health Effects Research Laboratory in Research Triangle Park, North Carolina.
[15] Devito, Steven (1995). Personal communication.

the responsible mechanisms of action, investigate the impacts on humans and the environment, and recommend critical areas for further research.[16]

Industry is also engaging in more research on this issue. The Chlorine Chemistry Council, an industry organization representing chemical companies involved in the use or manufacture of chlorine, has brought together industry professionals concerned about endocrine disruptors. This group is headed by Ron Miller of Dow Chemicals' Regulatory Affairs branch. The CCC has also commissioned a study to review the state of the art in short term in vivo and in vitro tests for estrogen-mimicking chemicals.[17]

Another industry organization, the Chemical Industry Institute of Toxicology (CIIT), has begun investigating this issue. CIIT conducts scientific research and is funded primarily by its 49 member chemical companies and receives additional support from government research grants. Currently, approximately 10 percent of its research funding is now being devoted to understanding endocrine disruption.

And at laboratories at individual companies, research is being conducted on this issue. Most interviewed scientists, both in industry and government, could not, however, quantify the funds being devoted to this topic. One reason for this is the endocrine disruption spans many different research topics and is only now being considered in a unified manner. For example, a company may already have research programs on carcinogenic, reproductive, and developmental effects, and the chemicals studies may be hormonally active. This makes it difficult to judge exactly how many resources are being spent on endocrine disruption.

Validation of the E-screen

One of the most critical lessons that can be learned from the case study of the Ames Test is the importance and problematic aspects of validation. The public pressure for cancer prevention and early studies of the Ames Test vaulted expectations of the test to a position that could not endure the scrutiny of the later validation studies. Some of the validation studies of the Ames Test may have been flawed or misleading. The objectives of a validation study and

16 NRC Commission on Life Sciences (1995). "Hormone-related Toxicants in the Environment," study prospectus.
17 Miller, Ron (1995). Personal communication.

the sample of chemicals chosen to be used in a validation study will significantly influence the conclusions.

The current knowledge about the E-screen is at a stage not unlike the Ames Test was in the late 1960s and early 1970s. The E-screen has been applied to a small number of chemicals, early results appear to be promising, and it has been suggested that it should be used in government and industrial settings. As the case study of the Ames Test vividly illustrates, the preliminary results from the E-screen are unlikely to be sufficiently persuasive for the test to be incorporated into chemical testing without further study. The case study suggests that there are several important issues relating to validation of short term tests that must be considered in evaluating the likely impact of the E-screen. In this section, we will examine the meaning of validation, the importance of an established test protocol, the selection of standards to which short term tests are compared, and the choice of chemicals used in validation studies.

The meaning of validation

Scientists, regulators, and industry can be expected to ask whether the E-screen is valid before its results are used as a basis for decision making. Disagreements may emerge from a misunderstanding of what validation means. This is illustrated in the testimony given during deliberations for OSHA's cancer policy that was described in Chapter 4. It became apparent that the witnesses, mostly scientists, had different conceptions of what validation meant, and whether the Ames Test could be considered as validated.

To some, validation meant the establishment of a *correlation* between the Ames Test and long term animal bioassays. To others, validation referred to the *standardization* of the test protocol. Some witnesses used validation to refer to whether the Ames Test exhibited *reproducibility*. Even the idea of reproducibility was ambiguous. Some thought reproducibility referred to agreement among different short term tests. Others believed that it meant that the test results could be repeated within the same laboratory.

OSHA stated that "it is clear...that much of the divergence of opinion about the state of validation of [short term] tests reflected different uses of the

term rather than different interpretations of the scientific data."[18] While the ambiguity of meanings contributed to the divergence of opinion, OSHA was too simplistic in their reasoning. Different interpretations of the scientific data, not just the different meanings of validation, probably had a great deal to do with the scientists' assessment of whether the Ames Test was validated. Determining whether a short term test is sufficiently validated involves subjective judgments about data. Does 95 percent concordance with long term animal bioassays suggest a validated test? Does 85 percent concordance indicate that the Ames Test is not validated? Although science is necessary to calculate the concordance, science cannot provide an answer as to whether a test is sufficiently valid.

The use of structure-activity relationships (SAR) in TSCA illustrates that the validity of a test or analysis method depends on the situation. SAR is used frequently by EPA in reviewing the potential risks posed by new chemicals in the pre-manufacturing notice (PMN) provisions under TSCA. In the absence of little data beyond chemical structure, EPA will often use SAR to ascertain a new chemical's likely hazard. EPA can require a manufacturer to implement measures to control exposure until the manufacturer provides additional information about the chemical. In this use of SAR, it is considered a sufficiently valid method of analysis. In contrast, if the same methodology was used as the basis for a restricting the use of an existing chemical used widely, industry would almost certainly question whether SAR was validated.

Validation studies of the Ames Test were frequently undertaken, yet little effort went into defining how validation should be performed. In the 250,000 pages of testimony given for the OSHA cancer policy, "the only clear description" of how to validate a short term test was given by Dr. Umberto Saffiotti of the National Cancer Institute.[19] He identified several important elements that should be considered. First, laboratory procedures and measurement of their reproducibility should be defined and standardized. Second, a large list of reference chemicals should be selected to be used with the test. The list should include representatives of a wide variety of chemical classes and a sufficiently large number of noncarcinogens. The chemicals on the list should also be evaluated for their reliability in terms of data from tests

[18] OSHA (1980). "Identification, Classification and Regulation of Potential Occupational Carcinogens," *Federal Register*, Vol. 45, No. 15, January 22, pp. 5165-5166.
[19] OSHA (1980). p. 5165.

that are going to used as the standard (e.g., long term animal bioassays). Third, systematic screening should be undertaken using a double-blind procedure. Discussion of these issues will be included in the sections below.

Protocol standardization

A study attempting to determine the correlation between the E-screen and long term tests in whole animals may rely on a review of the test results from various laboratories. Aggregating these data may not lead to informative conclusions if the protocol for the test is not consistent among laboratories.

The experience of the Ames Test shows that the importance of this apparently obvious point is easy to overlook. In 1978, Marvin Legator observed that results obtained from the Ames Test are highly variable from laboratory to laboratory because of differences in the way that the tests are performed. Joyce McCann, a collaborator of Bruce Ames, agreed with Legator's assessment and advocated greater use of standards among the various laboratories.[20] Even by the mid-1980s, some 15 years after Ames developed his test, the use of a standard protocol for the Ames Test was not taken for granted in validation studies. In the 1987 study by Tennant et al., the authors point out that their study differed from most previous validation studies in that it relied on the use of the Ames Test performed under a standard protocol.[21]

Protocol standardization, however, should not necessarily take place at the earliest possible moment. Experiments performed with variations allow for the optimization of test conditions and provides insight into the test system itself. Discouraging this type of experimentation by the establishment of a standard protocol may inhibit the development of the test. In the case of the Ames Test, scientists often varied the test conditions in order to make the test more sensitive to various chemical classes and to better reflect the different types of metabolic activity that take place in whole organisms.

Currently, Soto is in discussion with the American Society for Testing and Materials (ASTM) to establish guidelines for the use of the E-screen.

---

20 Maugh, Thomas (1978). "Chemical Carcinogenesis: The Scientific Basis for Regulation," *Science*, Vol. 201, September 29, p. 1203.

21 Tennant, Raymond, Barry Margolin, Michael Shelby, et al. (1987). "Prediction of Chemical Carcinogenicity in Rodents from in Vitro Genetic Toxicity Assays," *Science*, Vol. 236, May 22, p. 933.

Diana Hentschel of ASTM says that these guidelines will be an important step that will encourage laboratories to experiment with the test. Very few standards or guidelines for biomarkers have been established, and there is a desire to create more. At least one company has expressed interest in the guidelines for the E-screen, says Hentschel. She believes that laboratories have more confidence in assessing and using the test if guidelines have been established; without guidelines companies "have a tough time grappling with" the test's methodology and interpretation of results. She says that ASTM guidelines and standards also can influence regulators; EPA tends not to accept results from tests for which guidelines or standards do not exist. In addition, since EPA standards do not yet exist for the E-screen, ASTM guidelines are likely to influence the form of EPA's standards.[22]

Hentschel says that it is likely that ASTM will establish guidelines rather than a strict protocol for the E-screen. The science and technology for biomarkers is constantly in a state of flux; flexibility is an important attribute in these circumstances. Hentschel anticipates that the guidelines for the E-screen will include terminology, methodology, guidance for statistical analysis, and caveats for the test's use and interpretation. For this to be established as an ASTM guideline, a committee must vote on their acceptability. The committee consists of scientists mainly from industry and government and some from academia as well. ASTM encourages qualified persons to serve on the committee. Although some standards can take several years to establish, Hentschel believes that the guidelines for the E-screen will not be controversial or take very long because laboratories are anxious to have guidance.[23]

Standards of comparison

As discussed above, one type of validation study attempts to determine whether the results of the E-screen are correlated with some other type of data that is considered to be the standard of comparison. In the realm of cancer risk assessment, human epidemiology has had limited ability to ferret out carcinogens. Only some 20 chemicals have been found to be carcinogens through epidemiological studies. Because many more chemicals have been studied in animal models such as mice and rats, long term animal bioassays

22 Hentschel, Diana (1995). Personal communication.
23 Hentschel (1995). Personal communication.

serve as the most common data with which to judge the ability of the Ames Test to predict carcinogenicity. It is important to remember that studies in rodents are only proxies for what is usually of most concern—human health. Species differ, sometimes dramatically, in their metabolic and pharmacokinetic characteristics. Although one can question the relevance of rodent tests to human health effects, long term animal bioassays have been established as a standard of comparison.

It is not clear what will be used as the standard of comparison in validation studies of short term tests for estrogen-mimicking chemicals. What *is* an estrogenic chemical? One of the classical definitions of an estrogen is that an estrogen is an agent that causes the mitotic stimulation of the tissues of the female genital tract. Often this mitotic stimulation is observed typically through weight changes of the uterus (uterotropic activity) in rodents. Hertz argues that although estrogens are responsible for a wide range of responses, "the sine qua non of estrogenic activity remains" the mitotic stimulation of the female genital tract.[24] Uterotropic activity, however, is not a specific response of estrogen.[25] That is, there are other factors besides estrogen that can induce uterotropic activity.

This classical definition of estrogen is the basis for the claim that the E-screen is able to identify estrogenic chemicals. Soto states that cell proliferation observed in the E-screen "is recognized as biologically equivalent to the increase of mitotic activity in the rodent endometrium."[26] This is not to say, however, that such a chemical could be or is responsible for the well-publicized adverse effects in human populations, such as breast cancer or low sperm quality. Asking whether a chemical is estrogenic is a different question than asking whether a chemical is causing these adverse effects in human populations. The E-screen answers only the former question, Soto argues, while epidemiology can answer the latter question.[27]

---

[24] Hertz, Roy (1985). "The Estrogen Problem: Retrospect and Prospect," in *Estrogens in the Environment II*, John McLachlan, ed., New York: Elsevier Science Publishing, p. 1.

[25] Soto, Ana, Tien-Min Lin, Honorato Justicia, Renee Silvia, and Carlos Sonnenschein (1992a). "An 'In Culture' Bioassay to Assess the Estrogenicity of Xenobiotics (E-screen)," in *Chemically-Induced Alterations in Sexual and Functional Development: the Wildlife/Human Connection*, Colborn and Clement, eds., Princeton, New Jersey: Princeton Scientific Publishing, p. 296.

[26] Soto, Ana, Carlos Sonnenschein, Kerrie Chung, Mariana Fernandez, Nicolas Olea, and Fatima Olea Serrano (in press). "The E-screen Assay as a Tool to Identify Estrogens: An Update on Estrogenic Environmental Pollutants," *Environmental Health Perspectives*.

[27] Soto, Ana (1995). Personal communication.

Some scientists, however, believe that using uterotropic activity as a standard of comparison for the E-screen limits the test's use in regulatory and industry decision making. Donna Farmer, a developmental and reproductive toxicologist at Monsanto, argues that the public is concerned about chemicals that could be causing breast cancer and low sperm quality, not chemicals that cause uterotropic activity. There is a misperception, encouraged by press reports, that the E-screen has the ability to identify chemicals that cause breast cancer and low sperm quality. She believes that a more appropriate standard of comparison for the validation of the E-screen is long term in vivo tests like a multigenerational test, which examines a wide variety of effects from chemicals in several generations of rats or mice.

Using multigenerational tests (and other long term tests in whole animals) as the standard of comparison for the E-screen presents several problems, however. First, such a comparison may be overstepping the scientific basis of the E-screen. Just as the Ames Test identifies mutagens, but not necessarily carcinogens, the E-screen identifies chemicals that cause uterotropic activity, but not necessarily effects in multigenerational studies or adverse effects in human populations. It is critically important to understand what the E-screen can and cannot do, and attributing to the E-screen capabilities it does not have serves only to cause misunderstandings and unfulfilled expectations.

A second problem with using multigenerational studies as the standard of comparison is that they have different sensitivities than the E-screen in detecting estrogenic chemicals (as defined by uterotropic activity). One of the advantages of the E-screen is that it is highly sensitive, enabling it to detect the effects of even minute concentrations of test chemicals.[28] The sensitivity of multigenerational tests to detect effects, albeit effects very different from those seen in the E-screen, is limited by the statistical power of animal bioassays with relatively few subjects. For example, consider that in cancer risk assessment, the National Toxicology Program protocol for long term animal bioassays requires 600 animals. The largest experiment ever conducted used 24,000 rodents but was not sensitive enough to detect less

28 Soto et al. (1992a). p. 306.

than a one percent increase in tumor incidence; most animal cancer studies are only capable of detecting risks or perhaps one-in-a-hundred.[29]

The choice of a standard of comparison will effect how the E-screen may affect decision making. As discussed in Chapter 4, the regulation of existing chemicals produced in large volumes can be expected to require the most sophisticated evidence of a link between a chemical and adverse effects. If there is convincing evidence that the E-screen is well correlated with multigenerational tests, then the E-screen could be used in circumstances in which sophisticated multigenerational tests are desired to take regulatory action. It seems, at this point, that such a convincing correlation cannot yet be made, if ever. On the other hand, some decisions may require less sophisticated testing. Examples include decisions to prioritize which chemicals should be scrutinized in more sophisticated tests and decisions about which new chemicals a company should continue to develop. If uterotropic activity is a valid concern for these decisions, then the E-screen, if correlated with uterotropic activity, could be influential in these circumstances. Interviewed scientists from both industry and EPA stated that uterotropic activity and perhaps the E-screen could in fact be important areas of concern in developing testing priorities and new product development.

This suggests that there is not any inherent standard of comparison that must be used in validation studies of the E-screen. The standard will depend on what the E-screen will be used for. It is not necessary that the only appropriate way to validate the test is to use the most sophisticated standard like multigenerational tests. The standard(s) of comparison should match the purpose(s) for which the E-screen is used.

Selection of chemicals

The predictive power of a short term test is calculated using correlation studies of a sample of chemicals. Great care is required to generalize an observed correlation beyond the sample; such inferences depend strongly on the number, type, and variety of chemicals examined in the sample.

One frequently overlooked characteristic of the chemical sample is, in the case of the Ames Test, the number of noncarcinogens. If the relative

---

[29] Masters, Gilbert (1991). *Introduction to Environmental Engineering and Science*, Englewood, New Jersey: Prentice-Hall, Inc., p. 199.

number of noncarcinogens does not reflect their suspected prevalence in the larger set or universe of chemicals to which the test might be applied, then great care must be given to interpreting the predictive power of the short term test.

Table 5.1 illustrates this point. Predictive power is the ratio of the number of carcinogens detected by the Ames Test (i.e., observed to be a mutagen) to the number of total chemicals identified as mutagens in the sample. In Table 5.1, it is assumed that a validation study using a specific set of chemicals has shown the short term test has a predictive value of 90 percent (sensitivity and specificity are both 90 percent). Now suppose that the test is used by Company A and Company B to identify carcinogens among the new chemicals they are developing. Each company has 1000 different new chemicals to test. If the proportion of carcinogens in these different chemical sets is 1 percent for Company A and 10 percent for Company B, the predictive value of the test will be vastly different, even though the test has identical sensitivity and specificity. In one case, the predictive value is 8 percent and in the other case 50 percent.

Table 5.1 Expected results of a short term mutagenicity test used with two hypothetical samples

	Company A	Company B
Proportion of carcinogens in sample of 1000 chemicals	1% (10 carcinogens)	10% (100 carcinogens)
Carcinogens identified as mutagens	90% (9 of the 10 carcinogens)	90% (90 of the 100 carcinogens)
Carcinogens identified as nonmutagens (false negatives)	10% (1 of the 10 carcinogens)	10% (10 of the 100 carcinogens)
Noncarcinogens identified as nonmutagens	90% (891 of 990 noncarcinogens)	90% (810 of 900 noncarcinogens)
Noncarcinogens identified as mutagens (false positives)	10% (99 of 990 noncarcinogens)	10% (90 of 900 noncarcinogens)
Predictive value (carcinogens identified as mutagens/total carcinogens plus noncarcinogens identified as mutagens)	8.3% (9 of 108)	50% (90 of 180)

Source: OTA (1981).

Company A and Company B might have very different attitudes about the test. Company A can learn that 92 percent of the chemicals identified as mutagens are false positives (i.e., the mutagens are not in fact carcinogens). Company B can find that 50 percent of the mutagens are false positives. It would not, therefore, be surprising if these two companies were influenced differently by the Ames Test. Of course, these two companies probably do not know the underlying proportion of carcinogens in their new products.

The difference between the proportion of carcinogens found in the sample used to calculate the correlation and the proportion in a larger (or at least different) set of chemicals is not caused by authors of studies who are attempting to mislead the public. Tennant et al. recognized that attempts to identify carcinogens has led to a bias in the literature, which tends to focus most on those chemicals that are carcinogens. There are relatively few studies that show that a given chemical is not carcinogenic. Consequently, correlation studies often rely on a data base that contains a relatively large number of carcinogens and may not reflect the proportion of carcinogens found in a different set of chemicals.

Following this logic, one could therefore argue that correlation studies should be based on chemical samples that reflect a random sampling of the universe of all chemicals. This would be more indicative of the predictive power of the test—if the test is in fact used to test chemicals indiscriminately or on a random set of chemicals. One practical limitation of this argument in the case of the Ames Test is that there are a limited number of long term animal bioassays that can be used in the correlation study. The data base consists largely of carcinogenic chemicals.

This is not to say, however, that the Ames Test's predictive value, based on a sample containing many carcinogens is not informative. One could argue that it is incorrect to assume that the Ames Test should be used to test new chemicals indiscriminately or at random. The proportion of carcinogens in some chemical classes, such as alkylating agents, is higher than in the general universe of all chemicals. Similarly, some types of chemicals tend not to be carcinogenic, such as polymers. Therefore, information provided by the Ames Test when testing alkylating agents will probably be of more use than when

applied to polymers; the number of false positives and false negatives will be different for each of these two classes.

This is the logic used in some industrial research labs. One toxicologist for a large chemical and manufacturing company says that the Ames Test is not used indiscriminately on all new chemicals that are developed. New chemicals selected to be tested in the Ames Test are typically chosen on the basis of SAR analysis. This allows them to place more credence in the results of the Ames Test.

From the above discussion, we see two different approaches to selecting chemicals on which to base studies of correlation and predictive power. First, it can be argued that correlation studies should rely on a sample of chemicals that reflect the universe of chemicals. Focusing on a chemical sample with a high proportion of carcinogens can cause misleading predictions of the performance of the Ames Test when applied to a different set of chemicals. This viewpoint was voiced during the OSHA hearings for its 1980 cancer policy. A representative from CIIT downplayed the significance of correlation studies, arguing that the high correlation observed was misleading because of "judicious selection of chemicals used."[30]

On the other hand it can be argued that such correlation studies are not misleading if interpreted appropriately. Chemical companies tend to apply the Ames Test only to certain chemicals classes because the credibility of test results varies across chemical classes. Correlation studies can be designed to identify those chemical classes for which the Ames Test is most relevant. In contrast to the first approach, the goal here is not to test randomly in order to calculate the test's predictive value for the universe of chemicals, but rather the goal here is to find those chemicals for which a high correlation will in fact exist. If the Ames Test does not perform reliably for a certain chemical class, it makes little sense to include representatives of that class in the correlation study; a chemical company would not use the Ames Test for that chemical class anyway.

30 OSHA (1980). p. 5169.

The E-screen in the regulatory arena

The third major theme that emerges from the case study of the Ames Test is that there are only certain circumstances in which the Ames Test has been influential. It is not used by every company and regulatory agency in all situations. In this section, we discuss the likely niches for the E-screen and the situations where it is unlikely to have an effect.

Risk assessment guidelines

Along with the guidelines for cancer risk assessment discussed in Chapter 4, the EPA has also developed risk assessment guidelines for developmental toxicity. These guidelines encompass many of the non-cancer effects that have been attributed to estrogen-mimicking chemicals. Developmental toxicity refers to the adverse effects resulting from exposure prior to conception (either parent), during prenatal development, or postnatally to the time of sexual maturation. The adverse effects may be manifested at any time during the organism's lifetime. These adverse effects include death of the developing organism, structural abnormality, altered growth, and functional deficiency. In general, only the first three types of effects have been traditionally tested for in animal studies. The EPA states that endocrine and reproductive systems are subject to alterations in functional competence but testing for functional deficiencies "has not been required routinely" by regulatory agencies.[31]

Testing protocols have been established for developmental toxicity, but they may not include endpoints most relevant to the effects of estrogen-mimicking chemicals. This was a major topic of discussion at the April 1995 EPA workshop on endocrine disruption. Earl Gray, chairman of the breakout group for hazard identification, stated that there is little in the current testing scheme that could detect estrogenicity besides measurements of the weight of the developing organism, and that is not an effect specific to estrogens. He noted that the types of tests and the endpoints being measured are currently undergoing revision and are likely to include tests more informative for identifying estrogen-mimicking chemicals and other endocrine disruptors.

---

[31] EPA (1991). "Guidelines for Developmental Toxicity Risk Assessment," *Federal Register*, Vol. 56, No. 234, December 5, pp. 63801, 63805-63806.

In performing a full risk assessment for developmental toxicity, different types of data are weighed to determine whether there is a potential risk. EPA calls this process the characterization of the health-related data base; its purpose is to determine, in the context of expected exposure, whether there is sufficient data for proceeding further in the risk assessment process. Table 5.2 shows how data for chemical risks are weighed and categorized. Expectedly, human epidemiological studies are the preferred basis for inferring risk. Animal studies also contribute substantially, while short term tests have limited influence. Databases limited to short term tests, structure-activity

Table 5.2 Categorization of the health-related data base for hazard identification/dose-response evaluation for developmental toxicity

Sufficient evidence: The sufficient evidence category includes data that collectively provide enough information to judge whether or not a human developmental hazard could exist within the context of dose, duration, timing and route of exposure. This category includes both human and experimental animal evidence.

Sufficient Human Evidence: This category includes data from epidemiological studies (e.g., case control and cohort) that provide convincing evidence for the scientific community to judge that a causal relationship is or is not supported. A case series in conjunction with strong supporting evidence may also be used. Supporting animal data may or may not be available.

Sufficient Experimental Animal Evidence/Limited Human Data: This category includes data from experimental animal studies and/or limited human data that provide convincing evidence for the scientific community to judge if the potential for developmental toxicity exists. The minimum evidence necessary to judge that a potential hazard exists generally would be data demonstrating an adverse developmental effect in a single, appropriate, well-conducted study in a single experimental animal species. The minimum evidence needed to judge that a potential hazard does not exist would include data from appropriate, well-conducted laboratory animal studies in several species (at least two) which evaluated a variety of the potential manifestations of developmental toxicity, and showed no developmental effects at doses that were minimally toxic to the adult.

Insufficient evidence This category includes situations for which there is less than the minimum sufficient evidence necessary for assessing the potential for developmental toxicity, such as when no data are available on developmental toxicity, as well as for data bases from studies in animals or humans that have limited study design (e.g., small numbers, inappropriate dose selection/exposure information, other uncontrolled factors), or data from a single species reported to have no adverse developmental effects, or data form a single species reported to have no adverse developmental effects, or data bases limited to information on structure/activity relationships, short-term tests, pharmacokinetics, or metabolic precursors.

Source: EPA (1986).

relationships, and/or pharmacokinetics are categorized as insufficient evidence, but these data could be used to determine the need for further testing.[32]

It is not surprising that short term tests are not used as the sole basis in risk assessment for developmental toxicity. This does not imply, however, that short term tests like the E-screen has no potential for influence in environmental regulation. As was seen in the Ames Test case study, the Ames Test was not given much weight in EPA's cancer risk assessment guidelines, but still was used by EPA and industry. Vince Nabholz of EPA's Office of Toxic Substances says that the risk assessment guidelines were written by people who do not on a daily basis have to make decisions about chemical hazards with little or no data at hand.[33] While the E-screen does not have a prominent position in a full risk assessment, there are other circumstances where the E-screen may have a larger role, such as new chemical testing under FIFRA and TSCA statutes. These are discussed in the following section.

Environmental statutes and regulations

In Chapter 4, we discussed how the Ames Test has been used in the implementation of two environmental statutes, FIFRA and TSCA. These laws give EPA the authority to regulate the chemicals that cause cancer and reproductive and developmental defects—the kinds of effects suspected of being caused by estrogen-mimicking chemicals. In this section we discuss how the E-screen might fit into the current regulatory framework.

Lack of experience with the E-screen

Because estrogen-mimicking chemicals have come onto the environmental policy scene only recently, EPA is not yet in a position to regulate these chemicals on the basis of their estrogenic capabilities. Steven Devito, chairman of EPA's working group on environmental estrogens, stated that until more knowledge is gained about this phenomenon, it would be premature to take any regulatory actions. He pointed to the importance of critical assessments of the issue, such as the one currently being undertaken

32 EPA (1991). p. 63817.
33 Nabholz, Vince (1995). Personal communication.

by the NRC. Devito hopes that the NRC study will "separate the good science from the bad science" so that EPA's actions will be based only on sound science. One of the goals of the NRC study is to provide a scientific foundation on the issue "to guide appropriate responses to the concerns and to reduce the chances of misdirected legislation or regulatory action."[34] Devito stated that although the EPA is proceeding cautiously, the agency is very interested in both long term and short term tests for estrogen-mimicking chemicals.[35]

Two members of the organizing committee for the EPA's Endocrine Disruptor Research Needs Workshop added to Devito's remarks. Gary Ankley of EPA's Environmental Research Laboratory stated that discussions about how to regulate estrogen-mimicking chemicals "haven't gotten that far yet." And in regard to testing chemicals for their estrogenicity, Ankley claimed that it is too early for EPA to consider what testing would be necessary for regulatory action, although he said that the E-screen is one of many tests that are being investigated.[36] Earl Gray of EPA's Health Effects Research Laboratory said that it will probably take a long time before a test like the E-screen would be established as a routine part of EPA's decision making. Although it is very difficult to project what will happen with the E-screen, Gray said that it would not be unexpected if, like the Ames Test, 15 years of test development and validation are needed before the E-screen is accepted as part of EPA's decision making.[37]

Industry, like the EPA, is moving somewhat cautiously, attempting to avoid a rush to judgment. The industry is in the process of gathering more information. Several companies have experimented with the E-screen in their laboratories.[38] And, as mentioned previously, the Chlorine Chemistry Council has commissioned a study to review short term tests for estrogenic chemicals. According to Ron Miller, chairman of the Chlorine Chemistry Council's panel investigating these issues, industry is interested in the E-screen, not so much because of enthusiasm about the test, but rather because they are wary of the test. He thinks that many people in industry are worried that the E-screen will be accepted uncritically. He pointed to the Ames Test as an example of what

34 NRC Commission on Life Sciences (1995). p. 3.
35 Devito (1995). Personal communication.
36 Ankley, Gary (1995). Personal communication.
37 Gray, Earl (1995). Personal communication.
38 Soto (1995). Personal communication.

industry hopes to avoid. It would have been a serious error for regulators and industry if they had accepted the Ames Test after the first few years of its development, when there was still much enthusiasm about the ability of the Ames Test to identify carcinogens accurately. The Ames Test plays a useful role now, but this came about only after many years of serious scrutiny on the Ames Test and the relationship between mutagenicity and carcinogenicity. Miller said that there is a danger in trying to look for a quick fix to the problem of identifying estrogen-mimicking chemicals. Industry is interested in using short term tests, but it is far from clear which test or battery of tests would be most appropriate.[39]

It is clear that there is a great deal of uncertainty in how EPA and industry might use the E-screen in their decision making. The issue of estrogen-mimicking chemicals is still very new, and the E-screen is early in its development. Chapter 4 gives some shape to our expectations for the role of the E-screen in government and industry decision making. If the predictive power of the E-screen is not vastly different from that of the Ames Test, we can, for example, expect the E-screen to have a larger effect on decision making for new chemicals than on existing chemicals.

In addition, we can expect the E-screen to have varying effects on new chemical testing programs. In the development of FIFRA chemicals, the E-screen may be an important factor in company decisions about which closely related congeners of a new pesticide should continued to be developed, or whether further development should be halted in favor of other congeners. While at EPA, it is likely that the agency will not find very persuasive the results of the E-screen when it weighs all of the data submitted for FIFRA registration.

And as suggested by the Ames Test case study, the E-screen could have an effect in TSCA's PMN provisions. This will depend in part on the degree to which the testing is done; the paucity of data submitted for PMN submissions indicate that this is not an insignificant issue. Moreover, since both industry and EPA decision making for TSCA chemicals is based to a large extent on concerns for exposure, the effects of the E-screen may be more subtle, such as influencing decisions to control exposure. The link between the E-screen and decisions is likely to be less than fully clear or obvious.

---

39 Miller (1995). Personal communication.

Interviews with scientists and staff at both EPA and several large chemical companies help elucidate the important factors in determining how the E-screen will be used. While most people interviewed thought that the E-screen could be used in ways similar to the Ames Test, too little is known about the E-screen to make accurate predictions about its use and influence. Nevertheless, from these interviews emerged two important issues. The first deals with "conservatism" of the E-screen, and the second concerns the relationship between the E-screen and other tests used to identify estrogen-mimicking chemicals. Below we discuss these issues in turn.

Conservatism of the E-screen

The idea of conservatism was brought up by several people interviewed. If the results of the E-screen are to be used as a basis for deciding whether further testing should be performed, some argued that the E-screen should err on the side conservatism. That is, the test should be designed to avoid false negatives. This reflects a concern that a false negative result in the E-screen would lead, perhaps, to an incorrect assumption that a chemical is safe. Gary Ankley of the EPA argues that "it would be a disaster to have a test that wasn't completely conservative."

There is a cost, however, to a conservative test that avoids false negatives. That cost comes in the form of a higher rate of false positives. Positive results could lead to more sophisticated tests that are more expensive and time-consuming. It is not surprising that the EPA or a company would be reluctant to use a test that resulted in many false positives. Ankley added that a test would be less useful if it is *too* conservative. Designing a test to avoid, to the greatest extent possible, either false positives or false negatives may not be feasible or even desirable. Designers of the E-screen and decision makers using its results must confront the sticky issue of finding an appropriate balance between false negatives and false positives. It is far from clear what the appropriate balance might be; it is likely to differ depending on how the results will be used.

But a test that gives some false positives and some false negatives may not help matters much either. Paul Schlosser of the CIIT pointed out that a positive result in the E-screen may lead to more sophisticated testing. And if there is a reasonable possibility of false negatives, a negative result may not

necessarily imply that the chemical is safe; more testing may still need to be done. The result, said Schlosser, is that regardless of the whether the E-screen gives a positive or negative, further testing would have to be done. What then would be the use of the test?

Even though "no qualitative differences could be found when comparing" estrogenicity as measured in animal bioassays and the E-screen,[40] most industry scientists interviewed were skeptical of the early results of validation study. Further validation studies should be performed to examine a greater number of chemicals. Only then can judgments be made about the frequency with which the E-screen will produce false positive and false negatives.

Ankley, Gray, and Schlosser all emphasized the point that the results of the E-screen are most useful when there is other data available from other short term tests. There are a variety of mechanisms through which chemicals may exert hormonal effects; some may not interact directly with the estrogen receptor but work through complex inter-cellular signals. A false positive or a false negative result in one test may be correctly identified by other tests. Use of a battery of short term tests may seem to be an obvious point, but it is not a trivial task to actually design a battery of tests what will minimize false positives and/or false negatives. For example, the 1987 study by Tennant et al. argued that a battery of the four widely used in vitro assays used for genetic toxicity does not substantially improve the performance of the Ames Test alone.[41]

The E-screen's relationship to other tests

The development of other short term in vitro tests to be used with the E-screen in a battery of tests is just one example how other tests influence the way that the E-screen will be used. Chapter 4 illustrates how the use of the Ames Test was affected by the existence of a broad range of testing technologies. For example, in TSCA's PMN process, SAR analysis was used identify chemicals for which the Ames Test might be useful.

The development of SAR analysis for estrogen-mimicking chemicals has been limited. Chemicals with highly diverse structures have been associated with estrogenic properties; this makes the development of SAR analyses

40 Soto et al. (in press).
41 Tennant et al. (1987). p. 933.

difficult. Some progress is being made, however. At the EPA's Endocrine Disruptor Research Needs Workshop, Chris Waller of EPA's Health Effects Research Lab presented his work on developing an SAR model for predicting chemical binding affinities to the estrogen receptor. It is important to recognize that estrogen receptor binding may not be directly responsible for in vivo effects. Waller's model cannot predict whether the chemical will act as an agonist or antagonist.[42] Even so, there is significant interest in both industry and EPA for information on receptor binding, and Waller's SAR model could provide insight. The development of an this and other SAR models could prove useful as a "pre-screen" to identify chemicals for which the E-screen would be most relevant.

In addition to a pre-screen for the E-screen, development of in vivo tests for estrogenicity will effect how the E-screen is used. Ron Miller, chairman of the CCC's panel on environmental estrogens, stated that focusing on the E-screen, to the extent of ignoring other tests, will cause industry to resist the use of the E-screen. Donna Farmer of Monsanto argued that too much emphasis on the results of the E-screen will lead a chemical being labeled prematurely as a bad actor. That label is difficult to shed, even if other tests indicate that the chemical is not cause for concern. She noted that when results of the E-screen are published showing which chemicals are estrogenic, regulators, the press, and the public take it for granted that a given chemical is harmful, even if in vivo testing shows otherwise. It is very difficult to remove the label, Farmer asserted.

This "labeling" problem was seen also in the early use of the Ames Test, said Miller. We now know that cancer is a phenomenon much more complex than is reflected in the statement that "carcinogens are mutagens." Miller expressed his hope that the E-screen would not lead to similar "labeling" problems. He argued that a broader view must be taken in testing. While the E-screen may provide some information, Miller stated that the identification of estrogenic chemicals requires testing beyond the E-screen.

Miller went on to argue that when judged against other in vivo and in vitro tests, the E-screen does not appear to have any significant advantage. He said that the E-screen may be useful in a limited way, such as the assessment of closely related congeners of a new pesticide. But for the most

---

42 Soto (1995). Personal communication.

part, the E-screen would be a "waste of money" as a way to identify chemicals that cause uterotropic activity. He disagreed with the assertion that the E-screen's low cost and quick results make it more attractive to in vivo tests for the purpose of large scale screening. More reliable information about uterotropic activity can be obtained through short term in vivo tests, Miller claimed. Jerry Reel, the author of the CCC's review of short term tests for estrogenicity agreed, saying that a 72-hour uterotropic in vivo assay costs around $6500 per test.[43] While this is more expensive than the E-screen, both of them argue that in vivo testing incorporates metabolic activity that the E-screen cannot reliably mimic in vitro.

Miller did not say, however, that all in vitro assays have only limited use. He said that in vitro tests that measure estrogen receptor binding or use transfected cells to examine gene activation provide information that is not significantly improved upon by more expensive in vivo testing. Reel asserts that for binding and transfected cell assays, the mechanisms are fairly well understood. In contrast, Reel states that the mechanism through which cells proliferate in the E-screen is not well understood. Their main argument against the E-screen is that the test provides information about uterotropic activity that can be obtained more reliably through short term in vivo tests at costs not prohibitively higher. But because the cost is higher, it is possible that less testing would be done as compared to the less expensive E-screen. The utility of screening tests should be judged not only on the quality of data, but also on the quantity of data (i.e., how many chemicals will be tested).

Miller went on to say that there is a better strategy for test development. He stated that industry is more interested in developing in vivo tests that incorporate uterotropic activity as well as a wide variety of other endpoints. There is uncertainty about the various mechanisms of action through which estrogens may act, and a variety of effects may be observed. Consequently, industry would prefer to use a test that had several endpoints than to perform a series of isolated, narrowly focused tests based on individual endpoints or mechanisms of action.

Miller's stance on test development was echoed as well as criticized at the EPA Workshop. For example, Robert Chapin of the National Institute for Environmental Health Sciences argued that a screening test should be

---

43 Reel, Jerry (1995). Personal communication.

designed to detect a wide variety endpoints and mechanisms of action. He questioned the wisdom using tests that focus on specific endpoints and mechanisms of action when little is known about them. On the other hand, some scientists argued that tests such as estrogen receptor binding assays provide fundamental information about a chemical's characteristics.

Conclusions

Against the backdrop of the Ames Test case study, this chapter has critically examined the potential uses and influence of the E-screen. One major factor that will affect the future of the E-screen is the public, regulatory, and industrial concern for estrogen-mimicking chemicals and their effects. Although this concern is increasing, the issue has not achieved the political status attached to cancer several decades ago, a status that made the cure and prevention of cancer a national imperative. Given the fickle nature of the American public on environmental issues, it is difficult to gauge how this concern for estrogen-mimicking chemicals will progress. A high level of concern could lead to a more rapid acceptance of the E-screen, but that may or may not have a positive impact on the E-screen and environmental regulation. Both EPA and industry have shown interest in the E-screen, but they are likely to resist the acceptance of the E-screen or any other testing technology until more research is performed and more experience is gained.

A key issue that will affect the use and influence of the E-screen is the results of validation studies. It is clear that much more study is needed in this area. The design and interpretation of validation studies must be approached in a methodical and deliberate fashion. Standardization of the protocol, the standards of comparison, and the selection of chemicals will all affect the validation process. The choice of a standard of comparison is particularly problematic because of the diversity of endpoints related to estrogen-mimicking chemicals. Regulatory agencies and companies have different requirements for what constitutes a reasonable validation study; they ask different questions and have different circumstances. For example, some may want to know if the E-screen is correlated with uterotropic activity while others are more interested in whether it is correlated with multigenerational studies. There is no single standard of comparison that must be used in a validation study.

The third major factor is the existing regulatory context for estrogen-mimicking chemicals and their supposed effects. The utility of the E-screen is defined largely by existing statutes and regulations; there are only certain niches that the test could fill. Because the E-screen is so new, both EPA and industry are proceeding cautiously in regard to estrogen-mimicking chemicals. Information about the relative frequency of false positives and false negatives will help determine how the results of the E-screen are used.

The need for a testing infrastructure

The influence of the E-screen will depend not only on the characteristics of the test itself—its reliability, its predictive power, etc.—but also on the existence and shape of a testing infrastructure. Testing infrastructure refers to the different types of tests available to identify estrogen-mimicking chemicals. The first reason why a testing infrastructure is needed is obvious. The E-screen may not be able to detect all estrogenic chemicals because the relevant mechanisms or metabolic activities may not be incorporated in the E-screen.

There is a more subtle reason why a testing infrastructure is needed. On the one hand, the E-screen may fail to be embraced for the obvious reason that a better test exists. That was argued by some scientists who believed that short term in vivo tests were better than the E-screen, even for large scale testing needs. But, importantly, the E-screen may also fail to be embraced if other tests are *not* developed. The E-screen is likely to be more useful if there is a well developed testing infrastructure. The creation and use of SAR as a "pre-screen" can pinpoint the chemical classes for which the E-screen may be most appropriately applied.

In addition, tests more sophisticated than the E-screen are needed. For example, if the E-screen was very conservative, a large number of false positives may be generated. If the next, more sophisticated step in a testing program was not very expensive to perform, then the cost of a false positive from the E-screen would not be very high. In contrast, if the next step in testing was very expensive, then a screening test that generated a large number of false positives is less attractive.

The scope of testing for estrogen-mimicking chemicals should be large. A wide variety of tests and approaches will be beneficial. Focusing on a single test like the E-screen, to the exclusion of other tests, is likely to decrease the utility of the E-screen itself.

Chapter Six

Conclusions

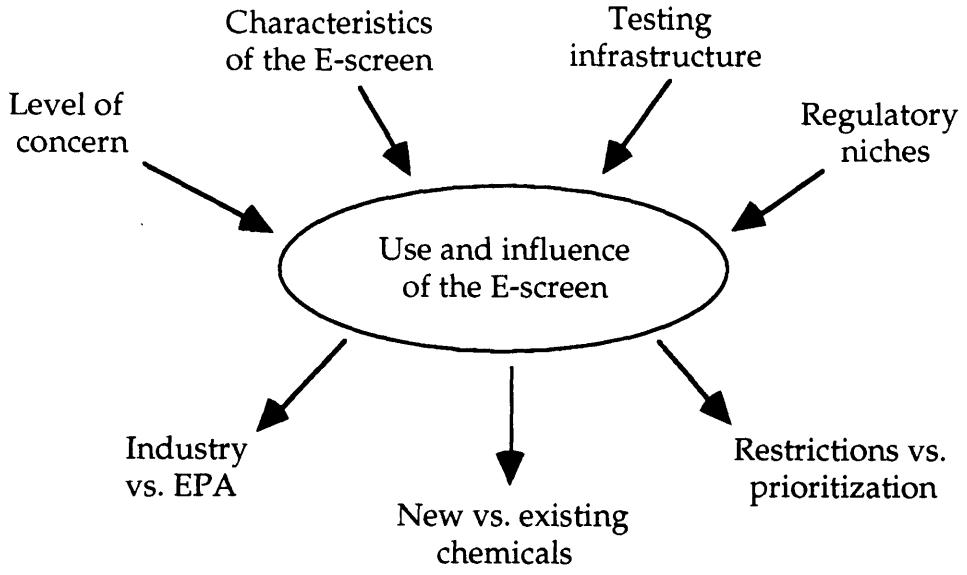
The underlying motive for this thesis is to understand and make recommendations about the development and future uses of short term tests for endocrine disrupting chemicals. In the scientific and regulatory communities, there does not seem to be a lack of recommendations. Some people argue that short term tests should be swiftly adopted by regulators and industry. Others believe such action is premature or that the tests are in some sense not good enough. It is not always clear what these people have based their recommendations on. The recommendations presented in this chapter are based on perhaps a different set of considerations. We have relied on the history of the Ames Test in cancer risk assessment. In addition, we have attempted to assess the concerns of scientists from academia, government and industry who are likely to have experience in the practical problems of chemical testing. Using the E-screen as a model, we review in this chapter the major findings of the research and conclude with several recommendations.

Findings

A useful way to organize the findings of this thesis is illustrated in Figure 6.1. The top portion shows four major factors that will affect the degree to which the E-screen is used and is influential. The lower part of the figure shows some of the important distinctions that should be made in assessing how the E-screen might be influential. Arrows, which indicate an effect, are not drawn between factors. We can expect, for example, that the level of concern affects the development of the testing infrastructure, but this link is not shown in the figure. The figure merely illustrates the factors related to the E-screen, the

focus of this research. We will first discuss the factors that will affect the E-screen, and then move to a discussion of the ways that it will be influential.

Figure 6.1 Factors affecting the E-screen and ways it may be influential



Factors affecting the E-screen

Level of concern

The first factor shown in Figure 6.1 relates to the level of concern that exists for estrogen-mimicking chemicals and their potential effects. In the Ames Test case study, we saw that the public concern for cancer was resounding. The campaign against cancer was sustained in the popular press, government bodies, and scientific laboratories. Policy makers responded by mounting against the disease a national effort relying significantly on scientific research. A solution was expected, and the Ames Test was touted as a significant part of that solution. Much interest in the Ames Test was generated, and thousands of chemicals were tested soon after it was first developed.

The initial high expectations for the Ames Test led to two things. First, because the test seemed so promising, many laboratories were eager to experiment with it. This allowed more data to be generated with which to judge the characteristics and utility of the Ames Test. The quick dissemination of the test also created confusion about comparing results between different laboratories. Second, much controversy was created by questions about whether the Ames Test could live up to the high expectations. A great deal of scrutiny followed, and some of the high expectations for the Ames Test were left unfulfilled. EPA did not begin using the test in its implementation of TSCA until some 15 years after the Ames Test was first developed. We saw in the case study that failure to meet some of those expectations did not mean that the Ames Test was useless, but rather that it filled different niches.

It is uncertain whether concern for estrogen-mimicking chemicals and their effects will reach a level as high as that for carcinogens and cancer during the early 1970s. It is difficult to predict what scientific research will uncover about estrogen-mimicking chemicals, and it is just as hard to guess how the public will respond to the issue. Estrogen-mimicking chemicals, and endocrine disruptors more generally, are gaining increased attention from both industry and government. More resources are being invested for research in this area, but it remains to be seen whether the effort will be sustained or if legislative measures will be taken—an uncertain proposition at best, given the current budget-cutting attempts of Congress.

It should not be assumed that the use of the E-screen will rise proportionally to a rise in public concern. Increased attention for the E-screen may help or hinder the test. On the one hand, wide dissemination of the test is important for generating more experimental data, understanding its limitations, creating confidence in the test, and possibly gaining broader support for the test.

On the other hand, growing public concern can lead to overselling the capabilities of the test. This overselling may be done by scientists, interest groups, or the media. Brian MacMahon of the Harvard School of Public Health, speaking in regard to the tentative reports of the link between organochlorine pesticides and breast cancer, notes that science does not "operate in a vacuum, and however cautiously the investigator may report his or her conclusions and stress the need for further evaluation, much of the press will pay little heed to such caution. Even when it does, by the time the

information reaches the public mind via print or screen, the tentative suggestion is likely to by interpreted as fact."[1] The result of overselling the E-screen could be a backlash against it. With the relatively wide publicity for the E-screen, this potential exists.

Characteristics of the test

This second factor includes items that can be considered to be the characteristics of the test itself. For example, it includes the characteristics of the E-screen's validity, predictive power, cost, and reliability. Although some validation studies have been performed, all interviewed scientists from government and industry uniformly believe that their acceptance of the E-screen requires much more research to ascertain the test's characteristics.

Great attention must be given to the design of validation studies. The concordance of the E-screen will depend a great deal on the characteristics of the chemical sample used, which may or may not be similar to the set of chemicals to which the E-screen may be applied in the future. Interpretation and application of the validation studies to specific circumstances must be done with great care.

Another important issue regarding validation studies is the choice of standards of comparison. Some scientists believe that the appropriate standard is uterotropic assays in vivo. Others argue that validating the E-screen against this endpoint may not be very informative if uterotropic activity is not what is of most concern or indicative of other effects. They might believe that more sophisticated tests like multigenerational tests should be used. The choice of an appropriate standard depends on the anticipated use of the test, and this should be made explicit to avoid any misunderstandings.

It is important to acknowledge that it is not necessary that the most sophisticated test be used as the standard of comparison in order for the E-screen to be useful. In a more ideal world, the E-screen would be perfectly predictive of human risk or correlated with multigenerational tests, but it is unrealistic to expect or demand that this would occur in the real world. The E-screen can still provide important information for many purposes if it is correlated well with endpoints such as uterotropic activity.

---

[1] MacMahon, Brian (1994). "Pesticide Residues and Breast Cancer?" *Journal of the National Cancer Institute*, Vol. 86, No. 8, April 20, p. 573.

Testing infrastructure

This factor relates to the existence of other tests—a testing infrastructure—that may complement the E-screen. Many scientists identified the need for other short term tests that could be used with the E-screen in a battery of tests. The variety and uncertainty of different mechanisms of action for estrogenic chemicals (and other endocrine disruptors) make other tests necessary. In addition to gaining new insight into the phenomena, the development of a testing infrastructure also provides a richer context with which to conduct validation studies.

A less obvious way to complement and even encourage the use of the E-screen is to develop tests that can serve as a pre-screen or as a higher level of testing. SAR analysis might be a useful pre-screen to identify the chemicals for which the E-screen would be informative. As we saw in the case study, the Ames Test is more reliable for some chemical classes but not others; SAR was used as a pre-screen. In addition, the E-screen can be more useful if there are low cost tests that can be used as another tier of testing following the E-screen. This is particularly important if the E-screen generates a high number of false positives. If the testing tier following the E-screen was expensive, then the consequences of false positives from the E-screen would be high monetary costs, possibly making the E-screen less attractive.

Although a testing infrastructure is likely to increase the utility of the E-screen, it is possible that it could also render the test less useful. As discussed in Chapter 5, some people believe that the E-screen provides less information than can be gained from short term in vivo testing. They argue that resources should not be spent on collecting data from the E-screen. Even if a short term test, such as one that measures uterotropic activity in vivo, provides more reliable information than the E-screen, that does not necessarily mean that the E-screen should not be used. Judgments about screening tests, whether in vivo or in vitro, should be made with consideration given to the number of chemicals that will be tested. There will be a tradeoff between the quality and quantity of test data. Therefore, it is not a foregone conclusion that short term uterotropic assays should be used instead of the E-screen.

Regulatory niches

The Ames Test case study suggests that the E-screen has the potential to be influential in certain ways. For example, it should not be expected that the E-screen will be used to test all new and existing chemicals. TSCA does not require industry to submit a minimum set of test data for PMN review. Although EPA has the authority after its review to require that industry conduct more testing, industry often does very little testing. In contrast, FIFRA regulations require an extensive array of tests to be performed for pesticide registration.

Ways that the E-screen can be influential

Influence is a slippery term. As the conclusion of Chapter 4 discussed, influence can have several different meanings and subtleties. Figure 6.1 shows some of the distinctions that should be made in assessing influence.

Industry vs. EPA

When assessing influence of the E-screen, one must be certain to understand who or what is being influenced. The test can be influential in industry in some circumstances, but largely irrelevant to EPA in other circumstances. Consequently, assessment of the influence of the E-screen depends on the perspective taken and the metrics of influence used.

For example, EPA is unlikely to give much weight to short term in vitro tests when reviewing the extensive data base for a new pesticide undergoing registration review. In contrast, a short term in vitro test may be relied upon heavily by the company because a large data base may not yet exist. In addition, the E-screen may have an important but "silent impact" in industry decision making that is not ordinarily observed by either EPA or the public.

New vs. existing chemicals

Much of the focus of this thesis has been on the use of the E-screen to test new chemicals, as opposed to existing chemicals. Part of the reason for this is the shift in environmental policy towards pollution prevention.

Addressing the EPA Endocrine Disruptors Research Needs Workshop, Theo Colborn stated that the "old" way of thinking was to try to control already existing hazards. The "new" approach is to look for opportunities to prevent pollution. In many cases, she says, it is "too late" to regulate existing chemicals already being released to the environment; more emphasis should be put into controlling new chemicals.

The E-screen is likely to have a larger influence on new chemicals than on existing chemicals. In general, a large body of evidence is required to restrict an existing chemical that is produced in large quantities. The E-screen could contribute to that body of evidence or serve to flag existing chemicals that should be tested further. However, decisions about restrictions on an existing chemical will be based on more sophisticated long term animal bioassays and/or human epidemiological evidence. When large investments have been made on an existing chemical, it can be expected that industry and/or EPA will be reluctant to stringently control the chemical without compelling evidence. For new chemicals, less evidence is often required for decisions to be made about its development.

Restrictions vs. prioritization

The E-screen can have different types of influence. On the one hand, the E-screen could be used as the basis for the EPA to decide that a chemical should be restricted or for a company to decide to halt the development of a new chemical. On the other hand, the E-screen may influence decision makers in prioritizing chemicals for further testing. In assessing the influence of the E-screen, it is critical to understand what exactly is meant by influence.

No scientist interviewed in this research argued that the E-screen should be used as the sole basis for a decision to ban an existing chemical. There was also some concern among industry scientists that results of the E-screen could lead to inappropriate "labeling" of a chemical. They were wary of the possibility that the E-screen could be used as the primary basis for a ban on an existing chemical. They emphasize that the E-screen can provide useful information, but it is only one piece of evidence that must be considered in deciding the fate of a chemical.

Most interviewed scientists agreed, however, that the E-screen could be influential by affecting decisions about whether to pursue more testing. In

some specific circumstances, the E-screen could influence decisions about whether development of a new product should be continued. One example that was cited was the situation in pesticide research where there are a number of different congeners being examined. Everything else being somewhat equal, results of a short term test may lead a company to halt development of some congeners and continue with others. Measuring the impact of the E-screen in these circumstances is difficult. A company's internal chemical development decisions are not open for the public to examine.

Furthermore, the Ames Test case study suggests that influence should not be measured solely by the number of chemicals that are tested in the Ames Test or how many chemicals are banned by EPA. This was shown in EPA's PMN process. The Ames Test is not used to test all chemicals because the test performs poorly for some classes of chemicals. EPA's listing of chemical categories of concern shows what tests are likely to be needed. In addition, chemical companies may seek to implement exposure controls for new chemicals rather than performing additional tests. Industry and EPA scientists suggest that in cases like this, the Ames Test was influential, even though additional testing was not done and the company still went ahead with chemical development. The E-screen may have a similarly subtle influence.

Recommendations

There is still considerable scientific uncertainty in our understanding of endocrine disruption and the ability of the E-screen to identify estrogen-mimicking chemicals. It cannot yet be determined whether policy makers and the public will demand that estrogen-mimicking chemicals be identified and controlled. Consequently, long term recommendations are not yet suitable. Several short term recommendations can be made here, however. At this point, there is no over-arching administrative entity responsible for guiding research in this area. This issue is at the cutting-edge of science and policy, and it may be too soon to expect or even to desire a centralized organizing body. The following recommendations, therefore, are addressed primarily to scientists and regulators who may be considering the establishment of a testing program for estrogen-mimicking chemicals.

It is premature for regulatory agencies and industry to adopt the E-screen as part of a formal testing program.

Scientists in academia, government, and industry have not yet gained enough experience or confidence in its use. Both EPA and industry do not want to misstep in the way they handle this issue. This means, on the one hand, "staying ahead of the curve," anticipating future developments. But on the other hand, this also means that they do not want to overreact by taking unwarranted actions.

A testing infrastructure should be developed.

This recommendation is important for two reasons. First, decision making regarding endocrine disrupting chemicals will benefit from the existence of a broad variety of testing tools. Second, the E-screen itself is likely to benefit from the development of other tests. Useful tests would include "pre-screens" and low cost tests that can serve as the next tier of testing after the E-screen. The E-screen should not be over emphasized to the extent of ignoring other types of tests.

The potential utility of the E-screen should be further explored.

This is necessary before scientists and policy makers can be expected to accept the test for use in a testing program. Additional validation studies are required. The call for more research is probably not surprising. Importantly, however, several qualifications listed below would encourage wise research and development of the E-screen during the next few years.

The E-screen should be widely disseminated.

Greater experimentation in academic, government, and industrial laboratories allows for more possibilities of optimizing the test. For example, an important area of research would be to optimize the use of metabolic activation in the E-screen.

Formal guidelines should be developed.

A strict protocol or standard is not yet appropriate for the E-screen because it is so new and is likely to be refined. Guidelines for its use and interpretation, however, will be important as more laboratories begin experimenting with the

E-screen. The guidelines currently being developed through ASTM are a good start.

Validation studies should be performed by many different research groups.

These studies should be funded and undertaken by many different research groups, including those in academia, industry, and government. These groups will examine the E-screen with different perspectives, goals, and uses in mind for the test. For example, the choice of a standard of comparison may depend on the intended use of the test. It is unlikely that studies that are done exclusively by industry, government, or even academia, with its supposed objectivity, will be persuasive to all parties. It is critical that scientists from each of these groups gain personal experience with the test.

Scientists should identify those chemical classes for which the E-screen is most predictive.

The emphasis here is subtle. Because there are likely to be a variety of mechanisms of action for estrogen-mimicking chemicals and other endocrine disruptors (e.g., androgenic substances), the E-screen may not detect all chemicals of concern. Rather than trying to see if the E-screen is valid for *all* chemicals, the study should be framed to investigate if the E-screen is valid for *certain* chemical classes. Even if the E-screen cannot identify all estrogen-mimicking chemicals, it can still provide useful information when applied to relevant chemical classes.

These recommendations are based primarily on our assessment of one test in particular, the E-screen. It must be emphasized, however, that the analysis, findings, and recommendations may be applied to short term tests for endocrine disruption in general. Of course, the specific details will differ from case to case, but the lessons learned here should be relevant elsewhere. As the issue of endocrine disruption grows, we can expect that more short term tests will be developed in the future. It is hoped that the research presented here will shed light on what we can expect from these tests and how they should be developed and implemented.

Bibliography

Ames, Bruce, Lois Gold, and Walter Willett (1995). "The Causes and Prevention of Cancer," *Proceedings of the National Academy of Sciences USA*, Vol. 92, June, pp. 5258-5265.

Ames, Bruce (1979). "Identifying Environmental Chemicals Causing Mutations and Cancer," *Science*, Vol. 204, May 11, pp. 587-593.

Ames, Bruce, William Durston, Edith Yamasaki, and Frank Lee (1973). "Carcinogens are Mutagens: A Simple Test System Combining Liver Homogenates for Activation and Bacteria for Detection," *Proceedings of the National Academy of Sciences USA*, Vol. 70, No. 8, August, pp. 2281-2285.

Ashford, Nicholas and Charles Caldart (1991). *Technology, Law, and the Working Environment*, New York: Van Nostrand Reinhold.

Auer, Charles, Vincent Nabholz, and Karl Baetcke (1990). "Mode of Action and the Assessment of Chemical Hazards in the Presence of Limited Data: Use of Structure Activity Relationships (SAR) under TSCA, Section 5," *Environmental Health Perspectives*, Vol. 87, pp. 183-197.

Auger, Jacques, Jean Marie Kunstmann, Francoise Czyglik, and Pierre Jouannet (1995). "Decline in Semen Quality Among Fertile Men in Paris During the Past Twenty Years," *New England Journal of Medicine*, Vol. 332, No. 5, February 2, pp. 281-285.

Begley, Sharon (1994). "The Estrogen Complex," *Newsweek*, March 21, pp. 76-77.

Bern, Howard (1995). Keynote address at the EPA's Endocrine Disruptor Research Needs Workshop, April 10-13, Raleigh, North Carolina.

Bern, Howard (1992). "The Fragile Fetus," in *Chemically-Induced Alterations in Sexual and Functional Development: The Wildlife/Human Connection*, Colborn and Clement, eds., Princeton, New Jersey: Princeton Scientific Publishing.

Bridges, Bryn (1976). "Short Term Screening Tests for Carcinogens," *Nature*, Vol. 261, May 20, pp. 195-200.

Brody, Jane (1994). "Strong Views on Origins of Cancer," *The New York Times*, July 5, pp. C1, C10.

Bromwich, Peter, Jack Cohen, Ian Stewart, and Andrew Walker (1994). "Decline in Sperm Counts: An Artefact of Changed Reference Range of 'Normal'?" *British Medical Journal*, Vol. 309, July 2, pp. 19-22.

Brotons, Jose, Maria Olea-Serrano, Mercedes Villalobos, Vincente Pedraza, and Nicolas Olea (1995). "Xenoestrogens Released from Lacquer Coatings in Food Cans," *Environmental Health Perspectives*, Vol. 103, No. 6, June, pp. 608-612.

Carlsen, Elisabeth, Aleksander Giwercman, Niels Keiding, and Niels Skakkebaek (1992). "Evidence for Decreasing Quality of Semen During the Past 50 Years," *British Medical Journal*, Vol. 305, September 12, pp. 609-613.

Clark, James, William Schrader, and Bert O'Malley (1985). "Mechanisms of Steroid Hormone Action," in *Textbook of Endocrinology*, Wilson and Foster, eds., Philadelphia: W.B. Saunders Company, pp. 33-75.

Cohrssen, John and Vincent Covello (1989). *Risk Analysis: A Guide to Principles and Methods for Analyzing Health and Environmental Risks.* Prepared for the U.S. Council on Environmental Quality.

Colborn, Theo (1995). Address to the Third Plenary Session at the EPA's Endocrine Disruptor Research Needs Workshop, April 10-13, Raleigh, North Carolina.

Colborn, Theo, Frederick vom Saal, and Ana Soto (1993). "Developmental Effects of Endocrine-Disrupting Chemicals in Wildlife and Humans," *Environmental Health Perspectives*, Vol. 101, No. 5, October, p. 378-384.

Colborn, Theo and Coralie Clement, eds. (1992). *Chemically-Induced Alterations in Sexual and Functional Development: The Wildlife/Human Connection*, Princeton, New Jersey: Princeton Scientific Publishing.

Davis, Devra Lee, H. Leon Bradlow, Mary Wolff, Tracey Woodruff, David Hoel, and Hoda Anton-Culver (1993). "Medical Hypothesis: Xenoestrogens as Preventable Causes of Breast Cancer," *Environmental Health Perspectives*, Vol. 101, No. 5, October, pp. 372-377.

Devoret, Raymond (1979). "Bacterial Tests for Potential Carcinogens," *Scientific American*, Vol. 241, No. 2, August, pp. 40-49.

EPA (1995). *New Chemicals Program Chemical Categories*, produced by the Office of Prevention, Pesticides and Toxic substances.

EPA (1991). *EPA's Pesticide Programs*. 21T-1005, May.

EPA (1991). "Guidelines for Developmental Toxicity Risk Assessment," *Federal Register*, Vol. 56, No. 234, December 5, pp. 63798-63824.

EPA (1986). "Guidelines for Carcinogen Risk Assessment," *Federal Register*, Vol. 51, September 24, pp. 33992-34003.

GAO (1994). *Toxic Substances Control Act: Legislative Changes Could Make the Act More Effective*, GAO/RCED-94-103, September.

GAO (1993). *Pesticide Reregistration May Not Be Completed Until 2006*, GAO/RCED-93-94, May.

GAO (1991). *Reproductive and Developmental Toxicants: Regulatory Actions Provide Uncertain Protection*, GAO/PEMD-92-3, October.

Gough, Michael (1990). "How Much Cancer Can EPA Regulate Away?" *Risk Analysis*, Vol. 10, No. 1, pp. 1-6.

Hathaway, Carolyne, David Hayes, and William Rawson (1994). "A Practitioner's Guide to the Toxic Substances Control Act: Part I," *Environmental Law Reporter News and Analysis*, Vol. 24, No. 5, May, p.10207-10230.

Hertz, Roy (1985). "The Estrogen Problem: Retrospect and Prospect," in *Estrogens in the Environment II: Influences on Development*, McLachlan, ed., New York: Elsevier Science Publishing.

Hileman, Bette (1994). "Environmental Estrogens Linked to Reproductive Abnormalities, Cancer," *Chemical & Engineering News*, January 31, pp. 19-23.

Hollstein, Monica and Joyce McCann (1979). "Short Term Tests for Carcinogens and Mutagens," *Mutation Research*, Vol. 65, pp. 133-226.

James, Robert, and Christopher Teaf (1985). "Carcinogenesis," in *Industrial Toxicology*, Williams and Burson, eds., New York: Van Nostrand Reinhold, pp. 301-342.

Kelce, William, Christy Stone, Susan Laws, Earl Gray, Jon Kemppalnen, and Elizabeth Wilson (1995). "Persistent DDT Metabolite p,p'-DDE is a Potent Androgen Receptor Antagonist," *Nature*, Vol. 375, June 15, pp. 581-585.

Liu, Hong, Mark Wormke, Stephen Safe, Leonard Bjeldanes (1994). "Indole [3,2-b]carbazole: a Dietary-Derived Factor That Exhibits Both Antiestrogenic and Estrogenic Activity," *Journal of the National Cancer Institute*, Vol. 86, No. 23, December 7, pp. 1758-1765.

MacMahon, Brian (1994). "Pesticide Residues and Breast Cancer?" *Journal of the National Cancer Institute*, Vol. 86, No. 8, April 20, pp. 572-573.

Masters, Gilbert (1991). *Introduction to Environmental Engineering and Science*, Englewood, New Jersey: Prentice-Hall, Inc.

Maugh, Thomas (1978). "Chemical Carcinogenesis: The Scientific Basis for Regulation," *Science*, Vol. 201, September 29, pp. 1200-1205.

Mayr, Ulrich, Annette Butsch, and Susanne Schneider (1992). "Validation of Two In Vitro Test Systems for Estrogenic Activities with Zearalenone, Phytoestrogens and Cereal Extracts," *Toxicology*, Vol. 74, pp. 135-149.

McCann, Joyce, Edmund Choi, Edith Yamasaki, and Bruce Ames (1976). "Detection of Carcinogens as Mutagens in the Salmonella/Microsome Test: Assay of 300 Chemicals: Discussion," *Proceedings of the National Academy of Sciences USA*, Vol. 73, No. 3, March, pp. 950-954.

McCann, Joyce, Edmund Choi, Edith Yamasaki, and Bruce Ames (1975). "Detection of Carcinogens as Mutagens in the Salmonella/Microsome Test: Assay of 300 Chemicals," *Proceedings of the National Academy of Sciences USA*, Vol. 72, No. 12, December 1975, pp. 5135-5139.

McLachlan, John (1993). "Functional Toxicology: A New Approach to Detect Biologically Active Xenobiotics," *Environmental Health Perspectives*, Vol. 101, No. 5, pp. 386-387.

McLachlan, John, ed. (1985). *Estrogens in the Environment II: Influences on Development*, New York: Elsevier Scientific Publishing.

McLachlan, John, ed. (1980). *Estrogens in the Environment*, New York: Elsevier Scientific Publishing.

Miller, James (1994). "Brief History of Chemical Carcinogenesis," *Cancer Letters*, Vol. 83, pp. 9-14.

NRC Commission on Life Sciences (1995). "Hormone-related Toxicants in the Environment," study prospectus.

NRC (1984). *Toxicity Testing: Strategies to Determine Needs and Priorities*, Washington, DC: National Academy Press.

NRC (1983). *Risk Assessment in the Federal Government: Managing the Process*, Washington, DC: National Academy Press.

OSHA (1980). "Identification, Classification and Regulation of Potential Occupational Carcinogens," *Federal Register*, Vol. 45, No. 15, January 22, pp. 5002-5282.

OTA (1987). *Identifying and Regulating Carcinogens*, Washington, DC: US Government Printing Office.

OTA (1981). *Assessment of Technologies for Determining Cancer Risks from the Environment*, Washington, DC: US Government Printing Office.

Paschkis, Karl, Abraham Rakoff, Abraham Cantarow, and Joseph Rupp (1967). *Clinical Endocrinology*, New York: Hoeber Medical Division of Harper & Row.

Patterson, James (1987). *The Dread Disease: Cancer and Modern American Culture*, Cambridge, Massachussetts: Harvard University Press.

Purchase, IFH, E Longstaff, John Ashby, JA Styles, Diana Anderson, PA Lefevre, and FR Westwood (1976). "Evaluation of Six Short Term Tests for Detecting Organic Chemical Carcinogens and Recommendations for Their Use," *Nature*, Vol. 264, December 16, pp. 624-627.

Rennie, John (1993). "Malignant Mimicry," *Scientific American*, September, pp. 34-35.

Rettig, Richard (1977). *Cancer Crusade: The Story of the National Cancer Act of 1971*, Princeton, NJ: Princeton University Press.

Rosen, Joseph (1990). "Much Ado About Alar," *Issues in Science and Technology*, Fall, pp. 85-90.

Rubin, Harry (1976). "Carcinogenicity Tests," *Science*, Vol. 191, January 23, p. 241.

Rushefsky, Mark (1986). *Making Cancer Policy*. Albany, NY: State University of New York Press.

Sharpe, Richard (1995). "Another DDT Connection," *Nature*, Vol. 375, June 15, pp. 538-539.

Sharpe, RM (1993). "Declining Sperm Counts in Men: Is There an Endocrine Cause?" *Journal of Endocrinology*, Vol. 136, pp. 357-360.

Soto, Ana, Carlos Sonnenschein, Kerrie Chung, Mariana Fernandez, Nicolas Olea, and Fatima Olea Serrano (in press). "The E-screen Assay as a Tool to Identify Estrogens: An Update on Estrogenic Environmental Pollutants," *Environmental Health Perspectives*.

Soto, Ana (1995). Presentation given at the MIT Seminar on the Management of the Future Uses of Chlorine, April 18.

Soto, Ana, Tien-Min Lin, Honorato Justicia, Renee Silvia, and Carlos Sonnenschein (1992a). "An 'In Culture' Bioassay to Assess the Estrogenicity of Xenobiotics (E-screen)," in *Chemically-Induced Alterations in Sexual and Functional Development: the Wildlife/Human Connection*, Colborn and Clement, eds., Princeton, New Jersey: Princeton Scientific Publishing

Soto, Ana, Renee Silvia and Carlos Sonnenschein (1992b). "A Plasm-Borne Specific Inhibitor of the Proliferation of Human Estrogen-Sensitive Breast Tumor Cells (Estrocolyone-I)," *Journal of Steroid Biochemistry and Molercular Biology*, Vol. 43, No. 7, pp. 703-712.

Soto, Ana, Honorato Justicia, Jonathan Wray, and Carlos Sonnenschein (1991). "p-Nonyl-Phenol: An Estrogenic Xenobiotic Released from 'Modified' Polystyrene," *Environmental Health Perspectives*, Vol. 92, pp. 167-173.

Soto, Ana and Carlos Sonnenschein (1985). "The Role of Estrogens on the Proliferation of Human Breast Tumor Cells (MCF-7)," *Journal of Steroid Biochemistry*, Vol. 23, No. 1, pp. 87-94.

Stone, Richard (1994). "Environmental Estrogens Stir Debate," *Science*, Vol. 265, July 15, pp. 308-310.

Strickland, Stephen (1972). *Politics, Science, and the Dread Disease.* Cambridge, Massachussetts: Harvard University Press.

Teaf, Christopher (1985). "Mutagenesis," in *Industrial Toxicology*, Williams and Burson, eds., New York: Van Nostrand Reinhold, pp. 273-300.

Tennant, Raymond, Barry Margolin, Michael Shelby, et al. (1987). "Prediction of Chemical Carcinogenicity in Rodents from in Vitro Genetic Toxicity Assays," *Science*, Vol. 236, May 22, pp. 933-941.

US Congress (1993). "Health Effects of Estrogenic Pesticides," hearing before the House of Representatives, Committee on Energy and Commerce, Subcommittee on Health and Environment, October 21.

US Congress (1992). "Toxic Substances Control: Still Waiting All These Years," hearing before the House of Representatives, Committee on Government Operations, Subcommittee on Environment, Energy, and Natural Resources, March 18.

Weiss, Rick (1994). "Estrogen in the Environment: Are Some Pollutants a Threat to Fertility?" *The Washington Post*, January 25.

Wildavsky, Aaron (1990). "No Risk is the Highest Risk of All," in *Readings in Risk*, Theodore Glickman and Michael Gough, eds., Washington, DC: Resources for the Future, p. 121.

Wolff, Mary, Paolo Toniolo, Eric Lee, Marilyn Rivera, and Neil Dubin (1993). "Blood Levels of Organochlorine Residues and Risk of Breast Cancer," *Journal of the National Cancer Institute*, Vol. 85, No. 8, April 21, pp. 648-652.